ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2023)

Rhodes Island, Greece
4-10 June 2023

Pages 1-365
# TABLE OF CONTENTS

Counterfactual Two-Stage Debiasing for Video Corpus Moment Retrieval
Sunjae Yoon, Ji Woo Hong, Soohwan Eom, Hee Suk Yoon, Eunseop Yoon, Daehyeok Kim, Junyeong Kim, Chanwoo Kim, Chang D. Yoo ................................................. 1

Test-Time Training-Free Domain Adaptation
Yongxiang Feng, Weihua He, Kaichao You, Bing Liu, Ziyang Zhang, Yaoyuan Wang, Minglei Li, Yihang Lou, Jiawei Li, Guoqi Li, Jianxing Liao ........................................ 6

AdapITN: A Fast, Reliable, and Dynamic Adaptive Inverse Text Normalization
Thai-Binh Nguyen, Le Duc Minh Nhat, Quang Minh Nguyen, Quoc Truong Do, Chi Mai Luong, Alexander Waibel .......................................................... 11

Compressed Distributed Regression Over Adaptive Networks
Marco Carpentiero, Vincenzo Matta, Ali H. Sayed .................................................. 16

The Role of Memory in Social Learning When Sharing Partial Opinions
Michele Cirillo, Virginia Bordignon, Vincenzo Matta, A. H. Sayed ...................... 21

IoU-Aware Multi-Expert Cascade Network Via Dynamic Ensemble for Long-Tailed Object Detection
Wan-Cyuan Fan, Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu ........................... 26

TransLink: Transformer-Based Embedding for Tracklets' Global Link
Yanting Zhang, Shuanghong Wang, Yuxuan Fan, Gaoang Wang, Cairong Yan .... 31

Burst Perception-Distortion Tradeoff: Analysis and Evaluation
Danna Xue, Luis Herranz, Javier Vazquez Corral, Yanning Zhang ...................... 36

Approximation Error Back-Propagation for Q-Function in Scalable Reinforcement Learning with Tree Dependency Structure
Yuzi Yan, Yu Dong, Kai Ma, Yuan Shen .............................................................. 41

EMC²-Net: Joint Equalization and Modulation Classification Based on Constellation Network
Hyun Ryu, Junil Choi ......................................................................................... 46

Improving Prosody for Cross-Speaker Style Transfer by Semi-Supervised Style Extractor and Hierarchical Modeling in Speech Synthesis
Chunyu Qiang, Peng Yang, Hao Che, Ying Zhang, Xiaorui Wang, Zhongyuan Wang .............................................................. 51

CyFi-TTS: Cyclic Normalizing Flow with Fine-Grained Representation for End-to-End Text-to-Speech
In-Sun Hwang, Young-Sub Han, Byoung-Ki Jeon .............................................. 56

Learning Graph Laplacian from Intrinsic Patterns Via Gaussian Process
Koshi Watanabe, Keisuke Maeda, Takahiro Ogawa, Miki Haseyama .................. 61

Overcoming the Seesaw in Monocular 3D Object Detection Via Language Knowledge Transferring
Weichen Xu, Tianhao Fu ................................................................................... 66

Improving Dropout in Graph Convolutional Networks for Recommendation Via Contrastive Loss
Hiroki Okamura, Keisuke Maeda, Ren Togo, Takahiro Ogawa, Miki Haseyama .... 71
Estimation of Visual Contents from Human Brain Signals Via VQA Based on Brain-Specific Attention
Ryo Shichida, Ren Togo, Keisuke Maeda, Takahiro Ogawa, Miki Haseyama

Combining the Silhouette and Skeleton Data for Gait Recognition
Likai Wang, Ruize Han, Wei Feng

Unobtrusive Respiratory Monitoring System for Intensive Care
Xudong Tan, Menghan Hu, Guangtao Zhai, Yan Zhu, Wenfang Li, Xiaoping Zhang

End-To-End Amp Modeling: from Data to Controllable Guitar Amplifier Models
Lauri Juvela, Eero-Pekka Damskägg, Aleksi Peussa, Jaakko Mäkinen, Thomas Sherson, Stylianos I. Mimilakis, Kimmo Rauhanen, Athanasios Gotsopoulos

Provably Convergent Plug & Play Linearized ADMM, Applied to Deblurring Spatially Varying Kernels
Charles Laroche, Andrés Almansa, Eva Coupeté, Matias Tassano

Robust Video Anomaly Detection Framework Via Prior Knowledge and Multi-Path Frame Prediction
Menghao Zhang, Jingyu Wang, Jing Wang, Qi Qi, Zirui Zhuang, Haiqiang Sun, Ning Xiao

Unrolled Fourier Disparity Layer Optimization for Scene Reconstruction from Few-Shots Focal Stacks
Brandon Le Bon, Mikaël Le Pendu, Christine Guillemot

Adaptive Axonal Delays in Feedforward Spiking Neural Networks for Accurate Spoken Word Recognition
Pengfei Sun, Ehsan EQlimi, Yansong Chua, Paul Devos, Dick Botteldooren

EEG2IMAGE: Image Reconstruction from EEG Brain Signals
Prajwal Singh, Pankaj Pandey, Krishna Miyapuram, Shammuganathan Raman

Multispectral Image Fusion Based on Super Pixel Segmentation
Nati Ofir

NF-PCAC: Normalizing Flow Based Point Cloud Attribute Compression
Rodrigo B. Pinheiro, Jean-Eudes Marvie, Giuseppe Valenzise, Frédéric Dufaux

An Adaptive DFE Using Light-Pattern-Protection Algorithm in 12 NM CMOS Technology
Shiyuan Xing, Changlong Lin, Yuchen Li, Huandong Wang

Hierarchical Transformer for Multi-Label Trailer Genre Classification
Zihui Cai, Hongwei Ding, Xuemeng Wu, Mohan Xu, Xiaohui Cui

A Contrastive Framework to Enhance Unsupervised Sentence Representation Learning
Haoyang Ma, Zeyu Li, Hongyu Guo

Self-Attention Based Action Segmentation Using Intra-And Inter-Segment Representations
Constantin Patsch, Eckehard Steinbach

A Deep Disentangled Approach for Interpretable Hyperspectral Unmixing
Ricardo Augusto Borsoi, Tales Imbiriba, Deniz Erdogmus

Residual Squeeze-And-Excitation U-Shaped Network for Minutia Extraction in Contactless Fingerprint Images
Anderson Nogueira Cotrim, Helio Pedrini
Learnable Flow Model Conditioned on Graph Representation Memory for Anomaly Detection
Ziyu Zhu, Wenlei Liu, Zhidong Deng

Building Blocks for a Complex-Valued Transformer Architecture
Florian Eilers, Xiaoyi Jiang

PCSalmix: Gradient Saliency-Based Mix Augmentation for Point Cloud Classification
Tao Hong, Zeren Zhang, Jinwen Ma

Entropy Based Feature Regularization to Improve Transferability of Deep Learning Models
Raphael Baena, Lucas Drumetz, Vincent Gripon

Exploring Complementary Features in Multi-Modal Speech Emotion Recognition
Suzhen Wang, Yifeng Ma, Yu Ding

Database-Aware ASR Error Correction for Speech-to-SQL Parsing
Yutong Shao, Arun Kumar, Ndapa Nakashole

Super-Resolution Harmonic Retrieval of Non-Circular Signals
Yu Zhang, Yue Wang, Zhi Tian, Geert Leus, Gong Zhang

Semi-Swinderain: Semi-Supervised Image Deraining Network Using SWIN Transformer
Chun Ren, Danfeng Yan, Yuanqiang Cai, Yangchun Li

Learning Generalizable Light Field Networks from Few Images
Qian Li, Franck Multon, Adnane Boukhayma

Distributionally Robust Multiclass Classification and Applications in Deep Image Classifiers
Ruidi Chen, Boran Hao, Ioannis Ch. Paschalidis

Adaptive Mask Co-Optimization for Modal Dependence in Multimodal Learning
Ying Zhou, Xuefeng Liang, Shiquan Zheng, Huijun Xuan, Takatsune Kumada

Improving Occluded Human Pose Estimation Via Linked Joints
Suhang Ye, Zebo Hong, Jiawen Zheng, Shengchuan Zhang

DST: Deformable Speech Transformer for Emotion Recognition
Weidong Chen, Xiaofen Xing, Xiangmin Xu, Jianxin Pang, Lan Du

Permutation Invariant Training for Paraphrase Identification
Jun Bai, Chuantao Yin, Hanhua Hong, Jianfei Zhang, Chen Li, Yanmeng Wang, Wenge Rong

Code-Enhanced Fine-Grained Semantic Matching for Tag Recommendation in Software Information Sites
Lin Li, Peipei Wang, Xinhao Zheng, Qing Xie

Audio Coding with Unified Noise Shaping and Phase Contrast Control
Byeongho Jo, Seungkwon Beack, Taejin Lee

Binary Image Fast Perfect Recovery from Sparse 2D-DFT Coefficients
Soo-Chang Pei, Kuo-Wei Chang

Reducing the GAP Between Streaming and Non-Streaming Transducer-Based ASR by Adaptive Two-Stage Knowledge Distillation
Haitao Tang, Yu Fu, Lei Sun, Jiabin Xue, Dan Liu, Yongchao Li, Zhiqiang Ma, Minghui Wu, Jia Pan, Genshun Wan, Ming En Zhao
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncertainty-Aware Few-Shot Class-Incremental Learning</td>
<td>251</td>
</tr>
<tr>
<td>Jiancai Zhu, Jiabao Zhao, Jiayi Zhou, Liang He, Jing Yang, Zhi Zhang</td>
<td></td>
</tr>
<tr>
<td>Flowreg: Latent Space Regularization Using Normalizing Flow for Limited Samples Learning</td>
<td>256</td>
</tr>
<tr>
<td>Chi Wang, Jian Gao, Yang Hua, Hui Wang</td>
<td></td>
</tr>
<tr>
<td>Promoting Cooperation in Multi-Agent Reinforcement Learning Via Mutual Help</td>
<td>261</td>
</tr>
<tr>
<td>Yunbo Qiu, Yue Jin, Lebin Yu, Jian Wang, Xudong Zhang</td>
<td></td>
</tr>
<tr>
<td>Semi-Supervised Semantic Segmentation with Structured Output Space Adaption</td>
<td>266</td>
</tr>
<tr>
<td>Weiquan Huang, Fu Zhang</td>
<td></td>
</tr>
<tr>
<td>Conditional LS-GAN Based Skylight Polarization Image Restoration and Application in Meridian Localization</td>
<td>271</td>
</tr>
<tr>
<td>Tian Yang, Hongbo Bo, Xinyu Yang, Jun Gao, Zijian Shi</td>
<td></td>
</tr>
<tr>
<td>Towards Trustworthy Multi-Label Sewer Defect Classification Via Evidential Deep Learning</td>
<td>276</td>
</tr>
<tr>
<td>Chenyang Zhao, Chuanfei Hu, Hang Shao, Zhe Wang, Yongxiong Wang</td>
<td></td>
</tr>
<tr>
<td>Learning from the Raw Domain: Cross Modality Distillation for Compressed Video Action Recognition</td>
<td>281</td>
</tr>
<tr>
<td>Yufan Liu, Jiajiong Cao, Weiming Bai, Bing Li, Weiming Hu</td>
<td></td>
</tr>
<tr>
<td>Identifying Opinion Influencers Over Social Networks</td>
<td>286</td>
</tr>
<tr>
<td>Valentina Shumovskaya, Mert Kayaalp, Ali H. Sayed</td>
<td></td>
</tr>
<tr>
<td>Beamforming Optimization in RIS-Aided Mimo Systems Under Multiple-Reflection Effects</td>
<td>291</td>
</tr>
<tr>
<td>Dilki Wijekoon, Amine Mezghani, Ekram Hossain</td>
<td></td>
</tr>
<tr>
<td>MLP-GAN for Brain Vessel Image Segmentation</td>
<td>296</td>
</tr>
<tr>
<td>Bin Xie, Hao Tang, Bin Duan, Dawen Cai, Yan Yan</td>
<td></td>
</tr>
<tr>
<td>Image Sharing Chain Detection VIA Sequence-To-Sequence Model</td>
<td>301</td>
</tr>
<tr>
<td>Jiaxiang You, Yuanman Li, Rongqin Liang, Yuxuan Tan, Jiantao Zhou, Xia Li</td>
<td></td>
</tr>
<tr>
<td>Input-Dependent Dynamical Channel Association for Knowledge Distillation</td>
<td>306</td>
</tr>
<tr>
<td>Qiankun Tang, Yuan Zhang, Xiaogang Xu, Jun Wang, Yimin Guo</td>
<td></td>
</tr>
<tr>
<td>An Adaptive Plug-And-Play Network for Few-Shot Learning</td>
<td>311</td>
</tr>
<tr>
<td>Hao Li, Li Li, Yunneng Huang, Ning Li, Yongtao Zhang</td>
<td></td>
</tr>
<tr>
<td>Decontamination Transformer for Blind Image Inpainting</td>
<td>316</td>
</tr>
<tr>
<td>Chun-Yi Li, Yen-Yu Lin, Wei-Chen Chiu</td>
<td></td>
</tr>
<tr>
<td>Prefix Tuning for Automated Audio Captioning</td>
<td>321</td>
</tr>
<tr>
<td>Minkyu Kim, Kim Sung-Bin, Tae-Hyun Oh</td>
<td></td>
</tr>
<tr>
<td>SCSGNet: Spatial-Correlated and Shape-Guided Network for Breast Mass Segmentation</td>
<td>326</td>
</tr>
<tr>
<td>Qingqiu Li, Jilan Xu, Runtian Yuan, Yuejie Zhang, Rui Feng</td>
<td></td>
</tr>
<tr>
<td>Hyperspectral Image Denoising Via Nonlocal Rank Residual Modeling</td>
<td>331</td>
</tr>
<tr>
<td>Zhiyuan Zha, Bihan Wen, Xin Yuan, Jiantao Zhou, Ce Zhu</td>
<td></td>
</tr>
<tr>
<td>PRRD: Pixel-Region Relation Distillation for Efficient Semantic Segmentation</td>
<td>336</td>
</tr>
<tr>
<td>Chen Wang, Jiang Zhong, Qizhu Dai, Yafei Qi, Rongzhen Li, Qin Lei, Bin Fang, Xue Li</td>
<td></td>
</tr>
</tbody>
</table>
Improved Indoor Localization with NLOS Signal Propagations
Wei Huang, Yixin Zhao, Xuechao Wu, Le Yin

HQP-MVS: High-Quality Plane Priors Assisted Multi-View Stereo for Low-Textured Areas
Zefan Tian, Rongjie Wang, Zhenyu Wang, Ronggang Wang

MSN-Net: Multi-Scale Normality Network for Video Anomaly Detection
Yang Liu, Di Li, Wei Zhu, Dingkang Yang, Jing Liu, Liang Song

High-Dynamic Range ADC for Finite-Rate-of-Innovation Signals
Satish Mulleti, Yonina C. Eldar

Lip-To-Speech Synthesis in the Wild with Multi-Task Learning
Minsu Kim, Joanna Hong, Yong Man Ro

NC-WAMKD: Neighborhood Correction Weight-Adaptive Multi-Teacher Knowledge Distillation for Graph-Based Semi-Supervised Node Classification
Jiahao Liu, Pengcheng Guo, Yonghong Song

Toward Asymptotic Optimality: Sequential Unsupervised Regression of Density Ratio for Early Classification
Akinori F. Ebihara, Taiki Miyagawa, Kazuyuki Sakurai, Hitoshi Imaoka

Improving Image Captioning with Control Signal of Sentence Quality
Zhangzi Zhu, Shuai Wang, Hong Qu

Multiple Acoustic Features Speech Emotion Recognition Using Cross-Attention Transformer
Yurun He, Nobuaki Minematsu, Daisuke Saito

Graph Neural Networks for Object Type Classification Based on Automotive Radar Point Clouds and Spectra
Loveneet Saini, Axel Acosta, Gor Hakobyan

Frame-Level Multi-Label Playing Technique Detection Using Multi-Scale Network and Self-Attention Mechanism
Dichucheng Li, Mingjin Che, Wenwu Meng, Yulun Wu, Yi Yu, Fan Xia, Wei Li

FedPrompt: Communication-Efficient and Privacy-Preserving Prompt Tuning in Federated Learning
Haodong Zhao, Wei Du, Fangqi Li, Peixuan Li, Gongshen Liu

Uncer2Natural: Uncertainty-Aware Unsupervised Image Denoising
Chenyu Huang, Weimin Tan, Jiaxing Shi, Zhen Xing, Bo Yan

Simple Pooling Front-Ends for Efficient Audio Classification
Xubo Liu, Haohe Liu, Qiuqiang Kong, Xinhao Mei, Mark D. Plumbley, Wenwu Wang

A Compensated Shrinkage Affine Projection Algorithm for Debiased Sparse Adaptive Filtering
Yi Zhang, Isao Yamada

CTCBERT: Advancing Hidden-Unit Bert with CTC Objectives
Ruchao Fan, Yiming Wang, Yashesh Gaur, Jinyu Li

Analysis and Re-Synthesis of Natural Cricket Sounds Assessing the Perceptual Relevance of Idiosyncratic Parameters
Marco Oliveira, Vitor Almeida, João Silva, Anibal Ferreira
Hierarchical Spatial-Temporal Transformer with Motion Trajectory for Individual Action and Group Activity Recognition ................................................................. 426
   Xiaolin Zhu, Dongli Wang, Yan Zhou

S3I-PointHop: SO(3)-Invariant PointHop for 3D Point Cloud Classification ................................................................. 431
   Pranav Kadam, Hardik Prajapati, Min Zhang, Jintang Xue, Shan Liu, C.-C. Jay Kuo

Salient Co-Speech Gesture Synthesizing with Discrete Motion Representation ................................................................. 436
   Zijie Ye, Jia Jia, Haozhe Wu, Shuo Huang, Shikun Sun, Junliang Xing

Hierarchical Diffusion Models for Singing Voice Neural Vocoder .................................................................................. 441
   Naoya Takahashi, Mayank Kumar, Singh, Yuki Mitsufuji

Period VITS: Variational Inference with Explicit Pitch Modeling for End-To-End Emotional Speech Synthesis ................................................................. 446
   Yuma Shirahata, Ryuichi Yamamoto, Eunwoo Song, Ryo Terashima, Jae-Min Kim, Kentaro Tachibana

Sanet: Spatial Attention Network with Global Average Contrast Learning for Infrared Small Target Detection ................................................................. 451
   Jiewen Zhu, Shengjia Chen, Lexiao Li, Luping Ji

An ASR-Free Fluency Scoring Approach with Self-Supervised Learning ................................................................. 456
   Wei Liu, Kaiqi Fu, Xiaohai Tian, Shuju Shi, Wei Li, Zejun Ma, Tan Lee

Memory-Augmented U-Transformer for Multivariate Time Series Anomaly Detection ................................................................. 461
   Shuxin Qin, Yongcan Luo, Gaofeng Tao

Optimizing Distributed Multi-Sensor Multi-Target Tracking Algorithm Based on Labeled Multi-Bernoulli Filter ........................................................................ 466
   Honggang Liu, Jinhong Yang, Yue Xu, Le Yang

Rethink Pair-Wise Self-Supervised Cross-Modal Retrieval from a Contrastive Learning Perspective .......... 471
   Tiantian Gong, Junsheng Wang, Liyan Zhang

Grad-StyleSpeech: Any-Speaker Adaptive Text-to-Speech Synthesis with Diffusion Models ................................................................. 476
   Minki Kang, Dongchan Min, Sung Ju Hwang

Classification Via Subspace Learning Machine (SLM): Methodology and Performance Evaluation .................. 481
   Hongyu Fu, Yijing Yang, Vinod K. Mishra, C.-C. Jay Kuo

T5-SR: A Unified Seq-to-Seq Decoding Strategy for Semantic Parsing ................................................................. 486
   Yuntao Li, Zhenpeng Su, Yutian Li, Hanchu Zhang, Sirui Wang, Wei Wu, Yan Zhang

Two-Stream Decoder Feature Normality Estimating Network for Industrial Anomaly Detection .................. 491
   Chaewon Park, Minhyeok Lee, Suhwan Cho, Donghyeong Kim, Sangyoun Lee

TriAAN-VC: Triple Adaptive Attention Normalization for Any-to-Any Voice Conversion .................. 496
   Hyun Joon Park, Seok Woo Yang, Jin Sob Kim, Wooseok Shin, Sung Won Han

AD-YOLO: You Look Only Once in Training Multiple Sound Event Localization and Detection .... 501
   Jin Sob Kim, Hyun Joon Park, Wooseok Shin, Sung Won Han

   Tomas Kerepecky, Jiaming Liu, Xue Wen Ng, David W. Piston, Ulugbek S. Kamilov
Global-Context Aware Generative Protein Design ................................................................. 511
Cheng Tan, Zhangyang Gao, Jun Xia, Bozhen Hu, Stan Z. Li

Semi-Supervised Speech Enhancement Based on Speech Purity .......................................... 516
Zihao Cui, Shilei Zhang, Yanan Chen, Yingying Gao, Chao Deng, Junlan Feng

Modify: Model-Driven Face Stylization Without Style Images ........................................... 521
Yuhe Ding, Jian Liang, Jie Cao, Aihua Zheng, Ran He

Precognition in Contextual Spoken Language Understanding Via Knowledge Distillation ........ 526
Nan Su, Bingzhu Du, Yuchi Zhang, Chao Liu, Yongliang Wang, Hong Chen, Xin Lu

GaPP: Multi-Target Tracking with Gaussian Processes ....................................................... 531
Fred Goodyer, Bashar I. Ahmad, Simon Godsill

Adversarial Network Pruning by Filter Robustness Estimation .......................................... 536
Xinlu Zhuang, Yunjie Ge, Baolin Zheng, Qian Wang

Leveraging Phone-Level Linguistic-Acoustic Similarity for Utterance-Level Pronunciation Scoring ........ 541
Wei Liu, Kaiqi Fu, Xiaohai Tian, Shuju Shi, Wei Li, Zejun Ma, Tan Lee

CTTSR: A Hybrid CNN-Transformer Network for Scene Text Image Super-Resolution ........... 546
Kaiwei Dai, Nan Kang, Li Kuang

CF-VTON: Multi-Pose Virtual Try-on with Cross-Domain Fusion ...................................... 551
Chenghu Du, Shengwu Xiong

In Search of Strong Embedding Extractors for Speaker Diarisation .................................... 556
Jee-Weon Jung, Hee-Soo Heo, Bong-Jin Lee, Jaesung Huh, Andrew Brown, Youngki Kwon, Shinji Watanabe, Joon Son Chung

Towards Making a Trojan-Horse Attack on Text-to-Image Retrieval ................................... 561
Fan Hu, Aozhu Chen, Xirong Li

Elastic Graph Transformer Networks for EEG-Based Emotion Recognition .......................... 566
Wei-Bang Jiang, Xu Yan, Wei-Long Zheng, Bao-Liang Lu

Raw Ultrasound-Based Phonetic Segments Classification Via Mask Modeling ....................... 571
Kang You, Bo Liu, Kele Xu, Yunsheng Xiong, Qisheng Xu, Ming Feng, Tamás Gábor Csapó, Boqing Zhu

Content-Insensitive Dynamic Lip Feature Extraction for Visual Speaker Authentication Against Deepfake Attacks ................................................................. 576
Zihao Guo, Shilin Wang

Sample-Aware Knowledge Distillation for Long-Tailed Learning ....................................... 581
Shanshan Zheng, Yachao Zhang, Hongyi Huang, Yanyun Qu

A3S: Adversarial Learning of Semantic Representations for Scene-Text Spotting .................. 586
Masato Fujitake

A Novel Cross-Component Context Model for End-to-End Wavelet Image Coding ................. 591
Anna Meyer, André Kaup

Multi-Head Uncertainty Inference for Adversarial Attack Detection .................................... 596
Yuqi Yang, Songyun Yang, Jiyang Xie, Zhongwei Si, Kai Guo, Ke Zhang, Kongming Liang
ESCL: Equivariant Self-Contrastive Learning for Sentence Representations .......................................................... 601
  Jie Liu, Yixuan Liu, Xue Han, Chao Deng, Junlan Feng

Vision Transformer-Based Feature Extraction for Generalized Zero-Shot Learning ............................................. 606
  Jiseob Kim, Kyuhong Shim, Junhan Kim, Byonghyo Shim

A Robust Kalman Filter Based Approach for Indoor Robot Positioning with Multi-Path Contaminated UWB Data ......................................................... 611
  Justin Cano, Yi Ding, Gaël Pages, Eric Chaumette, Jerome Le Ny

Semi-Supervised Sound Event Detection with Pre-Trained Model .............................................................................. 616
  Liang Xu, Lizhong Wang, Sijun Bi, Hanyue Liu, Jing Wang

Design Choices for Learning Embeddings from Auxiliary Tasks for Domain Generalization in Anomalous Sound Detection ........................................................................................................ 621
  Kevin Wilkinghoff

Tensor-Based Complex-valued Graph Neural Network for Dynamic Coupling Multimodal Brain Networks .......................................................... 626
  Yanwu Yang, Guoqing Cai, Chenfei Ye, Yang Xiang, Ting Ma

Sparse Graph Learning with Spectrum Prior for Deep Graph Convolutional Networks ........................................ 631
  Jin Zeng, Yang Liu, Gene Cheung, Wei Hu

Divcon: Learning Concept Sequences for Semantically Diverse Image Captioning ................................................. 636
  Yue Zheng, Ya-Li Li, Shengjin Wang

EBEN: Extreme Bandwidth Extension Network Applied to Speech Signals Captured with Noise-Resilient Body-Conduction Microphones ........................................................................... 641
  Julien Hauret, Thomas Joubaud, Véronique Zimpfer, Éric Bavu

Long Range Imaging Using Multispectral Fusion of RGB and NIR Images .......................................................... 646
  Hao Zhang, Lin Mei, Cheolkon Jung

FedEEG: Federated EEG Decoding Via inter-Subject Structure Matching .......................................................... 651
  Wenlong Hang, Jiaxing Li, Shuang Liang, Yuan Wu, Baiying Lei, Jing Qin, Yu Zhang, Kup-Sze Choi

Visual Onoma-To-Wave: Environmental Sound Synthesis from Visual Onomatopoeias and Sound-Source Images ......................................................................................... 656
  Hien Ohnaka, Shinnosuke Takamichi, Keisuke Imoto, Yuki Okamoto, Kazuki Fujii, Hiroshi Saruwatari

Asynchronous Social Learning ................................................................................................................................. 661
  Mert Cemri, Virginia Bordignon, Mert Kayaalp, Valentina Shumovskaya, Ali H. Sayed

Untargeted Backdoor Attack Against Object Detection .......................................................................................... 666
  Chengxiao Luo, Yiming Li, Yong Jiang, Shu-Tao Xia

Improving Contextual Biasing with Text Injection ................................................................................................. 671
  Tara N. Sainath, Rohit Prabhavalkar, Diamantino Caseiro, Pat Rondon, Cyril Allauzen

Noise PSD Insensitive RTF Estimation in a Reverberant and Noisy Environment .................................................. 676
  Changheng Li, Richard C. Hendriks

Error Analysis of Convolutional Beamspace Algorithms ......................................................................................... 681
  Po-Chih Chen, P. P. Vaidyanathan
Unitary Esprit for Coprime Arrays ................................................................. 686  
Po-Chih Chen, P. P. Vaidyanathan

Kashyap Patel, Anton Kovalyov, Issa Panahi

SARdBScene: Dataset and Resnet Baseline for Audio Scene Source Counting and Analysis .................. 696  
Michael Nigro, Sridhar Krishnan

CD-FSOD: A Benchmark for Cross-Domain Few-Shot Object Detection ............................................. 701  
Wuti Xiong

Fast and Accurate Factorized Neural Transducer for Text Adaption of End-To-End Speech  
Recognition Models ......................................................................................... 706  
Rui Zhao, Jian Xue, Partha Parthasarathy, Veljko Miljanic, Jinyu Li

Augmentation Robust Self-Supervised Learning for Human Activity Recognition ................................. 711  
Cong Xu, Yuhang Li, Dae Lee, Dae Hoon Park, Hongda Mao, Huyen Do, Jonathan Chung,  
Dinesh Nair

SPADE: Self-Supervised Pretraining for Acoustic Disentanglement .................................................. 716  
John Harvill, Jarred Barber, Arun Nair, Ramin Pishehvar

Unified Keyword Spotting and Audio Tagging on Mobile Devices with Transformers ....................... 721  
Heinrich Dinkel, Yongqing Wang, Zhiyong Yan, Junbo Zhang, Yujun Wang

Anomalous Sound Detection Using Audio Representation with Machine ID Based Contrastive  
Learning Pretraining ....................................................................................... 726  
Jian Guan, Feiyang Xiao, Youde Liu, Qiaoxi Zhu, Wenwu Wang

Cross-Modality Depth Estimation Via Unsupervised Stereo RGB-to-infrared Translation .................... 731  
Shi Tang, Xinchen Ye, Fei Xue, Rui Xu

Solving Jigsaw Puzzle of Large Eroded Gaps Using Puzzlet Discriminant Network ............................ 736  
Xingke Song, Xiaoying Yang, Jianfeng Ren, Ruibin Bai, Xudong Jiang

A New Semi-Supervised Classification Method Using a Supervised Autoencoder for Biomedical  
Applications .................................................................................................. 741  
Cyprien Gille, Frederic Guyard, Michel Barlaud

A Frequency-Domain Recursive Least-Squares Adaptive Filtering Algorithm Based on a Kronecker  
Product Decomposition .................................................................................. 746  
Hongsen He, Jingdong Chen, Jacob Benesty, Yi Yu

Asymptotic Bias and Variance of Kernel Ridge Regression ................................................................ 751  
Victor Solo

Rethinking Implicit Neural Representations for Vision Learners ......................................................... 756  
Yiran Song, Qianyu Zhou, Lizhuang Ma

Streaming Voice Conversion Via Intermediate Bottleneck Features and Non-Streaming Teacher  
Guidance ........................................................................................................... 761  
Yuanzhe Chen, Ming Tu, Tang Li, Xin Li, Qiuqiang Kong, Jiaxin Li, Zhichao Wang, Qiao  
Tian, Yuping Wang, Yuxuan Wang

Learning Sparse auto-Encoders for Green AI Image Coding ................................................................. 766  
Cyprien Gille, Frédéric Guyard, Marc Antonini, Michel Barlaud
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrared and Visible Image Fusion by Using Multi-Scale Transformation</td>
<td>856</td>
</tr>
<tr>
<td>and Fractional-Order Gradient Information</td>
<td></td>
</tr>
<tr>
<td>Shiwei Wu, Kang Zhang, Xia Yuan, Chunxia Zhao</td>
<td></td>
</tr>
<tr>
<td>Optimal Transport with a Diversified Memory Bank for Cross-Domain</td>
<td>861</td>
</tr>
<tr>
<td>Speaker Verification</td>
<td></td>
</tr>
<tr>
<td>Ruiteng Zhang, Jianguo Wei, Xugang Lu, Wenhuan Lu, Di Jin, Lin Zhang,</td>
<td></td>
</tr>
<tr>
<td>Junhai Xu</td>
<td></td>
</tr>
<tr>
<td>BHE-DARTS: Bilevel Optimization Based on Hypergradient Estimation</td>
<td>866</td>
</tr>
<tr>
<td>for Differentiable Architecture Search</td>
<td></td>
</tr>
<tr>
<td>Zicheng Cai, Lei Chen, Hai-Lin Liu</td>
<td></td>
</tr>
<tr>
<td>Not All Classes Are Equal: Adaptively Focus-Aware Confidence for</td>
<td>871</td>
</tr>
<tr>
<td>Semi-Supervised Object Detection</td>
<td></td>
</tr>
<tr>
<td>Hui Zhu, Yongchun Lu, Hongyu Zhao, Guoqing Zhao, Xiaofang Zhao</td>
<td></td>
</tr>
<tr>
<td>Combining Dual-Tree Wavelet Analysis and Proximal Optimization</td>
<td>876</td>
</tr>
<tr>
<td>for Anisotropic Scale-Free Texture Segmentation</td>
<td></td>
</tr>
<tr>
<td>Leo Davy, Nelly Pustelnik, Patrice Abry</td>
<td></td>
</tr>
<tr>
<td>Interaction-Assisted Multi-Modal Representation Learning for</td>
<td>881</td>
</tr>
<tr>
<td>Recommendation</td>
<td></td>
</tr>
<tr>
<td>Hao Wu, Jiajie Wang, Zhonglin Zu</td>
<td></td>
</tr>
<tr>
<td>Transaudio: Towards the Transferable Adversarial Audio Attack</td>
<td>886</td>
</tr>
<tr>
<td>Via Learning Contextualized Perturbations</td>
<td></td>
</tr>
<tr>
<td>Gege Qi, Yuefeng Chen, Yao Zhu, Binyuan Hui, Xiaodan Li, XiaoFeng</td>
<td></td>
</tr>
<tr>
<td>Mao, Rong Zhang, Hui Xue</td>
<td></td>
</tr>
<tr>
<td>Semi-Supervised Remote Sensing Image Change Detection Using Mean</td>
<td>891</td>
</tr>
<tr>
<td>Teacher Model for Constructing Pseudo-Labels</td>
<td></td>
</tr>
<tr>
<td>Zan Mao, Xinyu Tong, Ze Luo</td>
<td></td>
</tr>
<tr>
<td>SAR Image Despeckling with Residual-In-Residual Dense Generative</td>
<td>896</td>
</tr>
<tr>
<td>Adversarial Network</td>
<td></td>
</tr>
<tr>
<td>Yunpeng Bai, Yayuan Xiao, Xuan Hou, Ying Li, Changjing Shang, Qiang</td>
<td></td>
</tr>
<tr>
<td>Shen</td>
<td></td>
</tr>
<tr>
<td>Int-GNN: A User Intention Aware Graph Neural Network for</td>
<td>901</td>
</tr>
<tr>
<td>Session-Based Recommendation</td>
<td></td>
</tr>
<tr>
<td>Guangning Xu, Jinyang Yang, Jinjin Guo, Zhichao Huang, Bowen Zhang</td>
<td></td>
</tr>
<tr>
<td>Co-Design for Mimo Radar and Mimo Communication Aided by</td>
<td>906</td>
</tr>
<tr>
<td>Reconfigurable Intelligent Surface</td>
<td></td>
</tr>
<tr>
<td>Da Li, Bo Tang, Lei Xue</td>
<td></td>
</tr>
<tr>
<td>CLMAE: A Lighter and Faster Masked Autoencoders</td>
<td>911</td>
</tr>
<tr>
<td>Yiran Song, Lizhuang Ma</td>
<td></td>
</tr>
<tr>
<td>GCC-Speaker: Target Speaker Localization with Optimal</td>
<td>916</td>
</tr>
<tr>
<td>Speaker-Dependent Weighting in Multi-Speaker Scenarios</td>
<td></td>
</tr>
<tr>
<td>Guanjun Li, Wei Xue, Wenju Liu, Jiangyan Yi, Jianhua Tao</td>
<td></td>
</tr>
<tr>
<td>MID-Attribute Speaker Generation Using Optimal-Transport-Based</td>
<td>921</td>
</tr>
<tr>
<td>Interpolation of Gaussian Mixture Models</td>
<td></td>
</tr>
<tr>
<td>Aya Watanabe, Shinnosuke Takamichi, Yuki Saito, Detai Xin, Hiroshi</td>
<td></td>
</tr>
<tr>
<td>Saruwatari</td>
<td></td>
</tr>
<tr>
<td>M2-CTTS: End-to-End Multi-Scale Multi-Modal Conversational</td>
<td>926</td>
</tr>
<tr>
<td>Text-to-Speech Synthesis</td>
<td></td>
</tr>
<tr>
<td>Jinlong Xue, Yayue Deng, Fengping Wang, Ya Li, Yingming Gao, Jianhua</td>
<td></td>
</tr>
<tr>
<td>Tao, Jianqing Sun, Jiaen Liang</td>
<td></td>
</tr>
</tbody>
</table>
Nord: Non-Matching Reference Based Relative Depth Estimation from Binaural Speech ........................................... 931
Pranay Manocha, Israel D. Gebru, Anurag Kumar, Dejan Markovic, Alexander Richard

ROI-Based Deep Image Compression with Swin Transformers ................................................................. 936
Binglin Li, Jie Liang, Haisheng Fu, Jingning Han

Interpretability in the Context of Sequential Cost-Sensitive Feature Acquisition ........................................ 941
Yasitha Warahena Liyanage, Daphney–Stavroula Zois

Neural Transducer Training: Reduced Memory Consumption with Sample-Wise Computation ..................... 946
Stefan Braun, Erik McDermott, Roger Hsiao

Multi-Resolution Location-Based Training for Multi-Channel Continuous Speech Separation .......... 951
Hassan Taherian, Deliang Wang

CyPMLI: WISL-Minimized Unimodular Sequence Design Via Power Method-Like Iterations ............. 956
Arian Eamaz, Farhang Yeganegi, Mojtaba Soltanalian

Weight-Sharing Supernet for Searching Specialized Acoustic Event Classification Networks Across Device Constraints ................................................................................................................ 961
Guanting Lin, Qingming Tang, Chieh-Chi Kao, Viktor Rozgic, Chao Wang

Fast Convolution Algorithm for Real-Valued Finite Length Sequences ..................................................... 966
Weimei Wang, Victor Debrunner, Linda S. Debrunner

Do Prosody Transfer Models Transfer Prosody? ...................................................................................... 971
Atli Thor Sigurgeirsson, Simon King

Do Coarser Units Benefit Cluster Prediction-Based Speech Pre-Training? ....................................................... 976
Ali Elkahky, Wei-Ning Hsu, Paden Tomasello, Tu-Anh Nguyen, Robin Algayres, Yossi Adi, Jade Copet, Emmanuel Dupoux, Abdelrahman Mohamed

Learned Video Coding with Motion Compensation Mixture Model .......................................................... 981
Khanh Quoc Dinh, Kwang Pyo Choi

HEiMDaL: Highly Efficient Method for Detection and Localization of Wake-Words .................................. 986
Arnav Kundu, Mohammad Samragh, Minsik Cho, Priyanka Padmanabhan, Devang Naik

Spherical Vector Quantization for Spatial Direction Coding ......................................................................... 991
Stéphane Ragot, Adriana Vasilache

Sequential Datum–Wise Joint Feature Selection and Classification in the Presence of External Classifier ......................................................................................................................................... 996
Sachini Piyoni Ekanyake, Daphney–Stavroula Zois, Charalampos Chelmis

Multi-Scale Compositional Constraints for Representation Learning on Videos ....................................... 1001
Georgios Paraskevopoulos, Chandrashekhar LAVANIA, Lovish Chum, Shiva Sundaram

Active Learning of non-Semantic Speech Tasks with Pretrained Models ................................................ 1006
Harlin Lee, Aaqib Saeed, Andrea L. Bertozzi

A Magnetic Framelet-Based Convolutional Neural Network for Directed Graphs ................................ 1011
Lequan Lin, Junbin Gao

Multi-Output RNN-T Joint Networks for Multi-Task Learning of ASR and Auxiliary Tasks ......................... 1016
Weiran Wang, Ding Zhao, Shaojin Ding, Hao Zhang, Shuo-Yin Chang, David Rybach, Tara N. Sainath, Yanzhang He, Ian McGraw, Shankar Kumar
Visual-Aware Text-to-Speech
Mohan Zhou, Yalong Bai, Wei Zhang, Ting Yao, Tiejun Zhao, Tao Mei

Output-Dependent Gaussian Process State-Space Model
Zhidi Lin, Lei Cheng, Feng Yin, Lexi Xu, Shuguang Cui

A Mathematical Model for Neuronal Activity and Brain Information Processing Capacity
Yu Zheng, David Zhu, Jian Ren, Taosheng Liu, Karl Friston, Tongtong Li

Attention Based Relation Network for Facial Action Units Recognition
Yao Wei, Haoxiang Wang, Mingze Sun, Jiaawang Liu

Learning Hybrid Representations of Semantics and Distortion for Blind Image Quality Assessment
Xiaoxi Wang, Jian Xiong, Bo Li, Jinli Suo, Hao Gao

Affinity Learning with Blind-Spot Self-Supervision for Image Denoising
Yuhongze Zhou, Liguang Zhou, Issam Hadj Laradji, Tin Lun Lam, Yangsheng Xu

A Comprehensive Comparison of Projections in Omnidirectional Super-Resolution
Huicheng Pi, Senmao Tian, Ming Lu, Jianming Liu, Yandong Guo, Shunli Zhang

APGP: Accuracy-Preserving Generative Perturbation for Defending Against Model Cloning Attacks
Anda Cheng, Jian Cheng

HDNet: Hierarchical Dynamic Network for Gait Recognition Using Millimeter-Wave Radar
Yanyan Huang, Yong Wang, Kun Shi, Chaofie Gu, Yu Fu, Cheng Zhuo, Zhiguo Shi

Liveness Score-Based Regression Neural Networks for Face Anti-Spoofing
Youngjun Kwak, Minyoung Jung, Hunjae Yoo, Jinho Shin, Changick Kim

Efficient Online Convolutional Dictionary Learning Using Approximate Sparse Components
Farshad G. Veshki, Sergiy A. Vorobyov

Laryngeal Leukoplakia Classification Via Dense Multiscale Feature Extraction in White Light Endoscopy Images
Zhenzhen You, Yan Yan, Zhenghao Shi, Minghua Zhao, Jing Yan, Haiqin Liu, Xinhong Hei, Xiaoyong Ren

String-Based Molecule Generation Via Multi-Decoder VAE
Kisoo Kwon, Kuhwan Jeong, Junghyun Park, Hwidong Na, Jinwoo Shin

Contrastive Domain Adaptation Via Delimitation Discriminator
Xing Wei, Bin Wen, Lei Chen, Yujie Liu, Chong Zhao, Yang Lu

Towards Robust Audio-Based Vehicle Detection Via Importance-Aware Audio-Visual Learning
Jung Uk Kim, Seong Tae Kim

Modulation-Based Center Alignment and Motion Mining for Spatial Temporal Action Detection
Weiji Zhao, Kefeng Huang, Chongyang Zhang

Randmasking Augment: A Simple and Randomized Data Augmentation for Acoustic Scene Classification
Jubum Han, Mateusz Matuszewski, Olaf Sikorski, Hosang Sung, Hoonyoung Cho
MHSCNET: A Multimodal Hierarchical Shot-Aware Convolutional Network for Video Summarization

Wujiang Xu, Runzhong Wang, Xiaobo Guo, Shaoshuai Li, Qiongxu Ma, Yunan Zhao, Sheng Guo, Zhenfeng Zhu, Junchi Yan

Defending Against Universal Patch Attacks by Restricting Token Attention in Vision Transformers

Hongwei Yu, Jiansheng Chen, Huimin Ma, Cheng Yu, Xinlong Ding

Two-Stage Video De-Raining with Spatio-Temporal Fusion and Illumination-Invariant Detail Preservation

Yufeng Tan, Youjun Xiang, Lei Cai, Pengcheng Wang, Ying Zhang, Yuli Fu

AugTarget Data Augmentation for Infrared Small Target Detection

Shengjia Chen, Jiewen Zhu, Luping Ji, Hongjun Pan, Yuhaow Xu

MMATR: A Lightweight Approach for Multimodal Sentiment Analysis Based on Tensor Methods

Panagiotis Koromilas, Mihalis A. Nicolaou, Theodoros Giannakopoulos, Yannis Panagakis

Real-Time Speech Enhancement with Dynamic Attention Span

Chengyu Zheng, Yuan Zhou, Xiulian Peng, Yuan Zhang, Yan Lu

DQFORMER: Dynamic Query Transformer for Lane Detection

Hao Yang, Shuyuan Lin, Runqing Jiang, Yang Lu, Hanzi Wang

CROSSESPEECH: Speaker-Independent Acoustic Representation for Cross-Lingual Speech Synthesis

Ji-Hoon Kim, Hong-Sun Yang, Yoon-Cheol Ju, Il-Hwan Kim, Byeong-Yeol Kim

Focusing on Targets for Improving Weakly Supervised Visual Grounding

Viet-Quoc Pham, Nao Mishima

Lightweight Portrait Segmentation Via Edge-Optimized Attention

Xinyue Zhang, Guodong Wang, Lijuan Yang, Chenglizhao Chen

Leveraging Sparsity with Spiking Recurrent Neural Networks for Energy-Efficient Keyword Spotting

Manon Dampflhoffer, Thomas Mesquida, Emmanuel Hardy, Alexandre Valentin, Lorena Anghel

TSPTQ-ViT: Two-Scaled Post-Training Quantization for Vision Transformer

Yu-Shan Tai, Ming-Guang Lin, An-Yeu Andy Wu

Explanations for Automatic Speech Recognition

Xiaoliang Wu, Peter Bell, Ajitha Rajan

Contrast-PLC: Contrastive Learning for Packet Loss Concealment

Huaying Xue, Xiulian Peng, Yan Lu

Efficient Large-Scale Audio Tagging Via Transformer-to-CNN Knowledge Distillation

Florian Schmid, Khaled Koutini, Gerhard Widmer

Masking Speech Contents by Random Splicing: is Emotional Expression Preserved?

Felix Burkhardt, Anna Derington, Matthias Kahlau, Klaus Scherer, Florian Eyben, Björn Schuller

On Out-Of-Distribution Detection for Audio with Deep Nearest Neighbors

Zaharah Bukhsh, Aaqib Saeed
C2KD: Cross-Lingual Cross-Modal Knowledge Distillation for Multilingual Text-Video Retrieval

Andrew Rouditchenko, Yang-Sung Chuang, Nina Shvetsova, Samuel Thomas, Rogerio Feris,
Brian Kingsbury, Leonid Karlinsky, David Harwath, Hilde Kuehne, James Glass

Amicable Aid: Perturbing Images to Improve Classification Performance

Juyeop Kim, Jun-Ho Choi, Soobeom Jang, Jong-Seok Lee

Mutual Information Based Reweighting for Precipitation Nowcasting

Yuan Cao, Danchen Zhang, Xin Zheng, Hongming Shan, Junping Zhang

Structured State Space Decoder for Speech Recognition and Synthesis

Koichi Miyazaki, Masato Murata, Tomoki Koriyama

Implicit Vehicle Positioning with Cooperative Lidar Sensing

Luca Barbieri, Bernardo Camajori Tedeschini, Mattia Brambilla, Monica Nicoli

Channel-Driven Decentralized Bayesian Federated Learning for Trustworthy Decision Making in D2D Networks

Luca Barbieri, Osvaldo Simeone, Monica Nicoli

Multilevel FISTA for Image Restoration

Guillaume Lauga, Elisa Riccietti, Nelly Pustelnik, Paulo Gonçalves

TrimTail: Low-Latency Streaming ASR with Simple but Effective Spectrogram-Level Length Penalty

Xingchen Song, Di Wu, Zhiyong Wu, Binbin Zhang, Yuekai Zhang, Zhendong Peng, Wenpeng Li, Fuping Pan, Changbao Zhu

Cross-Device Federated Learning for Mobile Health Diagnostics: A First Study on COVID-19 Detection

Tong Xia, Jing Han, Abhirup Ghosh, Cecilia Mascolo

Masked Spectrogram Prediction for Self-Supervised Audio Pre-Training

Dading Chong, Helin Wang, Peilin Zhou, Qingcheng Zeng

Benchmarking Cross-Domain Face Recognition with Avatars, Caricatures and Sketches

A. Foroughi, C. Rathgeb, M. Ibsen, C. Busch

Unbiased Unsupervised Stimulus Reconstruction for EEG-Based Auditory Attention Decoding

Nicolas Heintz, Simon Geirnaert, Tom Francart, Alexander Bertrand

Projected Hierarchical ALS for Generalized Boolean Matrix Factorization

Rodrigo Cabral Farias, Sebastian Miron

Using Received Power in Microphone Arrays to Estimate Direction of Arrival

Gustav Zetterqvist, Fredrik Gustafsson, Gustaf Hendeby

Energy Regularized RNNS for Solving Non-Stationary Bandit Problems

Michael Rotman, Lior Wolf

Line Segment Matching Based on Intersection-Enhanced Point Correspondences

Zhiyu Liu, Baojiang Zhong

Text is All You Need: Personalizing ASR Models Using Controllable Speech Synthesis

Karren Yang, Ting-Yao Hu, Jen-Hao Rick Chang, Hema Swetha Koppula, Oncel Tuzel
On-The-Fly Text Retrieval for End-To-end ASR Adaptation ................................................................. 1360
Bolaji Yusuf, Aditya Gourav, Ankur Gandhe, Ivan Bulyko

Deep Learning-Based Stereo Camera Multi-Video Synchronization ...................................................... 1365
Nicolas Boizard, Kevin El Haddad, Thierry Ravet, François Cresson, Thierry Dutoit

Provable Computational and Statistical Guarantees for Efficient Learning of Continuous-Action
Graphical Games .............................................................................................................................. 1370
Adarsh Barik, Jean Honorio

Instance-Aware Hierarchical Structured Policy for Prompt Learning in Vision-Language Models ........... 1375
Xun Wu, Guolong Wang, Zhaoyuan Liu, Xuan Dang, Zheng Qin

Analyzing Acoustic Word Embeddings from Pre-Trained Self-Supervised Speech Models ................. 1380
Ramon Sanabria, Hao Tang, Sharon Goldwater

Probabilistic Back-Ends for Online Speaker Recognition and Clustering ............................................ 1385
Alexey Sholokhov, Nikita Kuzmin, Kong Aik Lee, Eng Siong Chng

A Unified Uncertainty-Aware Exploration: Combining Epistemic and Aleatory Uncertainty ............... 1390
Parvin Malekzadeh, Ming Hou, Konstantinos N. Plataniotis

Independent Vector Analysis with Multivariate Gaussian Model: A Scalable Method by Multilinear
Regression .............................................................................................................................................. 1395
Ben Gabrielson, Mingyu Sun, M. A. B. S. Akhonda, Vince D. Calhoun, Tülay Adali

Toward a Multimodal Approach for Disfluency Detection and Categorization ...................................... 1400
Amrit Romana, Kazuhiro Koishida

Towards Diverse and Coherent Augmentation for Time-Series Forecasting ....................................... 1405
Xiyuan Zhang, Ranak Roy Chowdhury, Jingbo Shang, Rajesh Gupta, Dezhi Hong

On Using the UA-Speech and Torgo Databases to Validate Automatic Dysarthric Speech
Classification Approaches ................................................................................................................ 1410
Guilherme Schu, Parvaneh Janbakhshi, Ina Kodrasi

Detection of Real-Time Deepfakes in Video Conferencing with Active Probing and Corneal
Reflection .............................................................................................................................................. 1415
Hui Guo, Xin Wang, Siwei Lyu

Sparse Mixture Once-For-all Adversarial Training for Efficient In-situ Trade-off Between Accuracy
and Robustness of DNNs ................................................................................................................. 1420
Souvik Kundu, Sairam Sundaresan, Sharath Nittur Sridhar, Shunlin Lu, Han Tang, Peter A. Beerel

Context-Aware End-to-end ASR Using Self-Attentive Embedding and Tensor Fusion .......................... 1425
Shuo-Yiin Chang, Chao Zhang, Tara N. Sainath, Bo Li, Trevor Strohman

EH-Enabled Distributed Detection Over Temporally Correlated Markovian MIMO Channels ........... 1430
Ghazaleh Ardehishi, Azadeh Vosoughi

Cross Modality Knowledge Distillation for Robust Pedestrian Detection in Low Light and Adverse
Weather Conditions ........................................................................................................................ 1435
Mazin Hnewa, Alireza Rahimpour, Justin Miller, Devesh Upadhyay, Hayder Radha

Direction-Of-Arrival Estimation Using Gaussian Process Interpolation .............................................. 1440
Ishan D. Khurjekar, Peter Gerstoft, Christoph F. Mecklenbräuker, Zoi-Heleni Michalopoulou
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQuid: Measuring Speech Naturalness in Many Languages</td>
<td>1445</td>
</tr>
<tr>
<td>Thibault Sellam, Ankur Bapna, Joshua Camp, Diana Mackinnon, Ankur P. Parikh, Jason Riesa</td>
<td></td>
</tr>
<tr>
<td>Linear Microphone Array Parallel to the Driving Direction for in-Car Speech Enhancement</td>
<td>1450</td>
</tr>
<tr>
<td>Masanori Tsujikawa, Akihiko Sugiyama, Ken Hanazawa, Yoshinobu Kajikawa</td>
<td></td>
</tr>
<tr>
<td>Facial Texture Perceiver: Towards High-Fidelity Facial Texture Recovery with Input-Level Inductive Biased Perceiver IO</td>
<td>1455</td>
</tr>
<tr>
<td>Seungeun Lee</td>
<td></td>
</tr>
<tr>
<td>Adaptive Noise Canceller Algorithm with SNR-Based Stepsize and Data-Dependent Averaging</td>
<td>1460</td>
</tr>
<tr>
<td>Akihiko Sugiyama</td>
<td></td>
</tr>
<tr>
<td>Exploring Instance Relation for Decentralized Multi-Source Domain Adaptation</td>
<td>1465</td>
</tr>
<tr>
<td>Yikang Wei, Yahong Han</td>
<td></td>
</tr>
<tr>
<td>Contrastive Learning with Dialogue Attributes for Neural Dialogue Generation</td>
<td>1470</td>
</tr>
<tr>
<td>Jie Tan, Hengyi Cai, Hongshen Chen, Hong Cheng, Helen Meng, Zhuoye Ding</td>
<td></td>
</tr>
<tr>
<td>Open-Set Automatic Target Recognition</td>
<td>1475</td>
</tr>
<tr>
<td>Bardia Safaei, Vibashan Vs, Celso M. De Melo, Shuowen Hu, Vishal M. Patel</td>
<td></td>
</tr>
<tr>
<td>Training Robust Spiking Neural Networks on Neuromorphic Data with Spatiotemporal Fragments</td>
<td>1480</td>
</tr>
<tr>
<td>Haibo Shen, Yihao Luo, Xiang Cao, Liangqi Zhang, Juyu Xiao, Tianjiang Wang</td>
<td></td>
</tr>
<tr>
<td>On Bidirectional Preestimates and Their Application to Identification of Fast Time-Varying Systems</td>
<td>1485</td>
</tr>
<tr>
<td>Maciej Niedzwiecki, Artur Gancza, Lu Shen, Yuriy Zakharov</td>
<td></td>
</tr>
<tr>
<td>Matrix Resolvent Eigenembeddings for Dynamic Graphs</td>
<td>1490</td>
</tr>
<tr>
<td>Vasileios Kalantzis, Panagiots A. Traganitis</td>
<td></td>
</tr>
<tr>
<td>Class-Aware Shared Gaussian Process Dynamic Model</td>
<td>1495</td>
</tr>
<tr>
<td>Ryosuke Sawata, Takahiro Ogawa, Miki Haseyama</td>
<td></td>
</tr>
<tr>
<td>Hierarchical Multi-Agent Reinforcement Learning with Intrinsic Reward Rectification</td>
<td>1500</td>
</tr>
<tr>
<td>Zhihao Liu, Zhiwei Xu, Guoliang Fan</td>
<td></td>
</tr>
<tr>
<td>On Multiple-Input/Binaural-Output Antiphase Speaker Signal Extraction</td>
<td>1505</td>
</tr>
<tr>
<td>Xianrui Wang, Ningning Pan, Jacob Benesty, Jingdong Chen</td>
<td></td>
</tr>
<tr>
<td>Speech Dereverberation with a Reverberation Time Shortening Target</td>
<td>1510</td>
</tr>
<tr>
<td>Rui Zhou, Wenye Zhu, Xiaofei Li</td>
<td></td>
</tr>
<tr>
<td>Spatially Informed Independent Vector Analysis for Source Extraction Based on the Convolutive Transfer Function Model</td>
<td>1515</td>
</tr>
<tr>
<td>Xianrui Wang, Andreas Brendel, Gongping Huang, Yichen Yang, Walter Kellermann, Jingdong Chen</td>
<td></td>
</tr>
<tr>
<td>Dewarping Documents Using C2 Continuous Boundary Estimation</td>
<td>1520</td>
</tr>
<tr>
<td>Prasenjit Mondal, Ayush Pant, Sachin Soni</td>
<td></td>
</tr>
<tr>
<td>Dynamic Selection of P-Norm in Linear Adaptive Filtering Via Online Kernel-based Reinforcement Learning</td>
<td>1525</td>
</tr>
<tr>
<td>Minh Vu, Yuki Akiyama, Konstantinos Slavakis</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>LiQuiD-MIMO Radar: Distributed MIMO Radar with Low-Bit Quantization</td>
<td>1530</td>
</tr>
<tr>
<td>Yikun Xiang, Feng Xi, Shengyao Chen</td>
<td></td>
</tr>
<tr>
<td>IAST: Instance Association Relying on Spatio-Temporal Features for</td>
<td>1535</td>
</tr>
<tr>
<td>Video Instance Segmentation</td>
<td></td>
</tr>
<tr>
<td>Junhao Chen, Sheng Liu, Ruixiang Chen, Bingnan Guo, Feng Zhang</td>
<td></td>
</tr>
<tr>
<td>UPGLADE: Unplugged Plug-And-Play Audio Decilper Based on Consensus</td>
<td>1540</td>
</tr>
<tr>
<td>Equilibrium of DNN and Sparse Optimization</td>
<td></td>
</tr>
<tr>
<td>Tomoro Tanaka, Kohei Yatabe, Yasuhiro Oikawa</td>
<td></td>
</tr>
<tr>
<td>Streaming Stroke Classification of Online Handwriting</td>
<td>1545</td>
</tr>
<tr>
<td>Jing-Yu Liu, Yan-Ming Zhang, Fei Yin, Cheng-Lin Liu</td>
<td></td>
</tr>
<tr>
<td>This Changes to That: Combining Causal and Non-Causal Explanations</td>
<td>1550</td>
</tr>
<tr>
<td>to Generate Disease Progression in Capsule Endoscopy</td>
<td></td>
</tr>
<tr>
<td>Anuja Vats, Ahmed Mohammed, Marius Pedersen, Nirmalie Wiratunga</td>
<td></td>
</tr>
<tr>
<td>Masked Modeling Duo: Learning Representations by Encouraging Both</td>
<td>1555</td>
</tr>
<tr>
<td>Networks to Model the Input</td>
<td></td>
</tr>
<tr>
<td>Daisuke Nizumi, Daiki Takeuchi, Yasunori Ohishi, Noboru Harada, Kunio</td>
<td></td>
</tr>
<tr>
<td>Kashino</td>
<td></td>
</tr>
<tr>
<td>CANet: Curved Guide Line Network with Adaptive Decoder for Lane</td>
<td>1560</td>
</tr>
<tr>
<td>Detection</td>
<td></td>
</tr>
<tr>
<td>Zhongyu Yang, Chen Shen, Wei Shao, Tengfei Xing, Runbo Hu, Pengfei</td>
<td></td>
</tr>
<tr>
<td>Xu, Hua Chai, Ruini Xue</td>
<td></td>
</tr>
<tr>
<td>Quantum Variational Bayes on Manifolds</td>
<td>1565</td>
</tr>
<tr>
<td>Anna Lopatnikova, Minh-Ngoc Tran</td>
<td></td>
</tr>
<tr>
<td>Robust Fir Filters for Wireless Low-Frequency Sound Zones</td>
<td>1570</td>
</tr>
<tr>
<td>Mo Zhou, Martin Bo Møller, Christian Sejer Pedersen, Jan Østergaard</td>
<td></td>
</tr>
<tr>
<td>UCONV-Conformer: High Reduction of Input Sequence Length for</td>
<td>1575</td>
</tr>
<tr>
<td>End-to-End Speech Recognition</td>
<td></td>
</tr>
<tr>
<td>Andrei Andrusenko, Rauf Nasretdinov, Aleksei Romanenko</td>
<td></td>
</tr>
<tr>
<td>SFEMGN: Image Denoising with Shallow Feature Enhancement Network and</td>
<td>1580</td>
</tr>
<tr>
<td>Multi-Scale ConvGRU</td>
<td></td>
</tr>
<tr>
<td>Qidong Wang, Lili Guo, Shifei Ding, Jian Zhang, Xiao Xu</td>
<td></td>
</tr>
<tr>
<td>Time-Aware Multiway Adaptive Fusion Network for Temporal Knowledge</td>
<td>1585</td>
</tr>
<tr>
<td>Graph Question</td>
<td></td>
</tr>
<tr>
<td>Yonghao Liu, Di Liang, Fang Fang, Sirui Wang, Wei Wu, Rui Jiang</td>
<td></td>
</tr>
<tr>
<td>FAPM: Fast Adaptive Patch Memory for Real-Time Industrial Anomaly</td>
<td>1590</td>
</tr>
<tr>
<td>Detection</td>
<td></td>
</tr>
<tr>
<td>Donghyeong Kim, Chaewon Park, Suhwan Cho, Sangyoun Lee</td>
<td></td>
</tr>
<tr>
<td>CRFAST: Clip-Based Reference-Guided Facial Image Semantic Transfer</td>
<td>1595</td>
</tr>
<tr>
<td>Ailin Li, Lei Zhao, Zhiwen Zuo, Zhizhong Wang, Wei Xing, Dongming Lu</td>
<td></td>
</tr>
<tr>
<td>Introducing Topography in Convolutional Neural Networks</td>
<td>1600</td>
</tr>
<tr>
<td>Maxime Poli, Emmanuel Dupoux, Rachid Riad</td>
<td></td>
</tr>
<tr>
<td>Structure-Aware Sparse Bayesian Learning-Based Channel Estimation</td>
<td>1605</td>
</tr>
<tr>
<td>for Intelligent Reflecting Surface-Aided MIMO</td>
<td></td>
</tr>
<tr>
<td>Yanbin He, Geethu Joseph</td>
<td></td>
</tr>
</tbody>
</table>

Hans Van Gorp, Merel M. Van Gilst, Pedro Fonseca, Sebastiaan Overeem, Ruud J. G. Van Sloun

Hardware Friendly Spline Sketched Lidar ................................................................. 1615

Michael P. Sheehan, Julián Tachella, Mike E. Davies

Explicit Ziv-Zakai Bound for Multiple Sources Doa Estimation ...................................... 1620

Zongyu Zhang, Yujie Gu, Zhiguo Shi

Adaptive Large Margin Fine-Tuning for Robust Speaker Verification ...................................... 1625

Leying Zhang, Zhengyang Chen, Yanmin Qian

Radio-Astronomy Imaging and Interference Excision Using Tensor Decomposition and Canonical Correlation Analysis ................................................................. 1630

Mikael Sørensen, Nicholas D. Sidiropoulos

Hybridformer: Improving Squeezeformer with Hybrid Attention and NSR Mechanism ...................... 1635

Yuguang Yang, Yu Pan, Jingjing Yin, Jiangyu Han, Lei Ma, Heng Lu

EGAN: A Neural Excitation Generation Model Based on Generative Adversarial Networks with Harmonics and Noise Input ......................................................... 1640

Yen-Ting Lin, Chen-Yu Chiang

Training Set Cleansing of Backdoor Poisoning by Self-Supervised Representation Learning ..................... 1645

Hang Wang, Sahar Karami, Ousmane Dia, Hippolyt Ritter, Ehsan Emamjomeh-Zadeh, Jiahui Chen, Zhen Xiang, David J. Miller, George Kesidis

HIFI++: A Unified Framework for Bandwidth Extension and Speech Enhancement ...................... 1650

Pavel Andreev, Aibek Alanov, Oleg Ivanov, Dmitry Vetrov

SUVR: A Search-Based Approach to Unsupervised Visual Representation Learning ...................... 1655

Yi-Zhan Xu, Chih-Yao Chen, Cheng-Te Li

Continuous Interaction with a Smart Speaker Via Low-Dimensional Embeddings of Dynamic Hand Pose ................................................................. 1660

Songpei Xu, Chaitanya Kaul, Xuri Ge, Roderick Murray-Smith

Making Synchrosqueezing Locally Adaptive in the Time-Frequency Plane ..................................... 1665

Marcelo A. Colominas, Sylvain Meignen

Refined Pseudo Labeling for Source-Free Domain Adaptive Object Detection ............................. 1670

Siqi Zhang, Lu Zhang, Zhiyong Liu

Learning Quantum Entanglement Distillation with Noisy Classical Communications ........................ 1675

Hari Hara Sutheran Chittoor, Osvaldo Simeone

Hybrid Ris-Assisted Interference Mitigation for Spectrum Sharing ............................................ 1680

Fangzhou Wang, A. Lee Swindlehurst

Disentangled Training with Adversarial Examples for Robust Small-Footprint Keyword Spotting .................. 1685

Zhenyu Wang, Li Wan, Biqiao Zhang, Yiteng Huang, Shang-Wen Li, Ming Sun, Xin Lei, Zhaojun Yang

Play it Back: Iterative Attention for Audio Recognition ..................................................... 1690

Alexandros Stergiou, Dima Damen
Anchored Speech Recognition with Neural Transducers ........................................... 1695
Desh Raj, Junteng Jia, Jay Mahadeokar, Chunyang Wu, Niko Moritz, Xiaohui Zhang, Ozlem Kalinli

Can Knowledge of End-To-End Text-To-Speech Models Improve Neural Midi-To-Audio Synthesis Systems? ......................................................... 1700
Xuan Shi, Erica Cooper, Xin Wang, Junichi Yamagishi, Shrikanth Narayanan

Perspective Projection-Based 3D CT Reconstruction from Biplanar X-Rays .................. 1705
Daeun Kyung, Kyungmin Jo, Jaegul Choo, Joonseok Lee, Edward Choi

Shadow Removal of Text Document Images Using Background Estimation and Adaptive Text Enhancement ......................................................... 1710
Wenjie Liu, Bingshu Wang, Jiangbin Zheng, Wenmin Wang

A Quantum Kernel Learning Approach to Acoustic Modeling for Spoken Command Recognition .......... 1715
Chao-Han Huck Yang, Bo Li, Yu Zhang, Nanxin Chen, Tara N. Sainath, Sabato Marco Siniscalchi, Chin-Hui Lee

Speech Emotion Recognition Via Two-Stream Pooling Attention with Discriminative Channel Weighting ......................................................... 1720
Ke Liu, Dekui Wang, Dongya Wu, Jun Feng

Identifying Entrainment in Task-Oriented Conversations ........................................... 1725
Run Chen, Seokhwan Kim, Alexandros Papangelis, Julia Hirschberg, Yang Liu, Dilek Hakkani-Tür

Spatial Active Noise Control Method Based on Sound Field Interpolation from Reference Microphone Signals ......................................................... 1730
Kazuyuki Arikawa, Shoichi Koyama, Hiroshi Saruwatari

A Perturbation-Based Policy Distillation Framework with Generative Adversarial Nets .......... 1735
Lihua Zhang, Quan Liu, Xiongzheng Zhang, Yapeng Xu

Domain Generalized Fundus Image Segmentation Via Dual-Level Mixing .................... 1740
Xin Luo, Wei Chen, Chen Li, Bin Zhou, Yusong Tan

A Holistic Cascade System, Benchmark, and Human Evaluation Protocol for Expressive Speech-To-Speech Translation ........................................... 1745
Wen-Chin Huang, Benjamin Peloquin, Justine Kao, Changhan Wang, Hongyu Gong, Elizabeth Salesky, Yossi Adi, Ann Lee, Peng-Jen Chen

Class-Incremental Learning on Multivariate Time Series Via Shape-Aligned Temporal Distillation .......... 1750
Zhongzheng Qiao, Minghui Hu, Xudong Jiang, Ponnuthurai Nagaratnam Suganthan, Ramasamy Savitha

Distributed Online Learning with Adversarial Participants in an Adversarial Environment .......... 1755
Xingrong Dong, Zhaoxian Wu, Qing Ling, Zhi Tian

A Geometric Surrogate for Simulation Calibration ..................................................... 1760
Lincon S. Souza, Bojan Batalo, Keisuke Yamazaki

Role of Bias Terms in Dot-Product Attention ......................................................... 1765
Mahdi Namazifar, Devamanyu Hazarika, Dilek Hakkani-Tür

Jazznet: A Dataset of Fundamental Piano Patterns for Music Audio Machine Learning Research .......... 1770
Tosiron Adegbija
Multi-Functional Reconfigurable Intelligent Surface ......................................................... 1775
Wen Wang, Wanli Ni, Hui Tian, Yonina C. Eldar

Heuristic Masking for Text Representation Pretraining .................................................... 1780
Yimeng Zhuang

Supervised Hierarchical Clustering Using Graph Neural Networks for Speaker Diarization ........... 1785
Prachi Singh, Amrit Kaul, Sriram Ganapathy

Kernel Interpolation of Acoustic Transfer Functions with Adaptive Kernel for Directed and
Residual Reverberations ........................................................................................................... 1790
Juliano G. C. Ribeiro, Shoichi Koyama, Hiroshi Saruwatari

SLBERT: A Novel Pre-Training Framework for Joint Speech and Language Modeling .................. 1795
Onkar Susladkar, Prajwal Gatti, Santosh Kumar Yadav

CANDY: Category-Kernelized Dynamic Convolution for Instance Segmentation ......................... 1800
Yao Lu, Zhiyi Chen, Zehui Chen, Jie Hu, Liujuan Cao, Shengchuan Zhang

A Study on Bias and Fairness in Deep Speaker Recognition .................................................. 1805
Amirhossein Hajavi, Ali Etemad

Thermal Infrared Image Inpainting Via Edge-Aware Guidance .............................................. 1810
Zeyu Wang, Haibin Shen, Changyou Men, Quan Sun, Kejie Huang

On Word Error Rate Definitions and Their Efficient Computation for Multi-Speaker Speech
Recognition Systems ................................................................................................................. 1815
Thilo Von Neumann, Christoph Boeddeker, Keisuke Kinoshita, Marc Delcroix, Reinhold
Haeb-Umbach

Deep Low Light Image Enhancement Via Multi-Scale Recursive Feature Enhancement and Curve
Adjustment ...................................................................................................................................... 1820
Haiyan Jin, Dawei Wei, Haonan Su

Log-Can: Local-Global Class-Aware Network for Semantic Segmentation of Remote Sensing
Images ........................................................................................................................................... 1825
Xiaowen Ma, Mengting Ma, Chenlu Hu, Zhiyuan Song, Ziyao Zhao, Tian Feng, Wei Zhang

TAMformer: Multi-Modal Transformer with Learned Attention Mask for Early Intent Prediction .......... 1830
Nada Osman, Guglielmo Camporese, Lamberto Ballan

Long-Tailed Recognition with Causal Invariant Transformation .............................................. 1835
Yahong Zhang, Sheng Shi, Chenchen Fan, Yixin Wang, Wenli Ouyang, Weifan, Jianpin Fan

“Prediction of Sleepiness Ratings from Voice by Man and Machine”: A Perceptual Experiment
Replication Study ......................................................................................................................... 1840
Vincent P. Martin, Aymeric Ferron, Jean-Luc Rouas, Pierre Philip

Online Residual-Based Key Frame Sampling with Self-Coach Mechanism and Adaptive Multi-
Level Feature Fusion ............................................................................................................... 1845
Rui Zhang, Yang Hua, Tao Song, Zhengui Xue, Ruhui Ma, Haibing Guan

Unsupervised Word Segmentation Using Temporal Gradient Pseudo-Labels ................................. 1850
Tzeviya Sylvia Fuchs, Yedid Hoshen
Exploiting PRNU and Linear Patterns in Forensic Camera Attribution Under Complex Lens Distortion Correction .......................................................... Andrea Montibeller, Fernando Pérez-González

Gradient Remedy for Multi-Task Learning in End-to-End Noise-Robust Speech Recognition ............... 1860 Yuchen Hu, Chen Chen, Ruizhe Li, Qiushi Zhu, Eng Siong Chng

Region-Aware Transformer with Asymmetric Loss in Multi-Label Classification .............................. 1865 Lei Zhang, Jie Liu, Yanqi Bao, Jie Wang

A Multi-Stage Triple-Path Method for Speech Separation in Noisy and Reverberant Environments .......... 1870 Zhaoxi Mu, Xinyu Yang, Xiangyuan Yang, Wenjing Zhu

Scaling Law Analysis for Covariance Based Activity Detection in Cooperative Multi-Cell Massive Mimo .......................................................... 1875 Ziyue Wang, Ya-Feng Liu, Zhaorui Wang, Wei Yu

Blind Polynomial Regression ...................................................................... 1880 Alberto Natali, Geert Leus

Empathetic Response Generation Via Emotion Cause Transition Graph ............................................ 1885 Yushan Qian, Bo Wang, Ting-En Lin, Yinhe Zheng, Ying Zhu, Dongming Zhao, Yuexian Hou, Yuchuan Wu, Yongbin Li

Semantic-Aware Gated Fusion Network for Interactive Colorization .................................................. 1890 Jie Zhang, Yi Xiao, Yan Zheng, Zhenni Wang, Chi-Sing Leung

CDHD: Contrastive Dreamer for Hint Distillation ............................................................................. 1895 Le Yu, Tongyan Hua, Wenming Yang, Peng Ye, Qingmin Liao

Masked Token Similarity Transfer for Compressing Transformer-Based ASR Models ...................... 1900 Euntae Choi, Youshin Lim, Byeong-Yeol Kim, Hyung Yong Kim, Hanbin Lee, Yunkyu Lim, Seung Woo Yu, Sungjoo Yoo

Which Country is This Picture From? New Data and Methods for Dnn-Based Country Recognition ...... 1905 Omran Alamayreh, Giovanna Maria Dimitri, Jun Wang, Benedetta Tondi, Mauro Barni

Improving Speech Enhancement Via Event-Based Query .............................................................. 1910 Yifei Xin, Xiudian Peng, Yan Lu

Capturing Cross-Scale Disparity for Stereo Image Super-Resolution .................................................. 1915 Kun He, Changyu Li, Dongyang Zhang, Jie Shao

Byzantine-Robust and Communication-Efficient Personalized Federated Learning ......................... 1920 Xuechao He, Jiaojiao Zhang, Qing Ling

Better Together: Dialogue Separation and Voice Activity Detection for Audio Personalization in TV .... 1925 Matteo Torcoli, Émanuël A. P. Habets

Decaying Contrast for Fine-Grained Video Representation Learning .................................................. 1930 Heng Zhang, Bing Su


The MBSTOI Binaural Intelligibility Metric Using a Close-Talking Microphone Reference ................. 1940 Pierre Guiraud, Alastair H. Moore, Rebecca R. Vos, Patrick A. Naylor, Mike Brookes
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>D²Q-DETR: Decoupling and Dynamic Queries for Oriented Object Detection with Transformers</td>
<td>1945</td>
</tr>
<tr>
<td>Qiang Zhou, Chaohui Yu, Zhibin Wang, Fan Wang</td>
<td></td>
</tr>
<tr>
<td>Cross-Head Supervision for Crowd Counting with Noisy Annotations</td>
<td>1950</td>
</tr>
<tr>
<td>Mingliang Dai, Zhizhong Huang, Jiaqi Gao, Hongming Shan, Junping Zhang</td>
<td></td>
</tr>
<tr>
<td>Digital Phenotype Representation by Statistical, Information Theory, Data-Driven Approach with Digital Health Data</td>
<td>1955</td>
</tr>
<tr>
<td>Binh Nguyen, Michael Nigro, Alice Rueda, Venkat Bhat, Sri Krishnan</td>
<td></td>
</tr>
<tr>
<td>Robust Data2VEC: Noise-Robust Speech Representation Learning for ASR by Combining Regression and Improved Contrastive Learning</td>
<td>1960</td>
</tr>
<tr>
<td>Qiu-Shi Zhu, Long Zhou, Jie Zhang, Shu-Jie Liu, Yu-Chen Hu, Li-Rong Dai</td>
<td></td>
</tr>
<tr>
<td>N2MVSNet: Non-Local Neighbors Aware Multi-View Stereo Network</td>
<td>1965</td>
</tr>
<tr>
<td>Zhe Zhang, Huachen Gao, Yuxi Hu, Ronggang Wang</td>
<td></td>
</tr>
<tr>
<td>Batch Normalization Damages Federated Learning on NON-IID Data: Analysis and Remedy</td>
<td>1970</td>
</tr>
<tr>
<td>Yanneng Wang, Qingjiang Shi, Tsung-Hui Chang</td>
<td></td>
</tr>
<tr>
<td>Pitch Mark Detection from Noisy Speech Waveform Using Wave-U-Net</td>
<td>1975</td>
</tr>
<tr>
<td>Hyun-Joon Nam, Hong-June Park</td>
<td></td>
</tr>
<tr>
<td>Zhepei Wang, Ritwik Giri, Devansh Shah, Jean-Marc Valin, Michael M. Goodwin, Paris Smaragdis</td>
<td></td>
</tr>
<tr>
<td>Look and Think: Intrinsic Unification of Self-Attention and Convolution for Spatial-Channel Specificity</td>
<td>1985</td>
</tr>
<tr>
<td>Xiang Gao, Honghui Lin, Yu Li, Ruiyan Fang, Xin Zhang</td>
<td></td>
</tr>
<tr>
<td>Jun Shi, Bingcai Wei, Gang Zhou, Liye Zhang</td>
<td></td>
</tr>
<tr>
<td>Filterbank Learning for Noise-Robust Small-Footprint Keyword Spotting</td>
<td>1995</td>
</tr>
<tr>
<td>Iván López-Espejo, Ram C. M. C. Shekar, Zheng-Hua Tan, Jesper Jensen, John H. L. Hansen</td>
<td></td>
</tr>
<tr>
<td>Learning ASR Pathways: A Sparse Multilingual ASR Model</td>
<td>2000</td>
</tr>
<tr>
<td>Mu Yang, Andros Tjandra, Chunxi Liu, David Zhang, Duc Le, Ozlem Kalinli</td>
<td></td>
</tr>
<tr>
<td>HIPI: A Hierarchical Performer Identification Model Based on Symbolic Representation of Music</td>
<td>2005</td>
</tr>
<tr>
<td>Syed Rifat Mahmud Rafee, György Fazekas, Geraint Wiggins</td>
<td></td>
</tr>
<tr>
<td>Neural Speech Phase Prediction Based on Parallel Estimation Architecture and Anti-Wrapping Losses</td>
<td>2010</td>
</tr>
<tr>
<td>Yang Ai, Zhen-Hua Ling</td>
<td></td>
</tr>
<tr>
<td>Adversarial Permutation Invariant Training for Universal Sound Separation</td>
<td>2015</td>
</tr>
<tr>
<td>Emilian Postolache, Jordi Pons, Santiago Pascual, Joan Serrà</td>
<td></td>
</tr>
<tr>
<td>Super-Resolution for Macro X-Ray Fluorescence Data Collected from Old Master Paintings</td>
<td>2020</td>
</tr>
<tr>
<td>Su Yarn, Herman Verinaz-Jadan, Jun-Jie Huang, Nathan Daly, Catherine Higgitt, Pier Luigi Dragotti</td>
<td></td>
</tr>
<tr>
<td>Audio-Text Models Do Not Yet Leverage Natural Language</td>
<td>2025</td>
</tr>
<tr>
<td>Ho-Hsiang Wu, Oriol Nieto, Juan Pablo Bello, Justin Salamon</td>
<td></td>
</tr>
</tbody>
</table>
ImagineNet: Target Speaker Extraction with Intermittent Visual Cue Through Embedding Inpainting .................................................................................................................. 2030

Zexu Pan, Wupeng Wang, Marvin Borsdorf, Haizhou Li

Audio-To-Intent Using Acoustic-Textual Subword Representations from End-To-End ASR .............................................................................................................. 2035

Pranay Dighe, Prateeth Nayak, Oggi Rudovic, Erik Marchi, Xiaochuan Niu, Ahmed Tewfik

DATA2VEC-SG: Improving Self-Supervised Learning Representations for Speech Generation Tasks ........................................................................................................... 2040

Heming Wang, Yao Qian, Hemin Yang, Naoyuki Kanda, Peidong Wang, Takuya Yoshioka, Xiaofei Wang, Yiming Wang, Shujie Liu, Zhuo Chen, Deliang Wang, Michael Zeng

On the Importance of Different Cough Phases for COVID-19 Detection .......................................................................................................................... 2045

Yi Zhu, Mahil Hussain Shaik, Tiago H. Falk

Privacy-Preserving Occupancy Estimation .......................................................................................................................... 2050

Jennifer Williams, Vahid Yazdanpanah, Sebastian Stein

Towards Simultaneous Segmentation of Liver Tumors and Intrahepatic Vessels Via Cross-Attention Mechanism .................................................................................................................. 2055

Haopeng Kuang, Dingkang Yang, Shunli Wang, Xiaoying Wang, Lihua Zhang

Robust Data-Driven Accelerated Mirror Descent .................................................................................................................................................. 2060

Hong Ye Tan, Subhadip Mukherjee, Junqi Tang, Andreas Hauptmann, Carola-Bibiane Schöntieb

Naturalistic Head Motion Generation from Speech .......................................................................................................................... 2065

Trisha Mittal, Zakaria Aldeneh, Masha Fedzechkina, Amrak Ranjan, Barry-John Theobald

On the Role of LIP Articulation in Visual Speech Perception .......................................................................................................................... 2070

Zakaria Aldeneh, Masha Fedzechkina, Skyler Seto, Katherine Metcalf, Miguel Sarabia, Nicholas Apostoloff, Barry-John Theobald

Accelerating Matrix Trace Estimation by Aitken’s Δ2 Process .......................................................................................................................... 2075

Vassilis Kalantzis, Georgios Kollias, Shashanka Ubaru, Theodoros Salonidis

Long-Short Attention Network for the Spectral Super-Resolution of Multispectral Images ....................................................................................................... 2080

Kai Zhang, Tian Jin, Feng Zhang, Jiande Sun

MADI: Inter-Domain Matching and Intra-Domain Discrimination for Cross-Domain Speech Recognition .................................................................................................................. 2085

Jiaming Zhou, Shiwan Zhao, Ning Jiang, Guoqing Zhao, Yong Qin

Weight-Based Mask for Domain Adaptation .......................................................................................................................... 2090

Eunseop Lee, Inhan Kim, Daijin Kim

Analysing Discrete Self Supervised Speech Representation for Spoken Language Modeling ......................................................................................................... 2095

Amitay Sicherman, Yossi Adi

Robust Multi-Object Tracking with Spatial Uncertainty .......................................................................................................................... 2100

Pin-Jie Liao, Yu-Cheng Huang, Chen-Kuo Chiang, Shang-Hong Lai

SADE: A Self-Adaptive Expert for Multi-Dataset Question Answering .................................................................................................................. 2105

Yixing Peng, Quan Wang, Zhendong Mao, Yongdong Zhang
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-Time Modelling of Observation Filter in the Remote Microphone Technique for an Active Noise Control Application</td>
<td>2110</td>
</tr>
<tr>
<td>Chung Kwan Lai, Bhan Lam, Dongyuan Shi, Woon-Seng Gan</td>
<td></td>
</tr>
<tr>
<td>To Wake-Up Or not To Wake-Up: Reducing Keyword False Alarm by Successive Refinement</td>
<td>2115</td>
</tr>
<tr>
<td>Yashas Malur Saidutta, Rakshith Sharma Srinivasa, Ching-Hua Lee, Chouchang Yang, Yilin Shen, Hongxia Jin</td>
<td></td>
</tr>
<tr>
<td>Continual Learning for On-Device Speech Recognition Using Disentangled Conformers</td>
<td>2120</td>
</tr>
<tr>
<td>Anuj Diwan, Ching-Feng Yeh, Wei-Ning Hsu, Paden Tomasello, Eunsol Choi, David Harwath, Abdelrahman Mohamed</td>
<td></td>
</tr>
<tr>
<td>Shuffled Autoregression for Motion Interpolation</td>
<td>2125</td>
</tr>
<tr>
<td>Shuo Huang, Jia Jia, Zongxin Yang, Wei Wang, Haozhe Wu, Yi Yang, Junliang Xing</td>
<td></td>
</tr>
<tr>
<td>Visual Answer Localization with Cross-Modal Mutual Knowledge Transfer</td>
<td>2130</td>
</tr>
<tr>
<td>Yixuan Weng, Bin Li</td>
<td></td>
</tr>
<tr>
<td>Learning to Build Reasoning Chains by Reliable Path Retrieval</td>
<td>2135</td>
</tr>
<tr>
<td>Minjun Zhu, Yixuan Weng, Shizhu He, Cunguang Wang, Kang Liu, Li Cai, Jun Zhao</td>
<td></td>
</tr>
<tr>
<td>Improving Bert Fine-Tuning Via Stabilizing Cross-Layer Mutual Information</td>
<td>2140</td>
</tr>
<tr>
<td>Jicun Li, Xingjian Li, Tianyang Wang, Shi Wang, Yanan Cao, Chengzhong Xu, Dejing Dou</td>
<td></td>
</tr>
<tr>
<td>TRUSTERA: A Live Conversation Redaction System</td>
<td>2145</td>
</tr>
<tr>
<td>Overcoming Posterior Collapse in Variational Autoencoders Via EM-Type Training</td>
<td>2150</td>
</tr>
<tr>
<td>Ying Li, Lei Cheng, Feng Yin, Michael Minyi Zhang, Sergios Theodoridis</td>
<td></td>
</tr>
<tr>
<td>Multicast Beamformer Design for Mimo Coded Caching Systems</td>
<td>2155</td>
</tr>
<tr>
<td>Mohammadjavad Salehi, Mohammad Naserihrani, Antti Tölli</td>
<td></td>
</tr>
<tr>
<td>Embrace Smaller Attention: Efficient Cross-Modal Matching with Dual Gated Attention Fusion</td>
<td>2160</td>
</tr>
<tr>
<td>Weikuo Guo, Xiangwei Kong</td>
<td></td>
</tr>
<tr>
<td>Online Vector Autoregressive Models Over Expanding Graphs</td>
<td>2165</td>
</tr>
<tr>
<td>Bishwadeep Das, Elvin Isufi</td>
<td></td>
</tr>
<tr>
<td>Estimation of Time-Varying Graph Topologies from Graph Signals</td>
<td>2170</td>
</tr>
<tr>
<td>Yuhao Liu, Chen Cui, Marzieh Ajirak, Petar M. Djuric</td>
<td></td>
</tr>
<tr>
<td>Exploiting One-Class Classification Optimization Objectives for Increasing Adversarial Robustness</td>
<td>2175</td>
</tr>
<tr>
<td>Vasileios Mygdalis, Ioannis Pitas</td>
<td></td>
</tr>
<tr>
<td>Meta-Learning for Image-Guided Millimeter-Wave Beam Selection in Unseen Environments</td>
<td>2180</td>
</tr>
<tr>
<td>Jerry Gu, Liam Collins, Debashri Roy, Aryan Mokhtari, Sanjay Shakkottai, Kaushik R. Chowdhury</td>
<td></td>
</tr>
<tr>
<td>Towards a Robust and Efficient Classifier for Real World Radio Signal Modulation Classification</td>
<td>2185</td>
</tr>
<tr>
<td>Dancheng Liu, Kazim Ergun, Tajana Simunic Rosing</td>
<td></td>
</tr>
<tr>
<td>Cleanformer: A Multichannel Array Configuration-Invariant Neural Enhancement Frontend for ASR in Smart Speakers</td>
<td>2190</td>
</tr>
<tr>
<td>Joseph Caroselli, Arun Narayanan, Nathan Howard, Tom O’Malley</td>
<td></td>
</tr>
</tbody>
</table>
Hybrid Transformers for Music Source Separation

Simon Rouard, Francisco Massa, Alexandre Défossez

Audio-Driven Facial Landmark Generation in Violin Performance Using 3DCNN Network with Self Attention Model

Ting-Wei Lin, Chao-Lin Liu, Li Su

Gesper: A Unified Framework for General Speech Restoration

Jun Chen, Yupeng Shi, Wenzhe Liu, Wei Rao, Shulin He, Andong Li, Yannan Wang, Zhiyong Wu, Shidong Shang, Chengshi Zheng

Flexible Beam Design for Vital Sign Monitoring Using a Phased Array Equipped with Double-Phase Shifters

Zhaoyi Xu, Donglin Gao, Shuping Li, Chung-Tse Michael Wu, Athina Petropulu

Graph Neural Networks for Sound Source Localization on Distributed Microphone Networks

Eric Grinstein, Mike Brookes, Patrick A. Naylor

FedSD: A New Federated Learning Structure Used in Non-Iid Data

Minmin Yi, Houchun Ning, Peng Liu

Controllable Music Inpainting with Mixed-Level and Disentangled Representation

Shiqi Wei, Ziyu Wang, Weiguo Gao, Gus Xia

Reliable Beamforming at Terahertz Bands: Are Causal Representations the Way Forward?

Christo Kurissummottil Thomas, Walid Saad

Light-Weight CNN-Attention Based Architecture for Hand Gesture Recognition Via Electromyography

Soheil Zabihi, Ehsan Rahimian, Amir Asif, Arash Mohammadi

Conformer-Based Target-Speaker Automatic Speech Recognition for Single-Channel Audio

Yang Zhang, Krishna C. Puvvada, Vitaly Lavrukhin, Boris Ginsburg

Effectiveness of Mining Audio and Text Pairs from Public Data for Improving ASR Systems for Low-Resource Languages

Kaushal Bhogale, Abhigyan Raman, Tahir Javed, Sumanth Dodapaneni, Anoop Kunchukuttan, Pratyush Kumar, Mitesh M. Khapra

Mixed Far-Field and Near-Field Source Localization Based on Low-Rank Matrix Reconstruction

Yunchang Liu, Hong Jiang, Qi Zhang

Cancelling Intermodulation Distortions for Otoacoustic Emission Measurements with Earbuds

Berken Utku Demirel, Khaldoon Al-Naimi, Fahim Kawser, Alessandro Montanari

Adversarial Guitar Amplifier Modelling with Unpaired Data

Alec Wright, Vesa Välimäki, Lauri Juvela

Deep Adaptive Superpixels for Hadamard Single Pixel Imaging in Near-Infrared Spectrum

Brayan Monroy, Jorge Bacca, Henry Arguello

Speechlmnscore: Evaluating Speech Generation Using Speech Language Model

Soumi Maiti, Yifan Peng, Takaaki Saeki, Shinji Watanabe

Adaptive Filtering Algorithms for Set-Valued Observations-Symmetric Measurement Approach to Unlabeled and Anonymized Data

Vikram Krishnamurthy
Training Large-Vocabulary Neural Language Models by Private Federated Learning for Resource-Constrained Devices ................................. 2277
Mingbin Xu, Congzheng Song, Ye Tian, Neha Agrawal, Filip Granqvist, Rogier Van Dalen, Xiao Zhang, Arturo Argueta, Shiyi Han, Yaqiao Deng, Leo Liu, Anmol Walia, Alex Jin

Bytecover3: Accurate Cover Song Identification on Short Queries ................................................................. 2282
Xingjian Du, Zijie Wang, Xia Liang, Huidong Liang, Bilei Zhu, Zejun Ma

A Multi-Modal Approach for Context-Aware Network Traffic Classification ...................................................... 2287
Bo Pang, Yongquan Fu, Siyuan Ren, Siqi Shen, Ye Wang, Qing Liao, Yan Jia

Wav2Seq: Pre-Training Speech-to-Text Encoder-Decoder Models Using Pseudo Languages ............................... 2292
Felix Wu, Kwangyoun Kim, Shinji Watanabe, Kyu J. Han, Ryan McDonald, Kilian Q. Weinberger, Yoav Artzi

Improved Mask-Based Neural Beamforming for Multichannel Speech Enhancement by Snapshot Matching Masking ..................................................... 2297
Ching-Hua Lee, Chouchang Yang, Yilin Shen, Hongxia Jin

Get: Gated Contextual Transformer for Sequential Audio Tagging ........................................................................ 2302
Yuanbo Hou, Yun Wang, Wenwu Wang, Dick Botteldooren

Diffusion Motion: Generate Text-Guided 3D Human Motion by Diffusion Model .................................................. 2307
Zhiyuan Ren, Zhihong Pan, Xin Zhou, Le Kang

Real-Time Wireless ECG-Derived Respiration Rate Estimation Using an Autoencoder with a DCT Layer ................................................................. 2312
Hongyi Pan, Xin Zhu, Zhilu Ye, Pai-Yen Chen, Ahmet Enis Cetin

Transceiver Design for MIMO-DFRC Systems ....................................................................................................... 2317
Cai Wen, Timothy N. Davidson

Explicit and Implicit Knowledge Distillation Via Unlabeled Data ........................................................................... 2322
Yuzheng Wang, Zuhao Ge, Zhaoyu Chen, Xian Liu, Chuangjia Ma, Yunquan Sun, Lizhe Qi

Robust Audio-Visual ASR with Unified Cross-Modal Attention .............................................................................. 2327
Jiahong Li, Chenda Li, Yifei Wu, Yanmin Qian

Similarity Relation Preserving Cross-Modal Learning for Multispectral Pedestrian Detection Against Adversarial Attacks ......................................................... 2332
Jung Uk Kim, Yong Man Ro

DailyTalk: Spoken Dialogue Dataset for Conversational Text-To-Speech .............................................................. 2337
Keon Lee, Kyumin Park, Daeyoung Kim

Resolving Doppler Ambiguity Via Spread Phase Alignment in FDA-MIMO Radar .............................................. 2342
Yanxing Wang, Shengqi Zhu, Guisheng Liao, Lan Lan, Zhuochen Chen, Feilong Liu

Adversarial Contrastive Distillation with Adaptive Denoising ............................................................................. 2347
Yuzheng Wang, Zhaoyu Chen, Dingkang Yang, Yang Liu, Siao Liu, Wenqiang Zhang, Lizhe Qi

Interpretable Multi-Scale Neural Network for Granger Causality Discovery ...................................................... 2352
Chenchen Fan, Yixin Wang, Yahong Zhang, Wenli Ouyang

Reliable Cluster-Based Framework for Open Set Domain Adaptation ............................................................... 2357
Xiu Zheng, Yuan Huang, Jie Tang
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust Your Partner’s Friends: Hierarchical Cross-Modal Contrastive Pre-Training for Video-Text</td>
<td>2362</td>
</tr>
<tr>
<td>Yuhan Xiang, Kaijian Liu, Shixiang Tang, Lei Bai, Feng Zhu, Rui Zhao, Xianming Lin</td>
<td></td>
</tr>
<tr>
<td>Relational Representation Learning for Zero-Shot Relation Extraction with Instance Prompting and Prototype Rectification</td>
<td>2367</td>
</tr>
<tr>
<td>Bin Duan, Xingxian Liu, Shusen Wang, Yajing Xu, Bo Xiao</td>
<td></td>
</tr>
<tr>
<td>BISVP: Building Footprint Extraction Via Bidirectional Serialized Vertex Prediction</td>
<td>2372</td>
</tr>
<tr>
<td>Mingming Zhang, Ye Du, Zhenghui Hu, Qingjie Liu, Yunhong Wang</td>
<td></td>
</tr>
<tr>
<td>TRICL: Triplet Continual Learning</td>
<td>2377</td>
</tr>
<tr>
<td>Xianchao Zhang, Guanglu Wang, Xiaotong Zhang, Han Liu, Zhengxi Yin, Wentao Yang</td>
<td></td>
</tr>
<tr>
<td>Chord-Conditioned Melody Harmonization with Controllable Harmonicity</td>
<td>2382</td>
</tr>
<tr>
<td>Shangda Wu, Xiaobing Li, Maosong Sun</td>
<td></td>
</tr>
<tr>
<td>Multilayer Subspace Learning with Self-Sparse Robustness for Two-Dimensional Feature Extraction</td>
<td>2387</td>
</tr>
<tr>
<td>Han Zhang, Maoguo Gong, Feiping Nie, Xuelong Li</td>
<td></td>
</tr>
<tr>
<td>Semantic Preserving Learning for Task-Oriented Point Cloud Downsampling</td>
<td>2392</td>
</tr>
<tr>
<td>Jianyu Xiong, Tao Dai, Yaohua Zha, Xin Wang, Shu-Tao Xia</td>
<td></td>
</tr>
<tr>
<td>Self-Supervised Representations in Speech-Based Depression Detection</td>
<td>2397</td>
</tr>
<tr>
<td>Wen Wu, Chao Zhang, Philip C. Woodland</td>
<td></td>
</tr>
<tr>
<td>A Novel Mode Selection-Based Fast Intra Prediction Algorithm for Spatial SHVC</td>
<td>2402</td>
</tr>
<tr>
<td>Dayong Wang, Yu Sun, Weisheng Li, Lele Xie, Xin Lu, Frederic Dufaux, Ce Zhu</td>
<td></td>
</tr>
<tr>
<td>Speech Emotion Recognition Via Heterogeneous Feature Learning</td>
<td>2407</td>
</tr>
<tr>
<td>Ke Liu, Dongya Wu, Dekui Wang, Jun Feng</td>
<td></td>
</tr>
<tr>
<td>Improving Contextual Spelling Correction by External Acoustics Attention and Semantic Aware Data Augmentation</td>
<td>2412</td>
</tr>
<tr>
<td>Xiaoqiang Wang, Yanqing Liu, Jinyu Li, Sheng Zhao</td>
<td></td>
</tr>
<tr>
<td>Deep Survival Analysis and Counterfactual Inference Using Balanced Representations</td>
<td>2417</td>
</tr>
<tr>
<td>Muskan Gupta, Gokul Kannan, Ranjitha Prasad, Garima Gupta</td>
<td></td>
</tr>
<tr>
<td>Decorrelating Language Model Embeddings for Speech-Based Prediction of Cognitive Impairment</td>
<td>2422</td>
</tr>
<tr>
<td>Lingfeng Xu, Kimberly D. Mueller, Julie Liss, Visar Berisha</td>
<td></td>
</tr>
<tr>
<td>The Edinburgh International Accents of English Corpus: Towards the Democratization of English ASR</td>
<td>2427</td>
</tr>
<tr>
<td>Ramon Sanabria, Nikolay Bogoychev, Nina Markl, Andrea Carmantini, Ondrej Klejch, Peter Bell</td>
<td></td>
</tr>
<tr>
<td>Prototype Knowledge Distillation for Medical Segmentation with Missing Modality</td>
<td>2432</td>
</tr>
<tr>
<td>Shuai Wang, Zipei Yan, Daohan Zhang, Haining Wei, Zhongsen Li, Rui Li</td>
<td></td>
</tr>
<tr>
<td>Multi-Temporal Lip-Audio Memory for Visual Speech Recognition</td>
<td>2437</td>
</tr>
<tr>
<td>Jeong Hun Yeo, Minsu Kim, Yong Man Ro</td>
<td></td>
</tr>
<tr>
<td>PQLM - Multilingual Decentralized Portable Quantum Language Model</td>
<td>2442</td>
</tr>
<tr>
<td>Shuyue Stella Li, Xiangyu Zhang, Shu Zhou, Hongchao Shu, Ruixing Liang, Hexin Liu, Leibny Paola Garcia</td>
<td></td>
</tr>
</tbody>
</table>
PAGE: A Position-Aware Graph-Based Model for Emotion Cause Entailment in Conversation
Xiaojie Gu, Renze Lou, Lin Sun, Shangxin Li

Switching Kronecker Product Linear Filtering for Multispeaker Adaptive Speech Dereverberation
Gongping Huang, Jacob Benesty, Israel Cohen, Emil Winebrand, Jingdong Chen, Walter Kellermann

Diffusionnet: An Efficient Framework to Classify Single-Molecule Images with Latent Entropy Minimization
Soumee Guha, Olivia De Cuba, Andreas Gahlmann, Scott T. Acton

Antenna Impedance Estimation in Correlated Rayleigh Fading Channels
Shaohan Wu, Brian L. Hughes

Hierarchical Filtering with Online Learned Priors for ECG Denoising
Timur Locher, Guy Revach, Nir Shlezinger, Ruud J. G. Van Sloun, Rik Vullings

Batch-Ensemble Stochastic Neural Networks for Out-of-Distribution Detection
Xiongjie Chen, Yinpeng Li, Yongxin Yang

Is Multi-Task Learning an Upper Bound for Continual Learning?
Zihao Wu, Huy Tran, Hamed Parsiavash, Soheil Kolouri

M^2TSR: Multi-Range and Mix-Grained Transformer for Single Image Super-Resolution
Zhong-Han Niu, Qing-Long Zhang, Yi Fan, Yu-Bin Yang

Large Covariance Matrix Estimation with Oracle Statistical Rate
Quan Wei, Ziping Zhao

Improving Performance of Real-Time Full-Band Blind Packet-Loss Concealment with Predictive Network
Viet-Anh Nguyen, Anh H. T. Nguyen, Andy W. H. Khong

Exploring Self-Supervised Pre-Trained ASR Models for Dysarthric and Elderly Speech Recognition
Shujie Hu, Xurong Xie, Zengrui Jin, Mengze Geng, Yi Wang, Mingyu Cui, Jiajun Deng, Xunying Liu, Helen Meng

Ultimate Negative Sampling for Contrastive Learning
Huijie Guo, Lei Shi

Designing a 3D-Aware Stylenerf Encoder for Face Editing
Songlin Yang, Wei Wang, Bo Peng, Jing Dong

Bias Reduced Semidefinite Relaxation Method for Multistatic Localization in the Absence of Transmitter Position and Its Synchronization
Jian Pei, Gang Wang, K. C. Ho, Lei Huang

Super-Resolution Information Enhancement for Crowd Counting
Jiahao Xie, Wei Xu, Dingkang Liang, Zhanyu Ma, Kongming Liang, Weidong Liu, Rui Wang, Ling Jin

Shadocnet: Learning Spatial-Aware Tokens in Transformer for Document Shadow Removal
Xuhang Chen, Xiaodong Cun, Chi-Man Pun, Shuqiang Wang

Fine-Grained Blind Face Inpainting with 3D Face Component Disentanglement
Yu Bai, Ruian He, Weimin Tan, Bo Yan, Yangle Lin
Sample-Adapt Fusion Network for RGB-D Hand Detection in the Wild.................................................. 2532
  Xingyu Liu, Pengfei Ren, Yuchen Chen, Cong Liu, Jing Wang, Hai Feng Sun, Qi Qi, Jingyu Wang

Inter-Pulse Estimation for Sperm Whale Click Detection ................................................................. 2537
  Guy Gubnitsky, Roee Diamant

Dynamic Vehicle Graph Interaction for Trajectory Prediction Based on Video Signals .......................... 2542
  Jian Chen, Wei Wang, Junxin Chen, Ming Cai

Pre-Trained Model Representations and Their Robustness Against Noise for Speech Emotion
Analysis.............................................................................................................................................. 2547
  Vikramjit Mitra, Vasudha Kowtha, Hsiang-Yun Sherry Chien, Erdrin Azemi, Carlos Avendano

Distributed Admm with Limited Communications Via Deep Unfolding ................................................. 2552
  Yoav Noah, Nir Shlezinger

Cross-Domain Learning with Normalizing Flow ..................................................................................... 2557
  Chi Wang, Jian Gao, Yang Hua, Hui Wang

A Fusion-Based and Multi-Layer Method for Low Light Image Enhancement ........................................ 2562
  Xueyan Zhou, Jiacen Guo, Hao Liu, Chao Wang

Assessing the Robustness of Deep Learning-Assisted Pathological Image Analysis Under Practical
Variables of Imaging System .............................................................................................................. 2567
  Yuxuan Sun, Chenglu Zhu, Yunlong Zhang, Hong Lin Li, Pingyi Chen, Lin Yang

SFR: Semantic-Aware Feature Rendering of Point Cloud ......................................................................... 2572
  Yaohua Zha, Rongsheng Li, Tao Dai, Jianyu Xiong, Xin Wang, Shu-Tao Xia

LQGNET: Hybrid Model-Based and Data-Driven Linear Quadratic Stochastic Control ......................... 2577
  Solomon Goldgraber Casspi, Oliver Hüsser, Guy Revach, Nir Shlezinger

Self-Supervised Guided Hypergraph Feature Propagation for Semi-Supervised Classification with
Missing Node Features .......................................................................................................................... 2582
  Chengxiang Lei, Sichao Fu, Yuetian Wang, Wenhao Qiu, Yachen Hu, Qinmu Peng, Xinge You

Contrastive Learning of Sentence Embeddings in Product Search ......................................................... 2587
  Bo-Wen Zhang, Yan Yan, Jiapei Yu

Boosting Bert Subnets with Neural Grafting ............................................................................................. 2592
  Ting Hu, Christoph Meinel, Haojin Yang

Enhanced GM-PHD Filter for Real Time Satellite Multi-Target Tracking .............................................. 2597
  Camilo Aguilar, Mathias Ortner, Josiane Zerubia

A Wavelet Scattering Approach for Load Identification with Limited Amount of Training Data .......... 2602
  Pascal A. Schirmer, Iosif Mporas

Automatic Severity Classification of Dysarthric Speech by Using Self-Supervised Model with
Multi-Task Learning ............................................................................................................................. 2607
  Eun Jung Yeo, Kwangheee Choi, Sunhee Kim, Minhwa Chung

Ternary Weight Networks ...................................................................................................................... 2612
  Bin Liu, Fengfu Li, Xiaoxing Wang, Bo Zhang, Junchi Yan
A Slot-Shared Span Prediction-Based Neural Network for Multi-Domain Dialogue State Tracking .......... 2617
Abibulla Atawulla, Xi Zhou, Yating Yang, Bo Ma, Fengyi Yang

CPA: Compressed Private Aggregation for Scalable Federated Learning Over Massive Networks .......... 2622
Natalie Lang, Elad Sofer, Nir Shlezinger, Rafael G. L. D’Oliveira, Salim El Rouayheb

BATT: Backdoor Attack with Transformation-Based Triggers ..................................................... 2627
Tong Xu, Yiming Li, Yong Jiang, Shu-Tao Xia

A Critical Look at Recent Trends in Compression of Channel State Information.............................. 2632
Marcus Valtonen Örnhag, Stefan Adalbjörnsson, Püren Güler, Mojtaba Mahdavi

Cross-Modal Adversarial Contrastive Learning for Multi-Modal Rumor Detection .................................. 2637
Ting Zou, Zhong Qian, Peifeng Li, Qiaoming Zhu

AMC-Net: An Effective Network for Automatic Modulation Classification ......................................... 2642
Jiawei Zhang, Tiantian Wang, Zhixi Feng, Shuyuan Yang

WHC: Weighted Hybrid Criterion for Filter Pruning on Convolutional Neural Networks ...................... 2647
Shaowu Chen, Weize Sun, Lei Huang

Transformer-Based Deep Hashing Method for Multi-Scale Feature Fusion ........................................... 2652
Chao He, Hongxi Wei

STACKMAPS: A Visualization Technique for Diabetic Retinopathy Grading ........................................ 2657
Ismail El-Yamany, Abdelrahman Wael, Noha Adly, Marwan Torki

TAPE: An End-To-End Timbre-Aware Pitch Estimator ........................................................................... 2662
Nazif Can Tamer, Yigitcan Özer, Meinard Müller, Xavier Serra

Towards Zero-Shot Personalized Table-to-Text Generation with Contrastive Persona Distillation ............ 2667
Haolan Zhan, Xuming Lin, Shaobo Cui, Zhongzhou Zhao, Wei Zhou, Haiqing Chen

Associative Learning Network for Coherent Visual Storytelling ......................................................... 2672
Xin Li, Chunping Liu, Yi Ji

Diabetic Retinopathy Grading with Weakly-Supervised Lesion Priors................................................... 2677
Junlin Hou, Fan Xiao, Jilan Xu, Rui Feng, Yuejie Zhang, Haidong Zou, Lina Lu, Wenwen Xue

Gridless Target Localization for FDA-Mimo Radar with Sparse Arrays ................................................ 2682
Xiaohuan Wu, Yaxin Liu, Xiaoyuan Jia

BECTRA: Transducer-Based End-To-End ASR with Bert-Enhanced Encoder ........................................... 2687
Yosuke Higuchi, Tetsuji Ogawa, Tetsunori Kobayashi, Shinji Watanabe

Robust GMM Parameter Estimation Via the K-BM Algorithm ..................................................................... 2692
Ori Kenig, Koby Todros, Tülay Adali

Interml: Momentum Pseudo-Labeling with Intermediate CTC Loss ..................................................... 2697
Yosuke Higuchi, Tetsuji Ogawa, Tetsunori Kobayashi, Shinji Watanabe

Brainnetformer: Decoding Brain Cognitive States with Spatial-Temporal Cross Attention ...................... 2702
Leheng Sheng, Wenhan Wang, Zhiyi Shi, Jichao Zhan, Youyong Kong

Convergence of Stochastic PDMM ............................................................................................................ 2707
Sebastian O. Jordan, Thomas W. Sherson, Richard Heusdens
Spatial Similarity Guidance for Few-Shot Segmentation .................................................. 2712
  Xiaoliu Luo, Zhao Duan, Taiping Zhang

Self-Distillation Hashing for Efficient Hamming Space Retrieval .................................. 2717
  Hongjia Zhai, Hai Li, Hanzhi Zhang, Hujun Bao, Guofeng Zhang

Benchmarking White Blood Cell Classification Under Domain Shift .................................. 2722
  Satoshi Tsutsui, Zhengyang Su, Bihan Wen

Exploiting Sparse Recovery Algorithms for Semi-Supervised Training of Deep Neural Networks for Direction-of-Arrival Estimation ........................................................................ 2727
  Murtiza Ali, Aditya Arie Nugraha, Karan Nathwani

QTROJAN: A Circuit Backdoor Against Quantum Neural Networks ...................................... 2732
  Cheng Chu, Lei Jiang, Martin Swany, Fan Chen

Deep Root Music Algorithm for Data-Driven DoA Estimation ............................................ 2737
  Dor H. Shmuel, Julian P. Merkofer, Guy Revach, Ruud J. G. Van Sloun, Nir Shlezinger

SDG-L: A Semiparametric Deep Gaussian Process Based Framework for Battery Capacity Prediction ......................................................................................................................... 2742
  Hanbing Liu, Yanru Wu, Yang Li, Ercan E. Kuruoglu, Xuan Zhang

RAT: Radial Attention Transformer for Singing Technique Recognition .................................. 2747
  Guan-Yuan Chen, Ya-Fen Yeh, Von-Wun Soo

Emodiff: Intensity Controllable Emotional Text-To-Speech with Soft-Label Guidance ............ 2752
  Yiwei Guo, Chenpeng Du, Xie Chen, Kai Yu

MEET: A Monte Carlo Exploration-Exploitation Trade-Off for Buffer Sampling .................... 2757
  Julius Ott, Lorenzo Servadei, Jose Arjona-Medina, Enrico Rinaldi, Gianfranco Mauro, Daniela Sanchez Lopera, Michael Stephan, Thomas Stadelmayer, Avik Santra, Robert Wille

Lego-Features: Exporting Modular Encoder Features for Streaming and Deliberation ASR .... 2762
  Rami Botros, Rohit Prabhavalkar, Johan Schalkwyk, Ciprian Chelba, Tara N. Sainath, Francoise Beaufays

Robust Subspace Tracking with Contamination Mitigation Via \(\alpha\)-Divergence ...................... 2767
  Le Trung Thanh, Aref Miri Rekavandi, Abd-Krim Seghouane, Karim Abed-Merai

CLAP Learning Audio Concepts from Natural Language Supervision .................................... 2772
  Benjamin Elizalde, Soham Deshmukh, Mahmoud Al Ismail, Huaming Wang

Self-Supervised Learning with Bi-Label Masked Speech Prediction for Streaming Multi-Talker Speech Recognition ................................................................................................. 2777
  Zili Huang, Zhuo Chen, Naoyuki Kanda, Jian Wu, Yiming Wang, Jinyu Li, Takuya Yoshioka, Xiaofei Wang, Peidong Wang

Rate Region Characterization for Semantics and Bits Based Multiuser Communications ........ 2782
  Xidong Mu, Yuanwei Liu

Cascading and Direct Approaches to Unsupervised Constituency Parsing on Spoken Sentences ................................................................................................................................. 2787
  Yuan Tseng, Cheng-I Jeff Lai, Hung-Yi Lee

A Nested Ensemble Method to Bilevel Machine Learning ...................................................... 2792
  Lisha Chen, Momin Abbas, Tianyi Chen
Police: Provably Optimal Linear Constraint Enforcement for Deep Neural Networks

Randall Balestriero, Yann Lecun

Heterogeneous Graph Learning for Acoustic Event Classification

Amir Shirian, Mona Ahmadian, Krishna Somandepalli, Tanaya Guha

Epic-Sounds: A Large-Scale Dataset of Actions that Sound

Jaesung Huh, Jacob Chalk, Evangelos Kazakos, Dima Damen, Andrew Zisserman

Ensemble Graph Q-Learning for Large Scale Networks

Talha Bozkus, Urbashi Mitra

Terminology-Aware Medical Dialogue Generation

Chen Tang, Hongbo Zhang, Tyler Loakman, Chenghua Lin, Frank Guerin

Prefix-Level Detection and Autocorrection of Keyboard Input Errors

Jerome R. Bellegarda

Enhanced Embeddings in Zero-Shot Learning for Environmental Audio

Ysobel Sims, Alexandre Mendes, Stephan Chalup

SD-PINN: Physics Informed Neural Networks for Spatially Dependent PDES

Ruixian Liu, Peter Gerstoft

EEG Emotion Recognition Via Ensemble Learning Representations

Bilal Taha, Dae Yon Hwang, Dimitrios Hatzinakos

Dual-Use Signal Design for MIMO Radcom with Inter-Pulse Index Modulation

Xue Yao, Guolong Cui, Xianxiang Yu

Unsupervised Vocal Dereverberation with Diffusion-Based Generative Models

Koichi Saito, Naoki Murata, Toshimitsu Uesaka, Chieh-Hsin Lai, Yuhta Takida, Takao Fukui, Yuki Mitsufuji

Towards Trustworthy Phoneme Boundary Detection with Autoregressive Model and Improved Evaluation Metric

Hyeongju Kim, Hyeong-Seok Choi

Dual-Head Fusion Network for Image Enhancement

Yuhong Zhang, Hengsheng Zhang, Li Song, Rong Xie, Wenjun Zhang

Spatial-Temporal Graph Convolutional Network Boosted Flow-Frame Prediction for Video Anomaly Detection

Kai Cheng, Xinhua Zeng, Yang Liu, Mengyang Zhao, Chengxin Pang, Xing Hu

Nasty-SFDA: Source Free Domain Adaptation from a Nasty Model

Jiajiong Cao, Yufan Liu, Weiming Bai, Jingting Ding, Liang Li

Removing Radio Frequency Interference from Auroral Kilometric Radiation with Stacked Autoencoders

Allen Chang, Mary Knapp, James Labelle, John Swoboda, Ryan Volz, Philip J. Erickson

Shift to Your Device: Data Augmentation for Device-Independent Speaker Verification Anti-Spoofing

Junhao Wang, Li Lu, Zhongjie Ba, Feng Lin, Kui Ren
Transformer-Based Tracking Network for Maneuvering Targets ........................................... 2882
Yushu Zhang, Gang Li, Xiao-Ping Zhang, You He

Multi-Stage Aggregation Transformer for Medical Image Segmentation ................................ 2887
Xiaoyan Wang, Minghan Shao, Dongyan Guo, Ying Cui, Xiaojie Huang, Ming Xia, Cong Bai

Ultrasound Image Quality Control Using Speech-Assisted Switchable CycleGAN .................. 2892
Jaeyoung Huh, Shujaat Khan, Eun Sun Lee, Jong Chul Ye

Efficient Quantized Constant Envelope Precoding for Multiuser Downlink Massive MIMO Systems ................................................................. 2897
Zheyu Wu, Ya-Feng Liu, Bo Jiang, Yu-Hong Dai

Image Fusion Via Slice-Based Convolutional Sparse Representation ................................... 2902
Jingchen Xu, Yali Zhang, Ze Li, Jinjia Wang

ICCRN: Inplace Cepstral Convolutional Recurrent Neural Network for Monaural Speech Enhancement .............................................................................. 2907
Jinjiang Liu, Xueliang Zhang

Channel State Information-Free Artificial Noise-Aided Location-Privacy Enhancement ............. 2912
Jianxiu Li, Urbashi Mitra

CAENet: Using Collaborative Attention Transformer and Add-Boost Strategy for Single Image Deraining .................................................................................. 2917
Shengdi Qin, Shunli Zhang, Yu Zhang, Haoyu Gao

Sparse Aggregation-Based Channel Estimation for Massive Mimo Systems with Decentralized Baseband Processing ........................................... 2922
Yanqing Xu, Enbin Song, Qingjiang Shi, Tsung-Hui Chang

A Few Shot Learning of Singing Technique Conversion Based on Cycle Consistency Generative Adversarial Networks ...................................................... 2927
Po-Wei Chen, Von-Wun Soo

Geogcn: Geometric Dual-Domain Graph Convolution Network for Point Cloud Denoising .......... 2932
Zhaowei Chen, Peng Li, Zeyong Wei, Honghua Chen, Haoran Xie, Mingqiang Wei, Fu Lee Wang

Wireless Location Tracking Via Complex-Domain Super MDS with Time Series Self-Localization Information .......................................................................... 2937
Yuya Nishi, Takumi Takahashi, Hiroki Ilmori, Giuseppe Abreu, Shinsuke Ibi, Seiichi Sampei

A Graph Neural Network Multi-Task Learning-Based Approach for Detection and Localization of Cyberattacks in Smart Grids ........................................................................... 2942
Abdurahman Takiddin, Rachad Atat, Muhammad Ismail, Katherine Davis, Erchin Serpedin

Deep Double Self-Expressive Subspace Clustering .................................................................... 2947
Ling Zhao, Yunpeng Ma, Shanxiong Chen, Jun Zhou

Dual-Stream Siamese Vision Transformer with Mutual Attention for Radar Gait Verification .......... 2952
Ran Ji, Jiarui Li, Wentao He, Jianfeng Ren, Xudong Jiang

Locality Preserving Multiview Graph Hashing for Large Scale Remote Sensing Image Search ........ 2957
Wenyun Li, Guo Zhong, Xingyu Lu, Chi-Man Pun
Dual Path Modeling for Semantic Matching by Perceiving Subtle Conflicts ................................................................. 2962
Chao Xue, Di Liang, Sirui Wang, Jing Zhang, Wei Wu

UNTAG: Learning Generic Features for Unsupervised Type-Agnostic Deepfake Detection ........................................... 2967
Nesryne Mejri, Enjie Ghorbel, Djamila Aouada

A Lightweight Fourier Convolutional Attention Encoder for Multi-Channel Speech Enhancement .......................... 2972
Siyu Sun, Jian Jin, Zhe Han, Xianjun Xia, Li Chen, Yijian Xiao, Piao Ding, Shenyi Song, Roberto Togneri, Haijian Zhang

Iterative Shallow Fusion of Backward Language Model for End-To-End Speech Recognition ............................... 2977
Atsunori Ogawa, Takafumi Moriya, Naoyuki Kamo, Naohiro Tawara, Marc Delcroix

Parallel 2D Seismic Ray Tracing Using Cuda on a Jetson Nano ................................................................. 2982
Ban-Sok Shin, Luis Wientgens, Dmitriy Shutin

Generalized Relative Harmonic Coefficients ................................................................. 2987
Yonggang Hu, Sharon Gannot, Thushara D. Abhayapala

Model-Matching Principle Applied to the Design of an Array-Based All-Neural Binaural Rendering System for Audio Telepresence ................................................................. 2992
Yicheng Hsu, Chenghong Ma, Mingsian R. Bai

Learning Task-Aligned Mask Query for Instance Segmentation ................................................................. 2997
Bin Fu, Hongliang He, Pengxu Wei, Jie Chen

PI-Trans: Parallel-Convmlp and Implicit-Transformation Based Gan for Cross-View Image Translation .......................... 3002
Bin Ren, Hao Tang, Yiming Wang, Xia Li, Wei Wang, McU Sebe

Self-Transriber: Few-Shot Lyrics Transcription with Self-Training ................................................................. 3007
Xiaoxue Gao, Xianghu Yue, Haizhou Li

Multi-Level Fusion for Burst Super-Resolution with Deep Permutation-Invariant Conditioning ................................ 3012
Martina Cilia, Diego Valsesia, Giulia Fracastoro, Enrico Magli

Source-Free Unsupervised Domain Adaptation for Question Answering ................................................................. 3017
Zishuo Zhao, Yuexiang Xie, Jingyou Xie, Zhenzhou Lin, Yaliang Li, Ying Shen

Doppler-Coded Joint Division Multiple Access Waveform for Automotive MIMO Radar .................................................. 3022
Yanhua Wang, Qiubo Pei, Xueyao Hu, Jiamin Long, Hao Yu, Le Zheng

Maximum Likelihood Distillation for Robust Modulation Classification ................................................................. 3027
Javier Maroto, Gérôme Bovet, Pascal Frossard

Speaker Diaphragm Excursion Prediction: Deep Attention and Online Adaptation .................................................. 3032
Yuwei Ren, Matt Zivney, Yin Huang, Eddie Choy, Chirag Patel, Hao Xu

A Novel Heart Rate Estimation Method Exploiting Heartbeat Second Harmonic Reconstruction Via Millimeter Wave Radar .................................................................................. 3037
Tao Li, Huayu Shou, Yuchen Deng, Yu Zhou, Chenqi Shi, Pengpeng Chen

Centroid Distance Distillation for Effective Rehearsal in Continual Learning .......................................................... 3042
Daofeng Liu, Fan Lyu, Linyan Li, Zhenping Xia, Fuyuan Hu

On Tracking a Stochastically Time-Varying Subspace ................................................................................................. 3047
Victor Solo
Difference Guided VHR Remote Sensing Image Change Detection............................................................... 3052
  Jiukai Sun, Ganchao Liu, Xuelong Li, Yuan Yuan

TT-Net: Dual-Path Transformer Based Sound Field Translation in the Spherical Harmonic Domain .......... 3057
  Yiwen Wang, Zijian Lan, Xihong Wu, Tianshu Qu

Parafac2-Based Coupled Matrix and Tensor Factorizations................................................................. 3062
  Carla Schenker, Xiulin Wang, Evrim Acar

False Alarm Regulation for Off-Grid Target Detection with the Matched Filter..................................... 3067
  P Develter, J. Bosse, O. Rabaste, P. Forster, J.-P. Ovarlez

MCKD: Mutually Collaborative Knowledge Distillation for Federated Domain Adaptation and
  Generalization ........................................................................................................................................ 3072
  Ziwei Niu, Hongyi Wang, Hao Sun, Shuyi Ouyang, Yen-Wei Chen, Lanfen Lin

A DNN-Based Hearing-Aid Strategy for Real-Time Processing: One Size Fits All ............................... 3077
  Fotios Drakopoulos, Arthur Van Den Broucke, Sarah Verhulst

Unsupervised Model-Based Speaker Adaptation of End-To-End Lattice-Free MMI Model for
  Speech Recognition ................................................................................................................................ 3082
  Xurong Xie, Xuanying Liu, Hui Chen, Hongan Wang

A Bandit Online Convex Optimization Approach to Distributed Energy Management in Networked
  Systems ...................................................................................................................................................... 3087
  Ioannis Tsetis, Xiaotong Cheng, Setareh Maghsudi

Query-Utterance Attention with Joint Modeuing for Query-Focused Meeting Summarization.................. 3092
  Xingxian Liu, Bin Duan, Bo Xiao, Yajing Xu

Optimized Dithering for Quantization Index Modulation ............................................................................ 3097
  Shanxiang Lyu

One-Shot Medical Action Recognition with a Cross-Attention Mechanism and Dynamic Time
  Warping ..................................................................................................................................................... 3102
  Leiyu Xie, Yuxing Yang, Zeyu Fu, Syed Mohsen Naqvi

Defense Against Black-Box Adversarial Attacks Via Heterogeneous Fusion Features ............................. 3107
  Jiahuan Zhang, Keisuke Maeda, Takahiro Ogawa, Miki Haseyama

Slot-Triggered Contextual Biasing for Personalized Speech Recognition Using Neural Transducers ........112
  Sibo Tong, Philip Harding, Simon Wiesler

Model-Free Online Learning for Waveform Optimization in Integrated Sensing and
  Communications ......................................................................................................................................... 3117
  Petteri Pulkkinen, Visa Koivunen

Graph Learning from Gaussian and Stationary Graph Signals ................................................................. 3122
  Andrei Buciulea, Antonio G. Marques

Hierarchical Softmax for End-To-End Low-Resource Multilingual Speech Recognition .......................... 3127
  Qianying Liu, Zhuo Gong, Zhengdong Yang, Yuhang Yang, Sheng Li, Chenchen Ding,
  Nobuaki Minematsu, Hao Huang, Fei Cheng, Chenhui Chu, Sadao Kurohashi

DMSA: Dynamic Multi-Scale Unsupervised Semantic Segmentation Based on Adaptive Affinity .......... 3132
  Kun Yang, Jun Lu
Image Completion Via Dual-Path Cooperative Filtering
Pourya Shamsolmoali, Masoumeh Zareapoor, Eric Granger

RGB-D Based Pose-Invariant Face Recognition Via Attention Decomposition Module
Wei-Chen Lin, Ching-Te Chiu, Kuan-Chang Shih

Parameter-Efficient Transfer Learning of Pre-Trained Transformer Models for Speaker Verification Using Adapters
Junyi Peng, Themos Stafylakis, Rongzhi Gu, Oldrich Pecht, Ladislav Mošner, Lukáš Burget, Jan Cernocký

A Dynamic Graph Interactive Framework with Label-Semantic Injection for Spoken Language Understanding
Zhihong Zhu, Wei yuan Xu, Xuxin Cheng, Tengtao Song, Yuexian Zou

Hyneter: Hybrid Network Transformer for Object Detection
Dong Chen, Duoqian Miao, Xuerong Zhao

Towards Domain Generalisation in ASR with Elitist Sampling and Ensemble Knowledge Distillation
Rehan Ahmad, Md Asif Jalal, Muhammad Umar Farooq, Anna Ollerenshaw, Thomas Hain

Structural Reparameterization Lightweight Network for Video Action Recognition
Anlei Zhu, Yinghui Wang, Wei Li, Pengjiang Qian

VLKP: Video Instance Segmentation with Visual-Linguistic Knowledge Prompts
Ruixiang Chen, Sheng Liu, Junhao Chen, Bingnan Guo, Feng Zhang

Inverse Reinforcement Learning with Graph Neural Networks for IoT Resource Allocation
Guangchen Wang, Peng Cheng, Zhuo Chen, Wei Xiang, Branka Vucetic, Yonghui Li

Cross-Site Generalization for Imbalanced Epileptic Classification
Tala Abdallah, Nisrine Jrad, Fahed Abdallah, Anne Humeau-Heurtier, Patrick Van Bogaert

Volumetric 3D Reconstruction with Window-Wise Global Feature Aggregation
Shihao Ren, Yikang Ding, Jinli Liao, Xinghui Li, Jia Guo, Wensen Feng, Xueqian Wang

FFFN: Fashion Feature Fusion Network by Co-Attention Model for Fashion Recommendation
Zhantu Lin, Xiaoyan Zhang

Low-Rank Constrained Memory Autoencoder for Hyperspectral Anomaly Detection
Yuyun Lian, Yongshan Zhang, Xuxiang Feng, Xinwei Jiang, Zhihua Cai

Effect of Lossy Compression Algorithms on Face Image Quality and Recognition
Torsten Schlett, Sebastian Schachner, Christian Rathgeb, Juan Tapia, Christoph Busch

Human Pose Estimation from Ambiguous Pressure Recordings with Spatio-Temporal Masked Transformers
Vandad Davoodnia, Ali Etemad

Fast Low-Latency Convolution by Low-Rank Tensor Approximation
Martin Jälmby, Filip Elvander, Toon Van Waterschoot

TrOMR: Transformer-Based Polyphonic Optical Music Recognition
Yixuan Li, Huaping Liu, Qiang Jin, Miaomiao Cai, Peng Li
Adversarial Data Augmentation Using VAE-GAN for Disordered Speech Recognition .............................................. 3222
Zengrui Jin, Xurong Xie, Mengzhe Geng, Tianzi Wang, Shujie Hu, Jiajun Deng, Guinan Li,
Xunying Liu

Scalable Multi-Task Semantic Communication System with Feature Importance Ranking ........................................ 3227
Jiangjing Hu, Fengyu Wang, Wenjion Xu, Hui Gao, Ping Zhang

Design and Performance of the Low-Power Noise Reduction Algorithm of the Med-El Sonnet 2™ Cochlear Implant Audio Processor ........................................................................................................... 3232
Ernst Aschbacher, Florian Frühauf, Anja Kurz, Peter Nopp

An Evaluation Platform to Scope Performance of Synthetic Environments in Autonomous Ground Vehicles Simulation ......................................................................................................................... 3237
Xiangyu Bai, Le Jiang, Yedi Luo, Aniket Gupta, Pushyami Kaveti, Hanumant Singh, Sarah Ostadabbas

Your Camera Improves Your Point Cloud Compression .......................................................................................... 3242
Yuhuan Lin, Tongda Xu, Ziyu Zhu, Yanghao Li, Zhe Wang, Yan Wang

Angle-Of-Arrival Target Tracking Using a Mobile Uav in External Signal-Denied Environment .......................... 3247
Bing Zhu, Sheng Xu, Feng Rice, Kutlayil Dogançay

Inductive Relation Prediction from Relational Paths and Context with Hierarchical Transformers .......................... 3252
Jiaang Li, Quan Wang, Zhendong Mao

MSFORMER: Multi-Scale Transformer with Neighborhood Consensus for Feature Matching ............................... 3257
Dongyue Li, Yaping Yan, Dong Liang, Songlin Du

ECG Artifact Removal from Single-Channel Surface EMG Using Fully Convolutional Networks ........................ 3262
Kuan-Chen Wang, Kai-Chun Liu, Sheng-Yu Peng, Yu Tsao

Volumetric Attribute Compression for 3D Point Clouds Using Feedforward Network with Geometric Attention ................................................................................................................................. 3267
Tam Thuc Do, Philip A. Chou, Gene Cheung

Continual Cell Instance Segmentation of Microscopy Images .................................................................................. 3272
Tzu-Ting Chuang, Ting-Yun Wei, Yu-Hsieh Hsieh, Chu-Song Chen, Huei-Fang Yang

High-Frequency Transformer Network Based on Window Cross-Attention for Pansharpening .......................... 3277
Chengjie Ke, Hao Liang, Duidui Li, Xin Tian

An Efficient Beam-Sharing Algorithm for RIS-aided Simultaneous Wireless Information and Power Transfer Applications .............................................................................................................. 3282
Nguyen Minh Tran, Muhammad Miftahul Amri, Je Hyeon Park, Dong In Kim, Kae Won Choi

Cumulative Attention Based Streaming Transformer ASR with Internal Language Model Joint Training and Rescoring .................................................................................................................. 3287
Mohan Li, Cong-Thanh Do, Rama Doddipatla

A 3D-Assisted Framework to Evaluate the Quality of Head Motion Replication by Reenactment DEEPFAKE Generators ............................................................................................................. 3292
Sahar Husseini, Jean-Luc Dugelay, Fabien Aili, Emmanuel Nars

Cross-Modal Audio-Visual Co-Learning for Text-Independent Speaker Verification .............................................. 3297
Meng Liu, Kong Aik Lee, Longbiao Wang, Hanyi Zhang, Chang Zeng, Jianwu Dang
Multiple Domain-Adversarial Ensemble Learning for Domain Generalization .............................................. 3302
Ze-Yu Mi, Kun Long, Yu-Bin Yang

Compose & Embellish: Well-Structured Piano Performance Generation Via a Two-Stage Approach ........ 3307
Shih-Lun Wu, Yi-Hsuan Yang

Multi-Scale Receptive Field Graph Model for Emotion Recognition in Conversations ......................... 3312
Jie Wei, Guanyu Hu, Luu Anh Tuan, Xinyu Yang, Wenjing Zhu

Meta-Dag: Meta Causal Discovery Via Bilevel Optimization .................................................................... 3317
Songtao Lu, Tian Gao

Efficient Monaural Speech Enhancement with Universal Sample Rate Band-Split RNN ......................... 3322
Jianwei Yu, Yi Luo

LA-VOCE: LOW-SNR Audio-Visual Speech Enhancement Using Neural Vocoders ............................ 3327
Rodrigo Mira, Buye Xu, Jacob Donley, Anurag Kumar, Stavros Petridis, Vamsi Krishna
Ithapu, Maja Pantic

Outlier-Insensitive Kalman Filtering Using NUV Priors ................................................................. 3332
Shunit Truzman, Guy Revach, Nir Shlezinger, Itzik Klein

Semantics-Disentangled Contrastive Embedding for Generalized Zero-Shot Learning ...................... 3337
Jian Ni, Yong Liao

Learnable Frontends that Do Not Learn: Quantifying Sensitivity to Filterbank Initialisation .............. 3342
Mark Anderson, Tomi Kinnunen, Naomi Harte

Ruize Xu, Ruoxuan Feng, Shi-Xiong Zhang, Di Hu

Balanced Mixup Loss for Long-Tailed Visual Recognition ................................................................... 3352
Haibo Ye, Fangyu Zhou, Xinjie Li, Qingheng Zhang

Local Feature Enhanced Adversarial Network for the Blind Image Quality Assessment .................. 3357
Xiaomei Shi, Min Zhang, Shouhai Xia, Ruxue Zhang, Jun Feng

A Multi-Signal Perception Network for Textile Composition Identification ....................................... 3362
Bo Peng, Liren He, Dong Wu, Mingmin Chi, Jintao Chen

Cross-Speaker Emotion Transfer by Manipulating Speech Style Latents .......................................... 3367
Suhee Jo, Younggun Lee, Yookyung Shin, Yeongtae Hwang, Taesu Kim

Neural Network Models with Integrated Training and Adaptation for Nonlinear Acoustic System
Identification ....................................................................................................................................... 3372
Svantje Voit, Gerald Enzner

Fast Online Source Steering Algorithm for Tracking Single Moving Source Using Online
Independent Vector Analysis ................................................................................................................. 3377
Taishi Nakashima, Rintaro Ikeshita, Nobutaka Ono, Shoko Araki, Tomohiro Nakatani

Priv-Aug-Shap-ECGResNet: Privacy Preserving Shapley-Value Attributed Augmented Resnet for
Practical Single-Lead Electrocardiogram Classification ........................................................................ 3382
Arijit Ukil, Leandro Marin, Antonio J. Jara

Single-Photon Image Super-Resolution Via Self-Supervised Learning ............................................. 3387
Yiwei Chen, Chen Jiang, Yu Pan
Column-Based Matrix Approximation with Quasi-Polynomial Structure ......................................................... 3476
   Jeongmin Chae, Praneeth Narayanamurthy, Selin Bac, Shaama Mallikarjun Sharada, Urbashi Mitra

Towards Polymorphic Adversarial Examples Generation for Short Text .......................................................... 3481
   Yuhang Liang, Zheng Lin, Fengcheng Yuan, Hanwen Zhang, Lei Wang, Weiping Wang

ERBNet: An Effective Representation Based Network for Unbiased Scene Graph Generation ......................... 3486
   Wenxi Ma, Tianxiang Hou, Qianji Di, Zhongang Qi, Ying Shan, Hanzi Wang

He-Gan: Differentially Private Gan Using Hamiltonian Monte Carlo Based Exponential Mechanism ......................... 3491
   Usman Hassan, Dongjie Chen, Sen-Ching S. Cheung, Chen-Nee Chuah

Order Reduction of Multi-Channel FIR Filters by Balanced Truncation .......................................................... 3496
   Florian Hilgemann, Peter Jax

Hyperbolic Audio Source Separation .................................................................................................................. 3501
   Darius Petermann, Gordon Wichern, Aswin Subramanian, Jonathan Le Roux

Learning Robust Self-Attention Features for Speech Emotion Recognition with Label-Adaptive Mixup ............ 3506
   Lei Kang, Lichao Zhang, Dazhi Jiang

An Analysis of Degenerating Speech Due to Progressive Dysarthria on ASR Performance .............................. 3511
   Katrin Tomanek, Katie Seaver, Pan-Pan Jiang, Richard Cave, Lauren Harrell, Jordan R. Green

Enhancing Representation Learning with Deep Classifiers in Presence of Shortcut ......................................... 3516
   Amirhossein Ahmadian, Fredrik Lindsten

Perceptual Analysis of Speaker Embeddings for Voice Discrimination Between Machine and Human Listening ........................................................................................................... 3521
   Iordanis Thoidis, Clément Gaultier, Tobias Goehring

An Online Algorithm for Chance Constrained Resource Allocation ................................................................. 3526
   Yuwei Chen, Zengde Deng, Yinzi Zhou, Zaiyi Chen, Yujie Chen, Haoyuan Hu

Learning with Multigraph Convolutional Filters ................................................................................................. 3531
   Landon Butler, Alejandro Parada-Mayorga, Alejandro Ribeiro

A New Probabilistic Distance Metric with Application in Gaussian Mixture Reduction .............................. 3536
   Ahmad Sajedi, Yuri A. Lawryshyn, Konstantinos N. Plataniotis

Cross-Domain Diffusion Based Speech Enhancement for Very Noisy Speech ............................................. 3541
   Heming Wang, Deliang Wang

Learned Kalman Filtering in Latent Space with High-Dimensional Data ......................................................... 3546
   Itay Buchnik, Damiano Steger, Guy Revach, Ruud J. G. Van Sloun, Tirza Routtenberg, Nir Shlezinger

Mendam: Multi-Expert Network with Distribution-Aware Momentum for Long-Tailed Recognition ........... 3551
   Qingheng Zhang, Haibo Ye, Kaicheng Yu

Cramér-Rao Bound on Lie Groups with Observations on Lie Groups: Application to SE(2) .......................... 3556
   Sámy Labsir, Alexandre Renaux, Jordi Vilà-Valls, Éric Chaumette
Signal Processing on Product Spaces ................................................................. 3561
T. Mitchell Roddenberry, Vincent P. Grande, Florian Frantzen, Michael T. Schaub, Santiago Segarra

Block-Based Color Constancy: The Deviation of Salient Pixels ........................................... 3566
Oguzhan Ulucan, Diclehan Ulucan, Marc Ebner

Difference Coarrays of Rational Arrays ........................................................................... 3571
Pranav Kulkarni, P. P. Vaidyanathan

Online Edge Flow Prediction Over Expanding Simplicial Complexes ................................ 3576
Maosheng Yang, Bishwadeep Das, Elvin Isufi

Spectral Clustering-Aware Learning of Embeddings for Speaker Diarisation .......................... 3581
Evonne P. C. Lee, Guangzhi Sun, Chao Zhang, Philip C. Woodland

Improvements to Embedding-Matching Acoustic-to-Word ASR Using Multiple-Hypothesis Pronunciation-Based Embeddings ....................................................... 3586
Hao Yen, Woojay Jeon

from Partial Observations ......................................................................................... 3591
Tomer Hershkovitz, Martin Haardt, Arie Yeredor

Implicit Bayes Adaptation: A Collaborative Transport Approach ......................................... 3596
Bo Jiang, Hamid Krim, Tianfu Wu, Derya Cansever

Adaptive Endpointing with Deep Contextual Multi-Armed Bandits ..................................... 3601
Do June Min, Andreas Stolcke, Anirudh Raju, Colin Vaz, Di He, Venkatesh Ravichandran, Viet Anh Trinh

Dynamic Speech Endpoint Detection with Regression Targets ............................................ 3606
Dawei Liang, Hang Su, Tarun Singh, Jay Mahadeokar, Shanil Puri, Jiedan Zhu, Edison Thomaz, Mike Seltzer

NVOC-22: A Low Cost Mel Spectrogram Vocoder for Mobile Devices .................................. 3611
Rakesh Iyer

Sparse and Structured Modelling of Underwater Acoustic Channel Impulse Responses .......... 3616
Chaoran Yang, Qing Ling, Xueli Sheng, Mengfei Mu, Andreas Jakobsson

Self-Supervised Representations for Singing Voice Conversion ............................................. 3621
Tejas Jayashankar, Jilong Wu, Leda Sari, David Kant, Vimal Manohar, Qing He

Improving Speech Prosody of Audiobook Text-To-Speech Synthesis with Acoustic and Textual Contexts .............................................................................................. 3626
Detai Xin, Sharath Adavanne, Federico Ang, Ashish Kulkarni, Shinnosuke Takamichi, Hiroshi Saruwatari

Utility Pole localization by Learning from Ambient Traces on Distributed Acoustic Sensing ............................................................... 3631
Zhuocheng Jiang, Yue Tian, Yangmin Ding, Sarper Ozharar, Ting Wang

Passive Acoustic Tracking of Whales in 3-D ........................................................................ 3636
Junsu Jang, Florian Meyer, Eric R. Snyder, Sean M. Wiggins, Simone Baumann-Pickering, John A. Hildebran
Customized Automatic Face Beautification

Wang Chen, Peizhen Chen, Weijie Chen, Luojun Lin

A Unified One-Shot Prosody and Speaker Conversion System with Self-Supervised Discrete Speech Units

Li-Wei Chen, Shinji Watanabe, Alexander Rudnicky

Phoneme-Level Bert for Enhanced Prosody of Text-To-Speech with Grapheme Predictions

Yinghao Aaron Li, Cong Han, Xilin Jiang, Nima Mesgarani

Surface-Sampling Based Objective Quality Assessment Metrics for Mes

Chunyang Fu, Xiang Zhang, Thuoq Nguyen-Canh, Xiaozhong Xu, Ge Li, Shan Liu

Generative Modeling Based Manifold Learning for Adaptive Filtering Guidance

Karim Helwani, Paris Smaragdis, Michael M. Goodwin

Applying Symmetrical Component Transform for Industrial Appliance Classification in Non-Intrusive Load Monitoring

Anthony Faustine, Lucas Pereira

AURA: Privacy-Preserving Augmentation to Improve Test Set Diversity in Speech Enhancement

Xavier Gitiaux, Aditya Khant, Ebrahim Beyrami, Chandan Reddy, Jayant Gupchup, Ross Cutler

SCA: Streaming Cross-Attention Alignment for Echo Cancellation

Yang Liu, Yangyang Shi, Yin Li, Kaustubh Kalgaonkar, Srim Srinivasan, Xin Lei

Quantpipe: Applying Adaptive Post-Training Quantization for Distributed Transformer Pipelines in Dynamic Edge Environments

Haonan Wang, Connor Imes, Souvik Kundu, Peter A. Beerel, Stephen P. Crago, John Paul Walters

Quantitative Evidence on Overlooked Aspects of Enrollment Speaker Embeddings for Target Speaker Separation

Xiaoyu Liu, Xu Li, Joan Serrà

Fast Cross-Correlation for TDoA Estimation on Small Aperture Microphone Arrays

François Grondin, Marc-Antoine Maheux, Jean-Samuel Lauzon, Jonathan Vincent, François Michaud

Interpolation Filter Model for Ramanujan Subspace Signals

Pranav Kulkarni, P. P. Vaidyanathan

Speaker Change Detection for Transformer Transducer ASR

Jian Wu, Zhuo Chen, Min Hu, Xiong Xiao, Jinyu Li

Wiener Filtering Without Covariance Matrix Inversion

Pranav U. Damale, Edwin K. P. Chong, Louis L. Scharf

Cough Detection Using Millimeter-Wave Fmcw Radar

Kawon Han, Songcheol Hong

RNN-Based Step-Size Estimation for the RLS Algorithm with Application to Acoustic Echo Cancellation

Ofer Schwartz, Ayal Schwartz
Bridging Speech and Textual Pre-Trained Models with Unsupervised ASR ................................................................. 3721
Jiatong Shi, Chan-Jan Hsu, Holam Chung, Dongji Gao, Paola Garcia, Shinji Watanabe, Ann Lee, Hung-Yi Lee

FED-3DA: A Dynamic and Personalized Federated Learning Framework .............................................................. 3726
Hui Wang, Jie Sun, Tianyu Wo, Xudong Liu

A New Approach to Extract Fetal Electrocardiogram Using Affine Combination of Adaptive Filters .......... 3731
Yu Xuan, Xiangyu Zhang, Shuyue Stella Li, Zihan Shen, Xin Xie, Leibny Paola Garcia, Roberto Togneri

Performing Neural Architecture Search Without Gradients .................................................................................... 3736
Pavel Rumiantsev, Mark Coates

Accelerating RNN-T Training and Inference Using CTC Guidance ................................................................. 3741
Yongqiang Wang, Zhewai Chen, Chengjian Zheng, Yu Zhang, Wei Han, Parisa Haghighi

GANstrument: Adversarial Instrument Sound Synthesis with Pitch-Invariant Instance Conditioning ........... 3746
Gaku Narita, Junichi Shimizu, Taketo Akama

Virtuoso: Massive Multilingual Speech-Text Joint Semi-Supervised Learning for Text-to-Speech .......... 3751
Takauki Saeki, Heiga Zen, Zhewai Chen, Nobuyuki Morioka, Gary Wang, Yu Zhang, Ankur Bapna, Andrew Rosenberg, Bhuvana Ramabhadran

Deep Fusion of Multi-Object Densities Using Transformer ................................................................................. 3756
Lechi Li, Chen Dai, Yuxuan Xia, Lennart Svensson

DASA: Difficulty-Aware Semantic Augmentation for Speaker Verification ...................................................... 3761
Yuanyuan Wang, Yang Zhang, Zhiyong Wu, Zhihan Yang, Tao Wei, Kun Zou, Helen Meng

Frequency Reciprocal Action and Fusion for Single Image Super-Resolution .................................................. 3766
Shuting Dong, Feng Lu, Chun Yuan

Augmenting Transformer-Transducer Based Speaker Change Detection with Token-Level Training Loss .................................................................................. 3771
Guanlong Zhao, Quan Wang, Han Lu, Yiling Huang, Ignacio Lopez Moreno

Parasympathetic-Sympathetic Causal Interactions and Perceived Workload for Varying Difficulty Affective Computing Tasks .................................................................................................................. 3776
Pravallika Lavanuru, Savon Pratisher, Karuna P. Sahoo, Mrinal Acharya, Sreejith S, Nirmalya Ghosh, Amit Patra

Optimizing Quantum Federated Learning Based on Federated Quantum Natural Gradient Descent .......... 3781
Jun Qi, Xiao-Lei Zhang, Javier Tejedar

DL-NET: Dilation Location Network for Temporal Action Detection .............................................................. 3786
Dianlong You, Houlin Wang, Bingxin Liu, Yang Yu, Zhiming Li

TAPLoss: A Temporal Acoustic Parameter Loss for Speech Enhancement .................................................... 3791
Yunyang Zeng, Joseph Konan, Shuo Han, David Bick, Muqiao Yang, Anurag Kumar, Shinji Watanabe, Bhiksha Raj

Adaptive Submanifold-Preserving Sparse Regression for Feature Selection and Multiclass Classification .......................................................................................................................................................... 3796
Rui Xu, Xin Liang
Improving CTC-Based ASR Models with Gated Interlayer Collaboration .................................................................................. 3801
Yuting Yang, Yuke Li, Binbin Du

Reducing Language Confusion for Code-Switching Speech Recognition with Token-Level Language Diarization ................................................................. 3806
Hexin Liu, Haihua Xu, Leibny Paola Garcia, Andy W. H. Khong, Yi He, Sanjeev Khudanpur

Training Stronger Spiking Neural Networks with Biomimetic Adaptive Internal Association Neurons .................................................................................. 3811
Haibo Shen, Yihao Luo, Xiang Cao, Liangqi Zhang, Juyu Xiao, Tianjiang Wang

ISmallNet: Densely Nested Network with Label Decoupling for Infrared Small Target Detection ................................................................. 3816
Zhiheng Hu, Yongzhen Wang, Peng Li, Jie Qin, Haoran Xie, Mingqiang Wei

Hierarchical Network with Decoupled Knowledge Distillation for Speech Emotion Recognition .................................................. 3821
Ziping Zhao, Huan Wang, Haishuai Wang, Björn Schuller

UAV Remote Sensing Image Dehazing Based on Multi-Dimensional Saliency Awareness Unequal Network ..................................................................... 3826
Ruohui Zheng, Libao Zhang

A Two-Branch Network for Video Anomaly Detection with Spatio-Temporal Feature Learning ................................................................. 3831
Guoqiu Li, Shengjie Chen, Yiju Yu, Zhenhua Guo

Zero-Shot Personalized Lip-To-Speech Synthesis with Face Image Based Voice Control ................................................................. 3836
Zheng-Yan Sheng, Yang Ai, Zhen-Hua Ling

Speech Reconstruction from Silent Tongue and Lip Articulation by Pseudo Target Generation and Domain Adversarial Training ................................................................. 3841
Rui-Chen Zheng, Yang Ai, Zhen-Hua Ling

Cross-Domain Object Classification Via Successive Subspace Alignment .................................................................................. 3846
Kecheng Chen, Haoliang Li, Hong Yan

Prosody-Aware SpeechT5 for Expressive Neural TTS .................................................................................................................. 3851
Yan Deng, Long Zhou, Yuanhao Yi, Shujie Liu, Lei He

DEHRFormer: Real-Time Transformer for Depth Estimation and Haze Removal from Varicolored Haze Scenes ................................................................. 3856
Sixiang Chen, Tian Ye, Jun Shi, Yun Liu, Jingxia Jiang, Erkang Chen, Peng Chen

Dual-Path Cross-Modal Attention for Better Audio-Visual Speech Extraction .................................................................................. 3861
Zhongweiyang Xu, Xulin Fan, Mark Hasegawa-Johnson

MSP-Former: Multi-Scale Projection Transformer for Single Image Desnowing .................................................................................. 3866
Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Jingxia Jiang, Erkang Chen, Peng Chen

Evaluating Variants of Wav2vec 2.0 on Affective Vocal Burst Tasks .................................................................................. 3871
Bagus Tris Atmaja, Akira Sasou

Robust Time Series Recovery and Classification Using Test-Time Noise Simulator Networks ................................................................. 3876
Eun Som Jeon, Suhas Lohit, Rushil Anirudh, Pavan Turaغا

Core: Transferable Long-Range Time Series Forecasting Enhanced by Covariates-Guided Representation ........................................................................ 3881
Xin-Yi Li, Pet-Nan Zhong, Di Chen, Yu-Bin Yang
PFT-SSR: Parallax Fusion Transformer for Stereo Image Super-Resolution ................................................. 3886
Hansheng Guo, Juncheng Li, Guangwei Gao, Zhi Li, Tieyong Zeng

Neural Optimization of Geometry and Fixed Beamformer for Linear Microphone Arrays .......................... 3891
Longfei Yan, Weilong Huang, W. Bastiaan Kleijn, Thushara D. Abhayapala

ECGT2T: Towards Synthesizing Twelve-Lead Electrocardiograms from Two Asynchronous Leads .......... 3896
Yong-Yeoon Jo, Young Sang Choi, Jong-Hwan Jang, Joon-Myoung Kwon

Auxiliary Pooling Layer for Spoken Language Understanding .......................................................... 3901
Yukun Ma, Trung Hieu Nguyen, Jinjie Ni, Wen Wang, Qian Chen, Chong Zhang, Bin Ma

Bilateral Coarse-To-Fine Network for Point Cloud Completion ...................................................... 3906
Tran Thanh Phong Nguyen, Son Lam Phung, Vinod Gopaladasani, Jane Whitelaw

SC-Net: Salient Point and Curvature Based Adversarial Point Cloud Generation Network ....................... 3911
Zihao Zhang, Nan Sang, Xupeng Wang, Mumuxin Cai

Light Projection-Based Physical-World Vanishing Attack Against Car Detection .................................. 3916
Huixiang Wen, Shan Chang, Luo Zhou

A Dual-Branch Adaptive Distribution Fusion Framework for Real-World Facial Expression Recognition .......................................................... 3921
Shu Liu, Yan Xu, Tongming Wan, Xiaoyan Kui

Face Recognition on Point Cloud with Cgan-Top for Denoising .................................................... 3926
Junyu Liu, Jianfeng Ren, Hongliang Sun, Xudong Jiang

Self-Paced Partial Domain-Aware Learning for Face Anti-Spoofing .................................................. 3931
Zhiyi Chen, Yao Lu, Xinze Deng, Jia Meng, Shengchuan Zhang, Liujuan Cao

Adaptive Knowledge Distillation Between Text and Speech Pre-Trained Models .................................... 3936
Jinjie Ni, Yukun Ma, Wen Wang, Qian Chen, Dianwen Ng, Han Lei, Trung Hieu Nguyen, Chong Zhang, Bin Ma, Erik Cambria

A Simple Yet Effective Approach to Structured Knowledge Distillation ................................................. 3941
Wenye Lin, Yangming Li, Lemao Liu, Shuming Shi, Hai-Tao Zheng

SQA: Strong Guidance Query with Self-Selected Attention for Human-Object Interaction Detection ........ 3946
Feng Zhang, Liu Sheng, Bingnan Guo, Ruixiang Chen, Junhao Chen

Picking the Underused Heads: A Network Pruning Perspective of Attention Head Selection for Fusing Dialogue Coreference Information .......................................................... 3951
Zhengyuan Liu, Nancy F. Chen

LIMI-VC: A Light Weight Voice Conversion Model with Mutual Information Disentanglement .................. 3956
Liangjie Huang, Tian Yuan, Yuming Liang, Zeyu Chen, Can Wen, Yanlu Xie, Jinsong Zhang, Dengfeng Ke

FCIR: Rethink Aerial Image Super Resolution with Fourier Analysis .................................................. 3961
Yan Zhang, Pengcheng Zheng, Jianan Jiang, Pu Xiao, Xinbo Gao

Selective Film Conditioning with CTC-Based ASR Probability for Speech Enhancement .......................... 3966
Da-Hee Yang, Joon-Hyuk Chang
Improved Deep Speaker Localization and Tracking: Revised Training Paradigm and Controlled Latency .......................................................... 3971
  Alexander Bohlender, Liesbeth Roelens, Nilesh Madhu

Multi-Layer Feature Division Transferable Adversarial Attack .......................................................... 3976
  Zikang Jin, Changchun Yin, Piji Li, Lu Zhou, Liming Fang, Xiangmao Chang, Zhe Liu

Improving Noisy Student Training on Non-Target Domain Data for Automatic Speech Recognition .......... 3981
  Yu Chen, Wen Ding, Junjie Lai

Clicker: Attention-Based Cross-Lingual Commonsense Knowledge Transfer ........................................ 3986
  Ruolin Su, Zhongkai Sun, Sixing Lu, Chengyuan Ma, Chenlei Guo

Efficient Data Loading with Quantum Autoencoder ............................................................................. 3991
  Siang-Ruei Wu, Chun-Tse Li, Hao-Chung Cheng

Spteae: A Soft Prompt Transfer Model for Zero-Shot Cross-Lingual Event Argument Extraction ............ 3996
  Huipeng Ma, Qiu Tang, Ni Zhang, Rui Xu, Yanhua Shao, Wei Yan, Yaojun Wang

Strategies for Enhanced Signal Modulation Classifications Under Unknown Symbol Rates and Noise Conditions .......................................................... 4001
  Ruixuan Wang, Yue Qi, Mojtaba Vaezi, Xin Jiao, Moeness Amin

Knowledge Distillation with Active Exploration and Self-Attention Based Inter-Class Variation Transfer for Image Segmentation ............................................. 4006
  Yifan Zhang, Shaofie Li, Xuan Yang

Oct Image Blind Despeckling Based on Gradient Guided Filter with Speckle Statistical Prior ............... 4011
  Sanqian Li, Muxing Xiong, Bing Yang, Xiaqing Zhang, Risa Higashita, Jiang Liu

Convolution-Based Channel-Frequency Attention for Text-Independent Speaker Verification .............. 4016
  Jingyu Li, Yusheng Tian, Tan Lee

Streaming Multi-Channel Speech Separation with Online Time-Domain Generalized Wiener Filter .... 4021
  Yi Luo

Invariant Adversarial Imitation Learning from Visual Inputs ............................................................... 4026
  Haoran Zhang, Yinghong Tian, Liang Yuan, Yue Lu

A Distributed Adaptive Algorithm for Non-Smooth Spatial Filtering Problems .................................... 4031
  Charles Hovine, Alexander Bertrand

Performance of Social Machine Learning Under Limited Data ............................................................ 4036
  Ping Hu, Virginia Bordignon, Mert Kayaalp, Ali H. Sayed

Single-Branch Network for Multimodal Training .................................................................................... 4041
  Muhammad Saad Saeed, Shah Nawaz, Muhammad Haris Khan, Muhammad Zaigham Zaheer,
  Karthik Nandakumar, Muhammad Haroon Yousaf, Arif Mahmood

Structure-Aware Multi-Feature Co-Learning for Dual Branch Face Super Resolution ......................... 4046
  Kangli Zeng, Zhongyuan Wang, Tao Lu, Jianyu Chen

Measuring Deviation from Stochasticity in Time-Series Using Autoencoder Based Time-Invariant Representation: Application to Black Hole Data ........................................... 4051
  Chakka Sai Pradeep, Neelam Sinha, Banibrata Mukhopadhay
Multi-Modal Domain Generalization for Cross-Scene Hyperspectral Image Classification
Yuxiang Zhang, Mengmeng Zhang, Wei Li, Ran Tao

Multi-Speaker Expressive Speech Synthesis Via Multiple Factors Decoupling
Xinfa Zhu, Yi Lei, Kun Song, Yongmao Zhang, Tao Li, Lei Xie

Multi-Speaker and Wide-Band Simulated Conversations as Training Data for End-to-End Neural Diarization
Federico Landini, Mireia Diez, Alicia Lozano-Diez, Lukáš Burget

Spatio-Temporal Structure Consistency for Semi-Supervised Medical Image Classification
Wentao Lei, Lei Liu, Li Liu

Hypothesis Test for Leakage Detection in Water Pipelines with High-Dimensional Sensor Signals
Liusha Yang, Matthew R. McKay, Xun Wang

Procter: Pronunciation-Aware Contextual Adapter for Personalized Speech Recognition in Neural Transducers
Rahul Pandey, Roger Ren, Qi Luo, Jing Liu, Ariya Rastrow, Ankur Gandhe, Denis Filimonov, Grant Strimel, Andreas Stolcke, Ivan Bulyko

Bebert: Efficient and Robust Binary Ensemble Bert
Jiayi Tian, Chao Fang, Haoran Wang, Zhongfeng Wang

End-To-End Unsupervised Sketch to Image Generation
Xingming Lv, Lei Wu, Zhenwei Cheng, Xiangyu Meng

Regularized Deep Generative Model Learning for Real-Time Massive MIMO Channel Tracking
Lixiang Lian, Ben Wang

Attention Localness in Shared Encoder-Decoder Model for Text Summarization
Li Huang, Hongmei Wu, Qiang Gao, Guisong Liu

PU-Edgeformer: Edge Transformer for Dense Prediction in Point Cloud Upsampling
Dohoon Kim, Minwoo Shin, Joonki Paik

LSSED: A Robust Segmentation Network for Inflamed Appendix from CT Images
Wing W Y. Ng, Peixin Zheng, Ting Wang, Jianjun Zhang, Yinhao Liang, Hui Zhou, Dan Liang, Guangming Li, Xinhua Wei

Tree-Like Interaction Learning for Bundle Recommendation
Haole Ke, Lin Li, Peipei Wang, Jingling Yuan, Xiaohui Tao

IfUNet++: Iterative Feedback UNet++ for Infrared Small Target Detection
Zhangying Weng, Peng Li, Xin Zhuang, Xuefeng Yan, Lina Gong, Haoran Xie, Mingqiang Wei

MmSense: Detecting Concealed Weapons with a Miniature Radar Sensor
Kevin Mitchell, Khaled Kassem, Chaitanya Kaul, Valentin Kapitany, Philip Binner, Andrew Ramsay, Daniele Faccio, Roderick Murray-Smith

Improving Learning Objectives for Speaker Verification from the Perspective of Score Comparison
Min Hyun Han, Sung Hwan Mun, Minchan Kim, Myeonghun Jeong, Sung Hwan Ahn, Nam Soo Kim

Adapter Tuning with Task-Aware Attention Mechanism
Jinliang Lu, Feihu Jin, Jiajun Zhang
Twitter Stance Detection Via Neural Production Systems .................................................. 4141
Bowen Zhang, Daijun Ding, Guangning Xu, Jinjin Guo, Zhichao Huang, Xu Huang

Zero-Shot Sound Event Classification Using a Sound Attribute Vector with Global and Local
Feature Learning .............................................................................................................. 4146
Yi-Han Lin, Xunquan Chen, Ryoichi Takashima, Tetsuya Takiguchi

Optimization for Robustness Evaluation Beyond lp Metrics ........................................... 4151
Hengyue Liang, Buyun Liang, Ying Cui, Tim Mitchell, Ju Sun

Semi-Supervised Contrastive Learning with Soft Mask Attention for Facial Action Unit Detection ........ 4156
Zhongling Liu, Ruijie Liu, Ziqiang Shi, Liu Liu, Xiaoyu Mi, Kentaro Murase

Speech Enhancement with Intelligent Neural Homomorphic Synthesis ................................. 4161
Shulin He, Wei Rao, Jinjiang Liu, Jun Chen, Yukai Ju, Xueliang Zhang, Yannan Wang, Shidong Shang

Deepspace: Dynamic Spatial and Source CUE Based Source Separation for Dialog Enhancement .... 4166
Aaron Master, Lie Lu, Jonas Samuelsson, Heidi-Maria Lehtonen, Scott Norcross, Nathan
Svedlow, Audrey Howard

Multiple Contrastive Learning for Multimodal Sentiment Analysis ....................................... 4171
Xiaocui Yang, Shi Feng, Daling Wang, Pengfei Hong, Soujanya Poria

Towards Real-Time Person Search with Invariant Feature Learning ..................................... 4176
Chengyou Jia, Minnan Luo, Zhuohang Dang, Xiaojun Chang, Qinghua Zheng

Unsupervised Fine-Tuning Data Selection for ASR Using Self-Supervised Speech Models ........... 4181
Reem Gody, David Harwath

PCF: ECAPA-TDNN with Progressive Channel Fusion for Speaker Verification ....................... 4186
Zhenduo Zhao, Zhuo Li, Wenchao Wang, Pengyuan Zhang

Short-Segment Speaker Verification Using ECAPA-TDNN with Multi-Resolution Encoder ............ 4191
Sangwook Han, Youngdo Ahn, Kyeongmuk Kang, Jong Won Shin

AST-SED: An Effective Sound Event Detection Method Based on Audio Spectrogram Transformer .... 4196
Kang Li, Yan Song, Li-Rong Dai, Ian McLoughlin, Xin Fang, Lin Liu

Ψ-Net: Point Structural Information Network for No-Reference Point Cloud Quality Assessment ....... 4201
Jian Xiong, Sifan Wu, Wang Luo, Jinli Suo, Hao Gao

Inter-Subnet: Speech Enhancement with Subband Interaction ............................................. 4206
Jun Chen, Wei Rao, Zilin Wang, Jiuxin Lin, Zhiyong Wu, Yannan Wang, Shidong Shang, Helen
Meng

Audio Cross Verification Using Dual Alignment Likelihood Ratio Test ..................................... 4211
Heidi Lei, Arm Wonghirundacha, Irmak Bukey, T J Tsai

Gender-Cartoon: Image Cartoonization Method Based on Gender Classification ....................... 4216
Long Feng, Guohua Geng, Chen Guo, Longquan Yan, Xingrui Ma, Zhan Li, Kang Li

CADET: Control-Aware Dynamic Edge Computing for Real-Time Target Tracking in UAV
Systems ........................................................................................................................... 4221
Luis Felipe Florenzan Reyes, Francesco Smarra, Alessandro D’Innocenzo, Marco Levorato
Select the Best: Enhancing Graph Representation with Adaptive Negative Sample Selection .......................... 4226 
Xiangping Zheng, Xun Liang, Bo Wu

Regression to Classification: Waveform Encoding for Neural Field-Based Audio Signal Representation ..................................................................................................................... 4231 
Taesoo Kim, Daniel Rho, Gahui Lee, Jaehan Park, Jong Hwan Ko

Exploiting Virtual Array Diversity for Accurate Radar Detection ................................................................................................................................. 4236 
Junfeng Guan, Sohrab Madani, Waleed Ahmed, Samah Hussein, Saurabh Gupta, Haitham Hassanieh

Nowcasting of Extreme Precipitation Using Deep Generative Models .................................................................................................................. 4241 
Haoran Bi, Maksym Kyryliuk, Zhiyi Wang, Cristian Meo, Yanbo Wang, Ruben Imhoff, Remko Uijlenhoet, Justin Dauwels

Deep Spatio-Temporal Multiplex Graph Learning for Cardiac Imaging Classification ................................................................. 4246 
Jaume Banus, Augustin Ogier, Roger Hullin, Philippe Meyer, Raud B. Van Heeswijk, Jonas Richiardi

JNDMix: Jnd-Based Data Augmentation for No-Reference Image Quality Assessment ................................................................. 4251 
Jiamu Sheng, Jiayuan Fan, Peng Ye, Jianjian Cao

Deep Generative Fixed-Filter Active Noise Control ................................................................................................................................. 4256 
Zhengding Luo, Dongyuan Shi, Xiaoyi Shen, Junwei Ji, Woon-Seng Gan

Leveraging Neural Koopman Operators to Learn Continuous Representations of Dynamical Systems from Scarce Data .................................................................................. 4261 
Anthony Frion, Lucas Drumetz, Mauro Dalla Mura, Guillaume Tochon, Abdeldjalil Aissa-El-Bey

Decoding Musical Pitch from Human Brain Activity with Automatic Voxel-Wise Whole-Brain FMRI Feature Selection ........................................ 4266 
Vincent K. M. Cheung, Yueh-Po Peng, Jing-Hua Lin, Li Su

RDO Candidate Selection for Maximizing Coding Efficiency in a Practical HEVC Encoder ................................. 4271 
Joose Sainio, Alexandre Mercat, Jarno Vanne

Residual Hybrid Attention Network for Compression Artifact Reduction ................................................................................................................................. 4276 
Bingchun Luo, Wei Yu

Robust Binary Component Decompositions ................................................................................................................................. 4281 
Christos Kolomvakis, Nicolas Gillis

A Computationally Efficient Algorithm for Distributed Adaptive Signal Fusion Based on Fractional Programs ........................................................................... 4286 
Cem Ataş Musluoglu, Alexander Bertrand

SL-MoE: A Two-Stage Mixture-of-Experts Sequence Learning Framework for Forecasting Rapid Intensification of Tropical Cyclone ........................................................................... 4291 
Jian Xu, Yang Lei, Guangqi Zhu, Yunling Feng, Bo Xiao, Qifeng Qian, Yajing Xu

Personalized Federated Learning on Long-Tailed Data Via Adversarial Feature Augmentation ........................................................................... 4296 
Yang Lu, Pixin Qian, Gang Huang, Hanzi Wang

Autotts: End-To-End Text-To-Speech Synthesis Through Differentiable Duration Modeling ........................................................................... 4301 
Bac Nguyen, Fabien Cardinaux, Stefan Uhlich
Semi-Supervised Learning with Per-Class Adaptive Confidence Scores for Acoustic Environment Classification with Imbalanced Data ........................................................................................................ 4306
   Luan Vinicius Fiorio, Boris Karanov, Johan David, Wim Van Houtum, Frans Widdershoven, Ronald M. Aarts

Central Nodes Detection from Partially Observed Graph Signals ........................................................................................................ 4311
   Yiran He, Hoi-To Wai

Compressive Channel Estimation for IRS-Aided Millimeter-Wave Systems Via Two-Stage Lamp Network ........................................................................................................................................ 4316
   Wen-Chiao Tsai, Chi-Wei Chen, An-Yeu Andy Wu

A Discriminative Multi-Channel Noise Feature Representation Model for Image Manipulation Localization ........................................................................................................................................ 4321
   Yang Zhou, Hongxia Wang, Qiang Zeng, Rui Zhang, Sijiang Meng

A Practical Distributed Active Noise Control Algorithm Overcoming Communication Restrictions ........................................................................................................ 4326
   Junwei Ji, Dongyuan Shi, Zhengding Luo, Xiaoyi Shen, Woon-Seng Gan

Group-Wise Co-Salient Object Detection with Siamese Transformers Via Brownian Distance Covariance Matching ........................................................................................................................................ 4331
   Yang Wu, Hao Zhang, Lingyan Liang, Yaqian Zhao, Kaihua Zhang

A Content-Based Multi-Scale Network for Single Image Super-Resolution ........................................................................................................................................ 4336
   Jiahuan Ji, Baojiang Zhong, Kai-Kuang Mu

Cross-Modal Fusion Techniques for Utterance-Level Emotion Recognition from Text and Speech ........................................................................................................................................ 4341
   Jiachen Luo, Huy Phan, Joshua Reiss

Lightvessel: Exploring Lightweight Coronary Artery Vessel Segmentation Via Similarity Knowledge Distillation ........................................................................................................................................ 4346
   Hao Dang, Yuekai Zhang, Xingqun Qi, Wanting Zhou, Muyi Sun

Enhancing Robustness and Imperceptibility of Blind Watermarking with Improved Message Processor ........................................................................................................................................ 4351
   Yufeng Wu, Baowei Wang, Changyu Dai, Yi Yuan, Bin Li, Weiqian Zheng, Hao Wu

Prior-Enhanced Temporal Action Localization Using Subject-Aware Spatial Attention ........................................................................................................................................ 4356
   Yifan Liu, Youbao Tang, Ning Zhang, Ruei-Sung Lin, Haoqian Wang

Dynamic Independent Component Extraction with Blending Mixing Vector: Lower Bound on Mean Interference-To-Signal Ratio ........................................................................................................................................ 4361
   Jaroslav Cmejla, Zbynek Koldovský, Václav Kautský, Tülay Adali

Integrating Syntactic and Semantic Knowledge in AMR Parsing with Heterogeneous Graph Attention Network ........................................................................................................................................ 4366
   Yikemaiti Sataer, Chuanqi Shi, Miao Gao, Yunlong Fan, Bin Li, Zhiqiang Gao

Contrastive Learning at the Relation and Event Level for Rumor Detection ........................................................................................................................................ 4371
   Yingrui Xu, Jingyuan Hu, Jingguo Ge, Yulei Wu, Tong Li, Hui Li

Multimodal Propaganda Detection Via Anti-Persuasion Prompt Enhanced Contrastive Learning ........................................................................................................................................ 4376
   Jian Cui, Lin Li, Xin Zhang, Jingling Yuan

Multitask Detection of Speaker Changes, Overlapping Speech and Voice Activity Using Wav2vec2.0 ........................................................................................................................................ 4381
   Marie Kunešová, Zbyněk Zajíc
Blood Oxygen Saturation Estimation from Facial Video Via DC and AC Components of Spatio-Temporal Map
Yusuke Akamatsu, Yoshifumi Onishi, Hitoshi Imaoka

Ensemble of Deep Neural Network Models for MOS Prediction
Marie Kunešová, Jindřich Matoušek, Jan Lehecka, Jan Švec, Josef Michálek, Daniel Tihelka, Martin Bulín, Zdenek Hanzlicek, Markéta Rezácková

Using Auxiliary Tasks in Multimodal Fusion of Wav2vec 2.0 and Bert for Multimodal Emotion Recognition
Dekai Sun, Yan-Cheng He, Ji-Qing Han

A2S-NAS: Asymmetric Spectral-Spatial Neural Architecture Search for Hyperspectral Image Classification
Lin Zhan, Jiyuan Fan, Peng Ye, Jianjian Cao

NCL: Textual Backdoor Defense Using Noise-Augmented Contrastive Learning
Shengfang Zhai, Qingni Shen, Xiaoyi Chen, Weilong Wang, Cong Li, Yuejian Fang, Zhonghai Wu

Prefallkd: Pre-Impact Fall Detection Via CNN-ViT Knowledge Distillation
Tin-Han Chi, Kai-Chun Liu, Chia-Yeh Hsieh, Yu Tsao, Chia-Tai Chan

Towards Reducing Patient Effort for the Automatic Prediction of Speech Intelligibility in Head and Neck Cancers
Sebastião Quintas, Alberto Abad, Julie Mauclair, Virginie Woisard, Julien Pinquier

Integrated Sensing and Full-Duplex Communication: Joint Transceiver Beamforming and Power Allocation
Zhenyao He, Wei Xu, Hong Shen, Derrick Wing Kwan Ng, Yonina C. Eldar, Xiao-Hu You

Ripple Sparse Self-Attention for Monaural Speech Enhancement
Qiquan Zhang, Hong-Xu Zhu, Qi Song, Xinyuan Qian, Zhaoheng Ni, Haizhou Li

Spectral Super-Resolution on the Unit Circle Via Gradient Descent
Xun-meng Wu, Zai Yang, Jian-Feng Cai, Zongben Xu

Interweaved Graph and Attention Network for 3D Human Pose Estimation
Ti Wang, Hong Liu, Runwei Ding, Wenhao Li, Yingxuan You, Xia Li

Scoreformer: Score Fusion-Based Transformers for Weakly-Supervised Violence Detection
Yang Xiao, Liejun Wang, Tongguan Wang, Huicheng Lai

Skillnet-NLG: General-Purpose Natural Language Generation with a Sparsely Activated Approach
Junwei Liao, Duyu Tang, Fan Zhang, Shuming Shi

Semantic Memory Guided Image Representation for Polyp Segmentation
Zijin Yin, Runpu Wei, Kongming Liang, Yiyang Lin, Wei Liu, Zhanyu Ma, Min Min, Jun Guo

OAFomer: Learning Occlusion Distinguishable Feature for Amodal Instance Segmentation
Zhixuan Li, Ruohua Shi, Tiejun Huang, Tingting Jiang

Rethink Long-Tailed Recognition with Vision Transforms
Zhengzhuo Xu, Shuo Yang, Xingjun Wang, Chun Yuan
Multi-Task Transformer with Relation-Attention and Type-Attention for Named Entity Recognition

Ying Mo, Hongyin Tang, Jiahao Liu, Qifan Wang, Zenglin Xu, Jingang Wang, Wei Wu, Zhoujun Li

Regularized Neural Detection for Millimeter Wave Massive Mimo Communication Systems with One-Bit Adcs

Aditya Sant, Bhaskar D. Rao

Decoupled Non-Parametric Knowledge Distillation for end-to-End Speech Translation

Hao Zhang, Nianwen Si, Yaqi Chen, Wenlin Zhang, Xukui Yang, Dan Qu, Zhen Li

Smoothing Complex-Valued Signals on Graphs with Monte-Carlo

Hugo Jaquard, Michaël Fanuel, Pierre-Olivier Amblard, Rémi Bardenet, Simon Barthelmé, Nicolas Tremblay

Constrained Dynamical Neural ODE for Time Series Modelling: A Case Study on Continuous Emotion Prediction

Ting Dang, Antoni Dimitriadis, Jingyao Wu, Vidhyasaharan Sethu, Eliathamby Ambikairajah

Parameter Efficient Transfer Learning for Various Speech Processing Tasks

Shinta Otake, Rei Kawakami, Nakamasa Inoue

Distortion-Aware Convolutional Neural Network-Based Interpolation Filter for AVS3

Ying Zhang, Liang Wen, Lisheng Wang, Yinji Piao, Weijing Shi, Kwang Pyo Choi

Yolo-Based Lightweight Object Detection with Structure Simplification and Attention Enhancement

Shuqi Sun, Xiaohui Yang, Jingliang Peng

UFO2: A Unified Pre-Training Framework for Online and Offline Speech Recognition

Li Fu, Siqi Li, Qingtao Li, Liping Deng, Fangzhu Li, Lu Fan, Meng Chen, Xiaodong He

Lightweight and High-Fidelity End-to-End Text-to-Speech with Multi-Band Generation and Inverse Short-Time Fourier Transform

Masaya Kawamura, Yuma Shirahata, Ryuichi Yamamoto, Kentaro Tachibana

Self-Adaptive Incremental Machine Speech Chain for Lombard TTS with High-Granularity ASR Feedback in Dynamic Noise Condition

Sashi Novitasari, Sakriani Sakti, Satoshi Nakamura

Cooperative Five Degrees of Freedom Motion Estimation for a Swarm of Autonomous Vehicles

Nikos Piperigkos, Aris S. Lalos, Kostas Berberidis, Christos Anagnostopoulos

LongFNT: Long-Form Speech Recognition with Factorized Neural Transducer

Xun Gong, Yu Wu, Jinyu Li, Shujie Liu, Rui Zhao, Xie Chen, Yamin Qian

RL-IFF: Indoor Localization Via Reinforcement Learning-Based Information Fusion

Mohamad Salimibeni, Arash Mohammadi

Feature Space Recovery for Incomplete Multi-View Clustering

Zhen Long, Ce Zhu, Pierre Comon, Yipeng Liu

On the Joint Estimation of Phase Noise and time-Varying Channels for OFDM Under High-Mobility Conditions

Francesco Linsalata, Nassar Ksairi
6G Integrated Sensing and Communication - Sensing Assisted Environmental Reconstruction and Communication ................................................................. 4546

Zhi Zhou, Xianjin Li, Jia He, Xiaoyan Bi, Yan Chen, Guangjian Wang, Peiying Zhu

A Novel Approach Based on Voronoi Cells to Classify Spectrogram Zeros of Multicomponent Signals ................................................................. 4551

N. Laurent, S. Meignen, M. A. Colominas, J. M. Miramont, F. Auger

DTTR: Detecting Text with Transformers ............................................................................. 4556

Jing Yang, Zhiqiang You, Zhiwei Zhong, Peng Liu, Langqi Mei, Shenguang Huang

Learning to Explain: A Gradient-Based Attribution Method for Interpreting Super-Resolution Networks ................................................................. 4561

Anni Yu, Yu-Bin Yang

Estimation of Cardiac Fibre Direction Based on Activation Maps ................................................. 4566

Johannes W. De Vries, Miao Sun, Natasja M. S. De Groot, Richard C. Hendriks

Bimodal Fusion Network for Basic Taste Sensation Recognition from Electroencephalography and Electromyography ......................................................... 4571

Han Gao, Shuo Zhao, Huiyan Li, Li Liu, You Wang, Ruifen Hu, Jin Zhang, Guang Li

Pushing the Limits of Self-Supervised Speaker Verification Using Regularized Distillation Framework ............................................................................. 4576

Yafeng Chen, Siqi Zheng, Hui Wang, Luyao Cheng, Qian Chen

A Multi-Scale Feature Aggregation Based Lightweight Network for Audio-Visual Speech Enhancement ................................................................................. 4581

Haitao Xu, Liangfa Wei, Jie Zhang, Jianming Yang, Yannan Wang, Tian Gao, Xin Fang, Lirong Dai

Multi-Lingual Pronunciation Assessment with Unified Phoneme Set and Language-Specific Embeddings ........................................................................ 4586

Binghuai Lin, Liyuan Wang

Robust Multi-Modal Speech Emotion Recognition with ASR Error Adaptation ................................ 4591

Binghuai Lin, Liyuan Wang

Matrix Recovery Using Deep Generative Priors with Low-Rank Deviations ................................ 4596

Pengbin Yu, Jianjun Wang, Chen Xu

Spectro-Temporal Post-Filtering Via Short-Time Target Cancellation for Directional Speech Enhancement in a Dual-Microphone Hearing AID ........................................................................ 4601

Marcos A. Cantu, Volker Hohmann

OTW: Optimal Transport Warping for Time Series ...................................................................... 4606

Fabian Latorre, Chenghao Liu, Doyen Sahoo, Steven C. H. Hoi

Narrow Down Before Selection: A Dynamic Exclusion Model for Multiple-Choice QA .............. 4611

Xiyan Liu, Yidong Shi, Ruiyang Liu, Ge Bai, Yanyi Chen

Voice Adversarial Network Using Feature Specific Loss Function Based Self-Attentive Generative Adversarial Network ................................................................................. 4616

Sandipan Dhar, Padmanabha Banerjee, Nanda Dulal Jana, Swagatam Das

Improving Text-Audio Retrieval by Text-Aware Attention Pooling and Prior Matrix Revised Loss ................................................................................. 4621

Yifei Xin, Dongchao Yang, Yuexian Zou
OPT: One-Shot Pose-Controllable Talking Head Generation ........................................... 4626
Jin Liu, Xi Wang, Xiaomeng Fu, Yesheng Chai, Cai Yu, Jiao Dai, Jizhong Han

Exploration of Language Dependency for Japanese Self-Supervised Speech Representation Models .......... 4631
Takanori Ashihara, Takaumi Mori, Kohei Matsuura, Tomohiro Tanaka

Multi-Modal ASR Error Correction with Joint ASR Error Detection ........................................... 4636
Binghuai Lin, Liyuan Wang

Joint Modeling for ASR Correction and Dialog State Tracking ..................................................... 4641
Deyuan Wang, Tiantian Zhang, Caixia Yuan, Xiaojie Wang

Optimising Different Feature Types for Inpainting-Based Image Representations ............................... 4646
Ferdinand Jost, Vassilen Chizhov, Joachim Weickert

Structured-Anchor Projected Clustering for Hyperspectral Images ................................................ 4651
Guozhu Jiang, Jie Zhang, Yongshan Zhang, Xianwei Jiang, Zhihua Cai

Target Velocity Estimation for Quantization-Based Cooperative MIMO Radar and Communications System ....................................................................................................................... 4656
Zhen Wang, Xuedan Yan, Qian He, Rick S. Blum

Unified Prompt Learning Makes Pre-Trained Language Models Better Few-Shot Learners ...................... 4661
Feihua Jin, Jinliang Lu, Jiajun Zhang

Self-Supervised Facial Action Unit Detection with Region and Relation Learning ............................... 4666
Juan Song, Zhilei Liu

Self-Supervised Learning with Explorative Knowledge Distillation ................................................... 4671
Tongtong Su, Jinsong Zhang, Gang Wang, Xiaoguang Liu

Robust and Globally Sparse Pca Via Majorization-Minimization and Variable Splitting ....................... 4676
Hugo Brehier, Arnaud Breloy, Mohammed Nabil El Korso, Sandeep Kumar

Boosting No-Reference Super-Resolution Image Quality Assessment with Knowledge Distillation and Extension ............................................................................................................. 4681
Haiyu Zhang, Shaolin Su, Yu Zhu, Jinqiu Sun, Yanning Zhang

Cutting Through the Noise: An Empirical Comparison of Psycho-Acoustic and Envelope-based Features for Machinery Fault Detection ................................................................. 4686
Peter Wißbrock, Yvonne Richter, David Pelkmann, Zhao Ren, Gregory Palmer

Backdoor Attack Against Automatic Speaker Verification Models in Federated Learning .................... 4691
Dan Meng, Xue Wang, Jun Wang

Adapting Exploratory Behaviour in Active Inference for Autonomous Driving ..................................... 4696
Sheida Nozari, Ali Krayani, Pablo Marin, Lucio Marcenaro, David Martin, Carlo Regazzoni

Multi-Microphone Speaker Separation by Spatial Regions .................................................................. 4701
Julian Wechsler, Srikanth Raj Chetupalli, Wolfgang Mack, Emanuël A. P. Habets

GTN-Bailando: Genre Consistent long-Term 3D Dance Generation Based on Pre-Trained Genre Token Network .................................................................................................................... 4706
Haolin Zhuang, Shun Lei, Long Xiao, Weiqin Li, Liyang Chen, Sicheng Yang, Zhiyong Wu, Shiyin Kang, Helen Meng
MSNet: A Deep Architecture Using Multi-Sentiment Semantics for Sentiment-Aware Image Style Transfer ................................................................. 4711
Shikun Sun, Jia Jia, Haozhe Wu, Zijie Ye, Junliang Xing

CNN Filter for Super-Resolution with RPR Functionality in VVC .................................................. 4716
Shimin Huang, Cheolkon Jung, Yang Liu, Ming Li

Mimo Radar Transmit Beampattern Matching Via Manifold Optimization ........................................ 4721
Weijie Xiong, Jinfeng Hu, Kai Zhong

Interpretable Nonnegative Incoherent Deep Dictionary Learning for FMRI Data Analysis ..................... 4726
Manuel Morante, Jan Østergaard, Sergios Theodoridis

Multi-Local Attention for Speech-Based Depression Detection............................................................ 4731
Fuxiang Tao, Xuri Ge, Wei Ma, Anna Esposito, Alessandro Vinciarelli

Daily Mental Health Monitoring from Speech: A Real-World Japanese Dataset and Multitask Learning Analysis ........................................................................... 4736
Meishu Song, Andreas Triantafyllopoulos, Zijiang Yang, Hiroki Takeuchi, Toru Nakamura, Akiyoshi Kishi, Tetsuro Ishizawa, Kazuhiro Yoshiiuchi, Xin Jing, Vincent Karas, Zhonghao Zhao, Kun Qian, Bin Hu, Björn W Schuller, Yoshinari Yamamoto

Audio-Driven Talking Head Video Generation with Diffusion Model ..................................................... 4741
Yizhe Zhua, Chihui Zhang, Qiong Liub, Xi Zhoub

Joint Pre-Training with Speech and Bilingual Text for Direct Speech to Speech Translation .................... 4746
Kun Wei, Long Zhou, Ziqiang Zhang, Liping Chen, Shujie Liu, Lei He, Jinyu Li, Furu Wei

Tracking Targets in Hyper-Scale Cameras Using Movement Predication ............................................. 4751
Jiaping Yu, Tongqing Zhou, Zhiping Cai, Wenyuan Kuang

CNN Filter for RPR-Based SR in VVC with Wavelet Decomposition .................................................... 4756
Hui Lan, Cheolkon Jung, Yang Liu, Ming Li

Sketch Less Face Image Retrieval: A New Challenge ............................................................................ 4761
Dawei Dai, Yutang Li, Liang Wang, Shiyu Fu, Shuyin Xia, Guoyin Wang

Comparing Decentralized Gradient Descent Approaches and Guarantees ............................................. 4766
Shana Moothedath, Namrata Vaswani

IQGAN: Robust Quantum Generative Adversarial Network for Image Synthesis on NISQ Devices........ 4771
Cheng Chu, Grant Skipper, Martin Swany, Fan Chen

Fine-Grained Emotional Control of Text-to-Speech: Learning to Rank Inter- and Intra-Class Emotion Intensities ........................................................................ 4776
Shijun Wang, Jón Guðnason, Damian Borth

Sharing Low Rank Conformer Weights for Tiny Always-On Ambient Speech Recognition Models ........ 4781
Steven M. Hernandez, Ding Zhao, Shaojin Ding, Antoine Bruguier, Rohit Prabhavalkar, Tara N. Sainath, Yanzhang He, Ian McGraw

Anomaly Detection in Optical Spectra VIA Joint Optimization ........................................................... 4786
Antonino Maria Rizzo, Luca Magri, Pietro Invernizzi, Enrico Sozio, Stefano Piciaccia, Alberto Tanzi, Stefano Binetti, Cesare Alippi, Giacomo Boracchi

Fast Single-Person 2D Human Pose Estimation Using Multi-Task Convolutional Neural Networks ........ 4791
Christos Papaoanndis, Ioannis Mademlis, Ioannis Pitas
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training Neural Networks for Sequential Change-Point Detection</td>
<td>4796</td>
</tr>
<tr>
<td>Junghwan Lee, Yao Xie, Xiuyuan Cheng</td>
<td></td>
</tr>
<tr>
<td>Extracting the Brain-Like Representation by an Improved Self-Organizing Map for Image Classification</td>
<td>4801</td>
</tr>
<tr>
<td>Jiahong Zhang, Lihong Cao, Moning Zhang, Wenlong Fu</td>
<td></td>
</tr>
<tr>
<td>Improving Spoken Language Identification with Map-Mix</td>
<td>4806</td>
</tr>
<tr>
<td>Shangeth Rajaa, Kriti Anandan, Swaraj Dalmia, Tarun Gupta, Eng Siong Chng</td>
<td></td>
</tr>
<tr>
<td>FEW-Shot Continual Learning with Weight Alignment and Positive Enhancement for Bioacoustic Event Detection</td>
<td>4811</td>
</tr>
<tr>
<td>Xiaoxiao Wu, Dongxing Xu, Haoran Wei, Yanhua Long</td>
<td></td>
</tr>
<tr>
<td>Unifying Speech Enhancement and Separation with Gradient Modulation for End-To-End Noise-Robust Speech Separation</td>
<td>4816</td>
</tr>
<tr>
<td>Yachen Hu, Chen Chen, Heqing Zou, Xionghu Zhong, Eng Siong Chng</td>
<td></td>
</tr>
<tr>
<td>Joint Symbol-Level Precoding and Sub-Block-Level RIS Design for Dual-Function Radar-Communications</td>
<td>4821</td>
</tr>
<tr>
<td>Linlong Wu, Bowen Wang, Ziyang Cheng, Bhavani Shankar M. R, Björn Ottersten</td>
<td></td>
</tr>
<tr>
<td>Speaker Recognition with Two-Step Multi-Modal Deep Cleansing</td>
<td>4826</td>
</tr>
<tr>
<td>Ruijie Tao, Kong Aik Lee, Zhan Shi, Haizhou Li</td>
<td></td>
</tr>
<tr>
<td>Conditioning and Sampling in Variational Diffusion Models for Speech Super-Resolution</td>
<td>4831</td>
</tr>
<tr>
<td>Chin-Yun Yu, Sung-Lin Yeh, György Fazekas, Hao Tang</td>
<td></td>
</tr>
<tr>
<td>Dynamic Multi-View Scene Reconstruction Using Neural Implicit Surface</td>
<td>4836</td>
</tr>
<tr>
<td>Decai Chen, Haofei Lu, Ingo Feldmann, Oliver Schreer, Peter Eisert</td>
<td></td>
</tr>
<tr>
<td>Activity-Infomed Industrial Audio Anomaly Detection Via Source Separation</td>
<td>4841</td>
</tr>
<tr>
<td>Jaechang Kim, Yunjoo Lee, Hyun Mi Cho, Dong Woo Kim, Chi Hoon Song, Jungseul Ok</td>
<td></td>
</tr>
<tr>
<td>CC-PoseNet: Towards Human Pose Estimation in Crowded Classrooms</td>
<td>4846</td>
</tr>
<tr>
<td>Zefang Yu, Yanping Hu, Suncheng Xiang, Ting Liu, Yuzhuo Fu</td>
<td></td>
</tr>
<tr>
<td>Robustness of Deep Equilibrium Architectures to Changes in the Measurement Model</td>
<td>4851</td>
</tr>
<tr>
<td>Junhao Hu, Shirin Shoushtari, Zihao Zou, Jiaming Liu, Zhixin Sun, Ulugbek S. Kamilov</td>
<td></td>
</tr>
<tr>
<td>TABLEIE: Capturing the Interactions Among Sub-Tasks in Information Extraction Via Double Tables</td>
<td>4856</td>
</tr>
<tr>
<td>Jiaxing Lin, Runxin Xu, Baobao Chang</td>
<td></td>
</tr>
<tr>
<td>Procontext: Exploring Progressive Context Transformer for Tracking</td>
<td>4861</td>
</tr>
<tr>
<td>Jin-Peng Lan, Zhi-Qi Cheng, Jun-Yan He, Chenyang Li, Bin Luo, Xu Bao, Wangmeng Xiang, Yifeng Geng, Xuan Song Xie</td>
<td></td>
</tr>
<tr>
<td>NNSVS: A Neural Network-Based Singing Voice Synthesis Toolkit</td>
<td>4866</td>
</tr>
<tr>
<td>Ryuichi Yamamoto, Reo Yoneyama, Tomoki Toda</td>
<td></td>
</tr>
<tr>
<td>Implicitly Rotation Equivariant Neural Networks</td>
<td>4871</td>
</tr>
<tr>
<td>Naman Khetan, Tushar Arora, Samee Ur Rehman, Deepak K. Gupta</td>
<td></td>
</tr>
<tr>
<td>Forecasting of Breathing Events from Speech for Respiratory Support</td>
<td>4876</td>
</tr>
<tr>
<td>Aki Härmä, Ulf Großekathöfer, Okke Ouweltjes, Venkata Srikant Nallanthighal</td>
<td></td>
</tr>
</tbody>
</table>
Optimal Transport in Diffusion Modeling for Conversion Tasks in Audio Domain
Vadim Popov, Amantur Amatov, Mikhail Kudinov, Vladimir Gogoryan, Tasnima Sadekova, Ivan Vovk

Longshortnet: Exploring Temporal and Semantic Features Fusion in Streaming Perception
Chenyang Li, Zhi-Qi Cheng, Jun-Yan He, Pengyu Li, Bin Luo, Hanyuan Chen, Yifeng Geng, Jin-Peng Lan, Xuansong Xie

Wordreg: Mitigating the Gap Between Training and Inference with Worst-Case Drop Regularization
Jun Xia, Ge Wang, Bozhen Hu, Cheng Tan, Jiangbin Zheng, Yongjie Xu, Stan Z. Li

SS-ADMM: Stationary and Sparse Granger Causal Discovery for Cortico-Muscular Coupling
Farwa Abbas, Verity McClelland, Zoran Cvetkovic, Wei Dai

LP-IOANet: Efficient High Resolution Document Shadow Removal
Konstantinos Georgiadis, M. Kerim Yucel, Evangelos Skartados, Valtia Dimaridou, Anastasios Drosou, Albert Saà-Garriga, Bruno Manganelli

Joint Ann-SNN Co-training for Object Localization and Image Segmentation
Marc Baltes, Nidal Abuhajar, Ye Yue, Charles D. Smith, Jundong Liu

Frame-Wise and Overlap-Robust Speaker Embeddings for Meeting Diarization
Tobias Cord-Landwehr, Christoph Boeddeker, Catalin Zorila, Rama Doddipatla, Reinhold Haeb-Umbach

Effective Graph-Based Modeling of Articulation Traits for Mispronunciation Detection and Diagnosis
Bi-Cheng Yan, Hsin-Wei Wang, Yi-Cheng Wang, Berlin Chen

Tensorized Neural Layer Decomposition for 2-D DOA Estimation
Hang Zheng, Chengwei Zhou, Sergiy A. Vorobyov, Zhiguo Shi

Non-Convex Approaches for Low-Rank Tensor Completion Under Tubal Sampling
Zheng Tan, Longxiu Huang, Hanqin Cai, Yifei Lou

Data Driven Joint Sensor Fusion and Regression Based on Geometric Mean Squared Error
Carlos Alejandro Lopez, Jaume Riba

Time-Resolved FMRI Shared Response Model Using Gaussian Process Factor Analysis
Mohammadreza Ebrahimi, Navona Calarco, Colin Hawco, Aristotle Voineskos, Ashish Khisti

Model Fingerprinting with Benign Inputs
Thibault Maho, Teddy Furon, Erwan Le Merrer

Automatic Camera Pose Estimation by Key-Point Matching of Reference Objects
Jinchen Zeng, Rick Butler, John J. Van Den Dobbelsteen, Benno H. W. Hendriks, Maarten Van Der Elst, Justin Dauwels

To Regularize Or Not to Regularize: The Role of Positivity in Sparse Array Interpolation with a Single Snapshot
Mehmet Can Hücümenoglu, Pulak Sarangi, Robin Rajamäki, Piya Pal

Style Modeling for Multi-Speaker Articulation-to-Speech
Miseul Kim, Zhenyu Piao, Jihyun Lee, Hong-Goo Kang
DRL Path Planning for UAV-Aided V2X Networks: Comparing Discrete to Continuous Action Spaces .......................................................... 4961

Leonardo Spampinato, Alessia Tarozzi, Chiara Buratti, Riccardo Marini

On Super-Resolution with Separation Prior .................................................................................................................. 4966

Xingyun Mao, Heng Qiao

A Patient Invariant Model Towards the Prediction of Freezing of Gait ........................................................................ 4971

Nasimuddin Ahmed, Shivam Singhal, Aniruddha Sinha, Avik Ghose

Multi-Speaker Speech Synthesis from Electromyographic Signals by Soft Speech Unit Prediction ......................... 4976

Kevin Scheck, Tanja Schultz

Deep Subband Network for Joint Suppression of Echo, Noise and Reverberation in Real-Time Fullband Speech Communication .............................................................................................................. 4981

Feifei Xiong, Minya Dong, Kechenyong Zhou, Houwei Zhu, Jinwei Feng

Joint Neural Representation for Multiple Light Fields .................................................................................................... 4986

Guillaume Le Guludec, Christine Guillemot

Fan-Net: Fourier-Based Adaptive Normalization for Cross-Domain Stroke Lesion Segmentation .................. 4991

Wei Yi Yu, Yiming Lei, Hongming Shan

Structured Errors-In-Variables Modelling for Cortico-Muscular Coherence Enhancement ..................................... 4996

Zhenghao Guo, Verity M. McClelland, Wei Dai, Zoran Cvetkovic

Multiscale Audio Spectrogram Transformer for Efficient Audio Classification ......................................................... 5001

Wentao Zhu, Mohamed Omar

CB-Conformer: Contextual Biasing Conformer for Biased Word Recognition .......................................................... 5006

Yaoxun Xu, Baiji Liu, Qiaochu Huang, Xingchen Song, Zhiyong Wu, Shiyan Kang, Helen Meng

SENER: Sentiment Element Named Entity Recognition for Aspect-Based Sentiment Analysis .......................... 5011

Sun-Kyung Lee, Jong-Hwan Kim

Exploiting Spatial Information with the Informed Complex-Valued Spatial Autoencoder for Target Speaker Extraction ......................................................................................................................... 5016

Annika Briegleb, Mhd Modar Halimeh, Walter Kellermann

Enrollment Rate Prediction in Clinical Trials Based on CDF Sketching and Tensor Factorization Tools ...................... 5021

Magda Amiridi, Cheng Qian, Nicholas D. Sidiropoulos, Lucas M. Glass

Domain Adaptation with External Off-Policy Acoustic Catalogs for Scalable Contextual End-to-End Automated Speech Recognition ........................................................................................................ 5026

David M. Chan, Shalini Ghosh, Ariya Rastrow, Björn Hoffmeister

Towards Robust Data-Driven Underwater Acoustic Localization: A Deep CNN Solution with Performance Guarantees for Model Mismatch .......................................................... 5031

Amir Weiss, Andrew C. Singer, Gregory W. Wornell

Raising the Limit of Image Rescaling Using Auxiliary Encoding ........................................................................ 5036

Chenzhong Yin, Zhihong Pan, Xin Zhou, Le Kang, Paul Bogdan
LSTM-Based Video Quality Prediction Accounting for Temporal Distortions in Videoconferencing
Gabriel Mittag, Babak Naderi, Vishak Gopal, Ross Cutler

Extended Kalman Filter for Graph Signals in Nonlinear Dynamic Systems
Guy Sagi, Nir Shlezinger, Tirza Routtenberg

Quantum Deep Recurrent Reinforcement Learning
Samuel Yen-Chi Chen

Prune Then Distill: Dataset Distillation with Importance Sampling
Anirudh S Sundar, Gokce Keskin, Chander Chandak, I-Fan Chen, Pegah Ghahremani, Shalini Ghosh

Pair DETR: Toward Faster Convergent DETR
Seyed Mehdi Iranmanesh, Sherry X. Chen, Kuo-Chin Lien

Enhanced Coprime Array Configuration for DoA Estimation of Non-Circular Signals
Nabil Mohsen, Ammar Hawbani, Xingfu Wang, Benjamin Bairrington, Liang Zhao, Saeed Alsamhi

Audio Barlow Twins: Self-Supervised Audio Representation Learning
Jonah Anton, Harry Coppock, Pancham Shukla, Björn W. Schuller

Interpretable, Unrolled Deep Radar Beampattern Design
Kareem Metwaly, Junho Kweon, Khaled Alhujaili, Maria Greco, Fulvio Gini, Vishal Monga

Adaptive Multi-Corpora Language Model Training for Speech Recognition
Yingyi Ma, Zhe Liu, Xuedong Zhang

Hardware-Limited Non-Uniform Task-Based Quantizers
Neil Irwin Bernardo, Jingge Zhu, Yonina C. Eldar, Jamie Evans

Expectation Propagation on Factor Graphs Based on Matrix Decomposition
Adam Mekhiche, Antonio Maria Cipriano, Charly Poulliat

An End-To-End Neural Network for Image-To-Audio Transformation
Liu Chen, Michael Deisher, Munir Georges

Lit the Darkness: Three-Stage Zero-Shot Learning for Low-Light Enhancement with Multi-Neighbor Enhancement Factors
Mariam Saeed, Marwan Torki

HuBERT-AGG: Aggregated Representation Distillation of Hidden-Unit Bert for Robust Speech Recognition
Wei Wang, Yanmin Qian

Single-Anchor UWB Localization Using Channel Impulse Response Distributions
Sitian Li, Alexios Balatsoukas-Stimming, Andreas Burg

Efficient Siamese Network for UAV Tracking
Xiaohan Zhang, Dong Wang, Xiaohong Ma

Optimal Carrier Frequency Design for Frequency Diverse Array Mimo Radar
Jie Cheng, Maria Juhlin, Wen-Qin Wang, Andreas Jakobsson
Learning Audio-Visual Dereverberation

Changan Chen, Wei Sun, David Harwath, Kristen Grauman

Can Spoofing Countermeasure and Speaker Verification Systems Be Jointly Optimised?

Wanying Ge, Hemlata Tak, Massimiliano Todisco, Nicholas Evans

Exploring Vision Transformer Layer Choosing for Semantic Segmentation

Fangjian Lin, Yizhe Ma, Shengwei Tian

Dataset Balancing Can Hurt Model Performance

R. Channing Moore, Daniel P. W. Ellis, Eduardo Fonseca, Shawn Hershey, Aren Jansen, Manoj Plakal

Mcrood: Multi-Class Radar Out-Of-Distribution Detection

Sabri Mustafa Kahya, Muhammet Sami Yavuz, Eckehard Steinbach

Data Augmentation Based on Invariant Shape Blending for Deep Learning Classification

Emna Ghorbel, Mahmoud Ghorbel, Slim Mhiri

Superem: Revisiting Clustering for Semi-Supervised Learning

Durgesh Singh, Aïchina Boubekki, Robert Jenssen, Michael C. Kampffmeyer

Multi-Rate Adaptive Transform Coding for Video Compression

Lyndon R. Duong, Bohan Li, Cheng Chen, Jingning Han

Towards Explainable Recommendation Via Bert-Guided Explanation Generator

Huijing Zhan, Ling Li, Shaohua Li, Weide Liu, Manas Gupta, Alex C. Kot

FedAudio: A Federated Learning Benchmark for Audio Tasks

Tuo Zhang, Tiantian Feng, Samiul Alam, Sunwoo Lee, Mi Zhang, Shrikanth S. Narayanan, Salman Avestimehr

Dynamic Signed Graph Learning

Abdullah Karaaslanli, Selin Aviyente

Equivalence of Aperture Reduction in Element Space and Constrained Combination of DFT Beams in Beamspace

Damir Rakhimov, Martin Haardt

Neurally Augmented State Space Model for Simultaneous Communication and Tracking with Low Complexity Receivers

Fernando Pedraza, Giuseppe Caire

Continilm: A Continual Learning Scheme for Non-Intrusive Load Monitoring

Stavros Sykiotis, Maria Kaselimi, Anastasios Doulamis, Nikolaos Doulamis

Efficiently Fusing Sparse Lidar for Enhanced Self-Supervised Monocular Depth Estimation

Yue Wang, Mingrong Gong, Lei Xia, Qieshi Zhang, Jun Cheng

Learned Generative Misspecified Lower Bound

Hai Victor Habi, Hagit Messer, Yoram Bresler

Robust and Parallelizable Tensor Completion Based on Tensor Factorization and Maximum Correntropy Criterion

Yicong He, George K. Atia
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Prompting for Adversarial Robustness</td>
<td>5211</td>
</tr>
<tr>
<td>Visual Prompting for Adversarial Robustness</td>
<td>5211</td>
</tr>
<tr>
<td>Aochuan Chen, Peter Lorenz, Yuguang Yao, Pin-Yu Chen, Sijia Liu</td>
<td></td>
</tr>
<tr>
<td>Sparse Convolution Based Octree Feature Propagation for Lidar Point Cloud Compression</td>
<td>5216</td>
</tr>
<tr>
<td>Muhammad Asad Lodhi, Jiahao Pang, Dong Tian</td>
<td></td>
</tr>
<tr>
<td>A Non-Contact SpO2 Estimation Using Video Magnification and Infrared Data</td>
<td>5221</td>
</tr>
<tr>
<td>Thomas Stogiannopoulos, Grigoris-Aris Cheimariotis, Nikolaos Mitianoudis</td>
<td></td>
</tr>
<tr>
<td>Speech Separation with Large-Scale Self-Supervised Learning</td>
<td>5226</td>
</tr>
<tr>
<td>Zhuo Chen, Naoyuki Kanda, Jian Wu, Yu Wu, Xiaofei Wang, Takuya Yoshioka, Jinyu Li, Sunit</td>
<td></td>
</tr>
<tr>
<td>Sivasankaran, Sefik Emre Eskimez</td>
<td></td>
</tr>
<tr>
<td>A Variational Inequality Model for Learning Neural Networks</td>
<td>5231</td>
</tr>
<tr>
<td>Patrick L. Combettes, Jean-Christophe Pesquet, Audrey Repetti</td>
<td></td>
</tr>
<tr>
<td>High-Dimensional Confidence Regions in Sparse MRI</td>
<td>5236</td>
</tr>
<tr>
<td>Frederik Hoppe, Felix Krahmer, Claudio Mayrink Verdun, Marion I. Menzel, Holger Rauhut</td>
<td></td>
</tr>
<tr>
<td>Jointly Visual- And Semantic-Aware Graph Memory Networks for Temporal Sentence Localization</td>
<td>5241</td>
</tr>
<tr>
<td>in Videos</td>
<td></td>
</tr>
<tr>
<td>Daizong Liu, Pan Zhou</td>
<td></td>
</tr>
<tr>
<td>Temporal Contrastive Learning with Curriculum</td>
<td>5246</td>
</tr>
<tr>
<td>Shuvendu Roy, Ali Etemad</td>
<td></td>
</tr>
<tr>
<td>Fine-Grained Textual Knowledge Transfer to Improve RNN Transducers for Speech Recognition</td>
<td>5251</td>
</tr>
<tr>
<td>and Understanding</td>
<td></td>
</tr>
<tr>
<td>Vishal Sunder, Samuel Thomas, Hong-Kwang J. Kuo, Brian Kingsbury, Eric Fosler-Lussier</td>
<td></td>
</tr>
<tr>
<td>Transformer-Based Bioacoustic Sound Event Detection on Few-Shot Learning Tasks</td>
<td>5256</td>
</tr>
<tr>
<td>Liwen You, Erika Pelaez Coyotl, Suren Gunuru, Maarten Van Segbroeck</td>
<td></td>
</tr>
<tr>
<td>RIS-Aided Wideband DFRC with Reconfigurable Holographic Surface</td>
<td>5261</td>
</tr>
<tr>
<td>Tong Wei, Linlong Wu, Kumar Vijay Mishra, M. R. Bhavani Shankar</td>
<td></td>
</tr>
<tr>
<td>HARQ Delay Minimization of 5G Wireless Network with Imperfect Feedback</td>
<td>5266</td>
</tr>
<tr>
<td>Weihang Ding, Mohammad Shikh-Bahaei</td>
<td></td>
</tr>
<tr>
<td>Guided Speech Enhancement Network</td>
<td>5271</td>
</tr>
<tr>
<td>Yang Yang, Shao-Fu Shih, Hakan Erdogan, Jamie Menjay Lin, Chehung Lee, Yinpeng Li, George</td>
<td></td>
</tr>
<tr>
<td>Sung, Matthias Grundsmut</td>
<td></td>
</tr>
<tr>
<td>JEIT: Joint End-To-End Model and Internal Language Model Training for Speech Recognition</td>
<td>5276</td>
</tr>
<tr>
<td>Zhong Meng, Weiran Wang, Rohit Prabhavalkar, Tara N. Sainath, Tongzhou Chen, Ehsan Variani,</td>
<td></td>
</tr>
<tr>
<td>Yu Zhang, Bo Li, Andrew Rosenberg, Bhuvana Ramabhadran</td>
<td></td>
</tr>
<tr>
<td>Bayesian Methods for Optical Flow Estimation Using a Variational Approximation, with</td>
<td>5281</td>
</tr>
<tr>
<td>Applications to Ultrasound</td>
<td></td>
</tr>
<tr>
<td>Jan Dorazil, Bernard H. Fleury, Franz Hlawatsch</td>
<td></td>
</tr>
<tr>
<td>Variable Attention Masking for Configurable Transformer Transducer Speech Recognition</td>
<td>5286</td>
</tr>
<tr>
<td>Pawel Swietojanski, Stefan Braun, Dogan Can, Thiago Fraga Da Silva, Arnab Ghoshal,</td>
<td></td>
</tr>
<tr>
<td>Takaaki Hori, Roger Hsiao, Henry Mason, Erik McDermott, Honza Silovsky, Ruchir Travadi,</td>
<td></td>
</tr>
<tr>
<td>Xiaodan Zhuang</td>
<td></td>
</tr>
</tbody>
</table>
Enhancing Multimodal Alignment with Momentum Augmentation for Dense Video Captioning
Yiwei Wei, Shaozu Yuan, Meng Chen, Longbiao Wang

Robust Hypothesis Testing with Moment Constrained Uncertainty Sets
Akshayaa Magesh, Zhongchang Sun, Venugopal V. Veeravalli, Shaofeng Zou

Adaptive Step-Size Methods for Compressed SGD
Adarsh M. Subramaniam, Akshayaa Magesh, Venugopal V. Veeravalli

Study on the Fairness of Speaker Verification Systems Across Accent and Gender Groups
Mariel Estevez, Luciana Ferrer

Improving Electric Load Demand Forecasting with Anchor-Based Forecasting Method
Maria Tzelepi, Paraskevi Nousi, Anastasios Tefas

Resource-Efficient Transfer Learning from Speech Foundation Model Using Hierarchical Feature Fusion
Zhouyuan Huo, Khe Chai Sim, Bo Li, Dongseong Hwang, Tara N. Sainath, Trevor Strohman

BEANS: The Benchmark of Animal Sounds
Masato Hagiwara, Benjamin Hoffman, Jen-Yu Liu, Maddie Cusimano, Felix Effenberger, Katie Zacarian

Towards Accurate and Real-Time End-of-Speech Estimation
Yifeng Fan, Colin Yaz, Di He, Jahn Heymann, Viet Anh Trinh, Zhe Zhang, Venkatesh Ravichandran

Gaussian Process Dynamical Modeling for Adaptive Inference Over Graphs
Qin Lu, Konstantinos D. Polyzos

Pop2Piano : Pop Audio-Based Piano Cover Generation
Jongho Choi, Kyogu Lee

Locale Encoding for Scalable Multilingual Keyword Spotting Models
Pai Zhu, Hyun Jin Park, Alex Park, Angelo Scorza Scarpati, Ignacio Lopez Moreno

Bayesian Optimization with Ensemble Learning Models and Adaptive Expected Improvement
Konstantinos D. Polyzos, Qin Lu, Georgios B. Giannakis

SINCO: A Novel Structural Regularizer for Image Compression Using Implicit Neural Representations
Harry Gao, Weijie Gan, Zhixin Sun, Ulugbek S. Kamilov

Physics-Informed Transfer Learning for Voltage Stability Margin Prediction
Manish K. Singh, Konstantinos D. Polyzos, Panagiotis A. Traganitis, Saiaj V. Dhople, Georgios B. Giannakis

Adapting a Self-Supervised Speech Representation for Noisy Speech Emotion Recognition by Using Contrastive Teacher-Student Learning
Seong-Gyun Leem, Daniel Fulford, Jukka-Pekka Onnela, David Gard, Carlos Busso

Evidence of Vocal Tract Articulation in Self-Supervised Learning of Speech
Cheol Jun Cho, Peter Wu, Abdelrahman Mohamed, Gopala K. Anumanchipalli

Active Subsampling Using Deep Generative Models by Maximizing Expected Information Gain
Koen C. E. Van De Camp, Hamdi Joudeh, Duarte J. Antunes, Ruud J. G. Van Sloun
A Synthetic Corpus Generation Method for Neural Vocoder Training .......................................................... 5376
Zilin Wang, Peng Liu, Jun Chen, Sipan Li, Jinfeng Bai, Gang He, Zhiyong Wu, Helen Meng

Alternating Constrained Minimization Based Approximate Message Passing .................................................. 5381
Christo Kurisummoottil Thomas, Dirk Slock

Progressive Diversifying Policy for Multi-Agent Reinforcement Learning ..................................................... 5386
Shaooqi Sun, Yuanzhao Zhai, Kele Xu, Dawei Feng, Bo Ding

Towards Real-Time Single-Channel Speech Separation in Noisy and Reverberant Environments .................. 5391
Julian Neri, Sebastian Braun

Accidental Learners: Spoken Language Identification in Multilingual Self-Supervised Models ....................... 5396
Travis M. Bartley, Fei Jia, Krishna C. Puvvada, Samuel Kriman, Boris Ginsburg

Implementing Continuous HRTF Measurement in Near-Field ................................................................. 5401
Ee-Leng Tan, Santi Peksi, Woon-Seng Gan

Efficient Protein Structural Class Prediction Via Chaos Game Representation and Recurrent Neural Networks .......................................................... 5406
Michaela Areti Zervou, Effrosyni Doutsi, Panagiotis Tsakalides

Efficient Similarity-Based Passive Filter Pruning for Compressing CNNS ................................................... 5411
Arshdeep Singh, Mark D. Plumbley

Leveraging Multiple Sources in Automatic African American English Dialect Detection for Adults and Children .......................................................... 5416
Alexander Johnson, Vishwas M. Shetty, Mari Ostendorf, Abeer Alwan

MTDL-NET: Morphological and Temporal Discriminative Learning for Heartbeat Classification .................. 5421
Can Han, Suncheng Xiang, Dahong Qian

Blind Source Counting and Separation with Relative Harmonic Coefficients ............................................ 5426
Huiyuan Sun, Prasanga Samarasinghe, Thushara Abhayapala

Global Matching-Optimization Network for Stereo Depth Estimation ....................................................... 5431
Yidi Zhang, Wenqi Huang, Wenming Yang

U-Beat: A Multi-Scale Beat Tracking Model Based on Wave-U-Net .......................................................... 5436
Tian Cheng, Masataka Goto

Low-Bitrate Redundancy Coding of Speech Using a Rate-Distortion-Optimized Variational Autoencoder ........... 5441
Jean-Marc Valin, Jan Büthe, Ahmed Mustafa

Zero-Shot Domain Adaptation of Anomalous Samples for Semi-Supervised Anomaly Detection .................. 5446
Tomoyaa Nishida, Takashi Endo, Yohei Kawaguchi

Estimation of High-Dimensional Differential Graphs from Multi-Attribute Data ........................................... 5451
Jitendra K. Tugnait

GAPter: Gray-Box Data Protector for Deep Learning Inference Services at User Side .................................. 5456
Hao Wu, Bo Yang, Xiaopeng Ke, Siyi He, Fengyuan Xu, Sheng Zhong

Text-To-Speech Synthesis Based on Latent Variable Conversion Using Diffusion Probabilistic Model and Variational Autoencoder ........................................... 5461
Yusuke Yasuda, Tomoki Toda
RD-NAS: Enhancing One-Shot Supernet Ranking Ability Via Ranking Distillation from Zero-Cost Proxies
Peijie Dong, Xin Niu, Lujun Li, Zhiliang Tian, Xiaodong Wang, Zimian Wei, Hengyue Pan,
Dongsheng Li .................................................. 5466

ADHD Classification with Biomarker Identification Using a Triplet Loss Attention Auto-Encoding Network
Yibin Tang, Ying Chen, Yuan Gao, Aimin Jiang, Lin Zhou .................................................. 5471

Phase Retrieval for Rydberg Quantum Arrays
Peter Vouras, Kumar Vijay Mishra, Alexandra Artusio-Glimpse .................................................. 5476

Tempo Vs. Pitch: Understanding Self-Supervised Tempo Estimation
Giovana Morais, Matthew E. P. Davies, Marcelo Queiroz, Magdalena Fuentes .................................. 5481

Perceptual Quality Assessment for Digital Human Heads
Zicheng Zhang, Yingjie Zhou, Wei Sun, Xiongkuo Min, Yuzhe Wu, Guangtao Zhai .................................. 5486

Leveraging Language Embeddings for Cross-Lingual Self-Supervised Speech Representation Learning
Tomohiro Tanaka, Ryo Masumura, Mana Itori, Hiroshi Sato, Taiga Yamane, Takanori Ashihara, Kohei Matsuura, Takafumi Moriya .................................................. 5491

Progressive Meta-Pooling Learning for Lightweight Image Classification Model
Peijie Dong, Xin Niu, Zhiliang Tian, Lujun Li, Xiaodong Wang, Zimian Wei, Hengyue Pan,
Dongsheng Li .................................................. 5496

Cross-Training: A Semi-Supervised Training Scheme for Speech Recognition
Soheil Khorram, Anshuman Tripathi, Jaeyoung Kim, Han Lu, Qian Zhang, Rohit Prabhavalkar, Hasim Sak .................................................. 5501

Multi-View Millimeter-Wave Imaging Over Wireless Cellular Network
Xin Tong, Zhaoyang Zhang, Zhaohui Yang .................................................. 5506

SADI: A Self-Adaptive Decomposed Interpretable Framework for Electric Load Forecasting Under Extreme Events
Hengbo Liu, Ziqing Ma, Linxiao Yang, Tian Zhou, Rui Xia, Yi Wang, Qingsong Wen, Liang Sun .................................................. 5511

Spatial-Domain Object Detection Under Mimo-Fmcw Automotive Radar Interference
Sian Jin, Pu Wang, Petros Boufounos, Ryuhei Takahashi, Sumit Roy .................................................. 5516

Speech Summarization of Long Spoken Document: Improving Memory Efficiency of Speech/Text Encoders
Takatomo Kano, Atsunori Ogawa, Marc Delcroix, Roshan Sharma, Kohei Matsuura, Shinji Watanabe .................................................. 5521

UNeXt: A Low-Dose CT Denoising UNet Model with the Modified ConvNeXt Block
Farzan Niknejad Mazandarani, Paul Babyn, Javad Alirezaie .................................................. 5526
Flow-Guided Deformable Alignment Network with Self-Supervision for Video Inpainting
 Zhiliang Wu, Kang Zhang, Changchang Sun, Hanyu Xuan, Yan Yan

A Novel State Connection Strategy for Quantum Computing to Represent and Compress Digital Images
 Md Ershadul Haque, Manoranjan Paul, Anwar Ulhaq, Tanmoy Debnath

Light-Weight Sequential SBL Algorithm: An Alternative to OMP
 Rohan R. Pote, Bhaskar D. Rao

Robust Video Object Segmentation with Restricted Attention
 Huaizheng Zhang, Pinxue Guo, Zhongwen Le, Wenqiang Zhang

MmWave Wi-Fi Trajectory Estimation with Continuous-Time Neural Dynamic Learning
 Cristian J. Vaca-Rubio, Pu Wang, Toshiaki Koike-Akino, Ye Wang, Petros Boufounos, Petar Popovski

Multi-Stream Facial Adaptive Network for Expression Recognition from a Single Image
 Baichuan Zhang, Fanyang Meng, Runwei Ding, Mengyuan Liu

On the Role of Visual Context in Enriching Music Representations
 Kleanthis Avramidis, Shanti Stewart, Shrikanth Narayanan

TINYCOD: Tiny and Effective Model for Camouflaged Object Detection
 Haozhe Xing, Shuyong Gao, Hao Tang, Tsui Qin Mok, Yanlan Kang, Wenqiang Zhang

Soft Dynamic Time Warping for Multi-Pitch Estimation and Beyond
 Michael Krause, Christof Weiß, Meinard Müller

Classification of the Cervical Vertebrae Maturation (CVM) Stages Using the Tripod Network
 Salih Atici, Hongyi Pan, Mohammed H. Elnagar, Veerasathpurush Allareddy, Omar Suhaym, Rashid Ansari, Ahmet Enis Cetin

Multi-Agent Reinforcement Learning for Covert Semantic Communications Over Wireless Networks
 Yining Wang, Ye Hu, Hongyang Du, Tao Luo, Dusit Niyato

Framewise Wavegan: High Speed Adversarial Vocoder in Time Domain with Very Low Computational Complexity
 Ahmed Mustafa, Jean-Marc Valin, Jan Büthe, Paris Smaragdis, Mike Goodwin

Prompttts: Controllable Text-To-Speech with Text Descriptions
 Zhifang Guo, Yichong Leng, Yihan Wu, Sheng Zhao, Xu Tan

Intermediate Fine-Tuning Using Imperfect Synthetic Speech for Improving Electrolaryngeal Speech Recognition
 Lester Phillip Violeta, Ding Ma, Wen-Chin Huang, Tomoki Toda

Semantic Preprocessor for Image Compression for Machines
 Mingyi Yang, Luis Herranz, Fei Yang, Luka Murn, Marc Gorriz Blanch, Shuai Wan, Fuzheng Yang, Marta Mrak

Privacy Preserving Face Recognition with Lensless Camera
 Chris Henry, M. Salman Asif, Zhu Li

High Quality Audio Coding with Mdctnet
 Grant Davidson, Mark Vinton, Per Ekstrand, Cong Zhou, Lars Villemoes, Lie Lu
Search for Efficient Deep Visual-Inertial Odometry Through Neural Architecture Search
Yu Chen, Mingyu Yang, Hun-Seok Kim

An Antispoofing Approach in Biometric Authentication System for a Smartcard
Han-Sol Lee, Moon-Kyu Song, Junseo Lee, Yeolmin Seong, Ducksoo Kim, Kwanghyuk Bae, Seongwook Song

Decoupled Visual Causality for Robust Detection
Ping Jiang, Xiaoheng Deng, Shichao Zhang

Factorized Blank Thresholding for Improved Runtime Efficiency of Neural Transducers
Duc Le, Frank Seide, Yuhao Wang, Yang Li, Kjell Schubert, Ozlem Kalinli, Michael L. Seltzer

Nonparallel Emotional Voice Conversion for Unseen Speaker-Emotion Pairs Using Dual Domain
Adversarial Network & Virtual Domain Pairing
Nirmesh Shah, Mayank Singh, Naoya Takahashi, Naoyuki Onoe

Stacking-Based Attention Temporal Convolutional Network for Action Segmentation
Liu Yang, Yu Jiang, Junkun Hong, Zhenjie Wu, Zhan Yang, Jun Long

Lightweight Fisher Vector Transfer Learning for Video Deduplication
Chris Henry, Rifan Liao, Ruiyuan Lin, Zhebin Zhang, Hongyu Sun, Zhu Li

Iterative Water-Filling Power and Subcarrier Allocation for Multicarrier NOMA Downlink
Chin Choy Chai, Xiao-Ping Zhang

Self-Adaptive Reasoning on Sub-Questions for Multi-Hop Question Answering
Zekai Li, Wei Peng

Optimal Compression for Minimizing Classification Error Probability: An Information-Theoretic Approach
Jingchao Gao, Ao Tang, Weiyu Xu

Quickest Change Detection with Leave-One-Out Density Estimation
Yuchen Liang, Venugopal V. Veeravalli

Self-Supervised Learning-Based Source Separation for Meeting Data
Yuang Li, Xianrui Zheng, Philip C. Woodland

The R3VIVAL Dataset: Repository of Room Responses and 360 Videos of a Variable Acoustics Lab
Florian Klein, Sebastià V. Amengual Gari

Unrestricted Anchor Graph Based GCN for Incomplete Multi-View Clustering
Liang Zhao, Zihao Wang, Yukun Yuan, Feng Ding

DecomFormer: Decompose Self-Attention Via Fourier Transform for VHR Aerial Image Scene Classification
Yan Zhang, Xiuyuan Gao, Xiao Pu, Tao Wang, Xinbo Gao

Nonparallel High-Quality Audio Super Resolution with Domain Adaptation and Resampling
CycleGANs
Reo Yoneyama, Ryuichi Yamamoto, Kentaro Tachibana

COVID-19 Detection from Speech in Noisy Conditions
Shuo Liu, Adria Mallol-Ragolta, Björn W. Schuller
Sparse Black-Box Inversion Attack with Limited Information ................................................................. 5711
Yixiao Xu, Xiaolei Liu, Teng Hu, Bangzhuo Xin, Run Yang

Small-Footprint Slimmable Networks for Keyword Spotting ................................................................. 5716
Zuhaib Akhtar, Mohammad Omar Khursheed, Dongsu Du, Yuzong Liu

BAUENet: Boundary-Aware Uncertainty Enhanced Network for Infrared Small Target Detection .......... 5721
Tianxiang Chen, Qi Chu, Zhentao Tan, Bin Liu, Nenghai Yu

A Novel Extrapolation Technique to Accelerate WMMSE ......................................................................... 5726
Kaiwen Zhou, Zilin Chen, Guochen Liu, Zhitang Chen

Bagging R-CNN: Ensemble for Object Detection in Complex Traffic Scenes .......................................... 5731
Pengteng Li, Ying He, Dongfu Yin, F Richard Yu, Pinhao Song

Possibilistic Bernoulli Filter for Extended Target Tracking ........................................................................ 5736
Zhijin Chen, Branko Ristic, Du Yong Kim

Does Human Speech Follow Benford’s Law? ............................................................................................. 5741
Leo Hsu, Visar Berisha

Image Adversarial Steganography Based on Joint Distortion ...................................................................... 5746
Zexin Fan, Kejiang Chen, Chuan Qin, Kai Zeng, Weiming Zhang, Nenghai Yu

End-To-End Neural Audio Coding in the MDCT Domain ........................................................................... 5751
Hyungsob Lim, Jihyun Lee, Byeong Hyeon Kim, Inseon Jang, Hong-Goo Kang

Multimodal Knowledge Distillation for Arbitrary-Oriented Object Detection in Aerial Images ............ 5756
Zhanchao Huang, Wei Li, Ran Tao

Simultaneous Reconstruction and Uncertainty Quantification for Tomography ....................................... 5761
Agnimitra Dasgupta, Carlo Graziani, Zichao Wendy Di

Joint Waveform and Passive Beamformer Design in Multi-IRS-Aided Radar ........................................ 5766
Zahra Esmaeilbeig, Arian Eamaz, Kumar Vijay Mishra, Mojtaba Soltanalian

SingNet: A Real-Time Singing Voice Beat and Downbeat Tracking System .......................................... 5771
Mojtaba Heydari, Ju-Chi Wang, Zhiyao Duan

RCDPT: Radar-Camera Fusion Dense Prediction Transformer ............................................................. 5776
Chen-Chou Lo, Patrick Vandewalle

Local to Global Prior Learning for Blind Unsupervised Image Super Resolution .................................... 5781
Kazuhiro Yamawaki, Xian-Hua Han

Reducing the Computational Complexity of Learning with Random Convolutional Features ................ 5786
M. A. Omidi, B. Seyfe, S. Valaee

On Unsupervised Uncertainty-Driven Speech Pseudo-Label Filtering and Model Calibration .............. 5791
Nauman Dawalatabad, Sameer Khurana, Antoine Laurent, James Glass

Scalable Weight Reparametrization for Efficient Transfer Learning ....................................................... 5796
Byeongjeun Kim, Jun-Tae Lee, Seunghan Yang, Simyung Chang

Eigen-Decomposition-Free Directed Graph Sampling Via Gershgorin Disc Alignment ....................... 5801
Yuejiang Li, H. Vicky Zhao, Gene Cheung
Stay in the Middle: A Semi-Supervised Model for CT Metal Artifact Reduction .................................................. 5806
Tao Wang, Hui Yu, Zexin Lu, Zhongzhou Zhang, Jiliu Zhou, Yi Zhang

A Bayesian Perspective for Determinant Minimization Based Robust Structured Matrix Factorization .................................................. 5811
Gokcan Tatli, Alper T. Erdogan

Unsupervised Domain Adaptation Via Subspace Interpolating Deep Dictionary Learning: A Case Study in Machine Inspection .................................................. 5816
Kriti Kumar, Anshul Majumdar, A Anil Kumar, M Girish Chandra

Dual-Attention Neural Transducers for Efficient Wake Word Spotting in Speech Recognition .................................................. 5821
Saumya Y. Sahai, Jing Liu, Thejaswi Muniyappa, Kanthashree M. Sathyendra, Anastasios Alexandridis, Grant P. Strimel, Ross McGowan, Ariya Rastrow, Feng-Ju Chang, Athanasios Moutcharis, Siegfried Kunzmann

LiNuIQA: Lightweight No-Reference Image Quality Assessment Based on Non-Uniform Weighting .................................................. 5826
Wook-Hyung Kim, Cheul-Hee Hahm, Anant Baijal, Namuk Kim, Ilhyun Cho, Jayoon Koo

MFAT: A Multi-Level Feature Aggregated Transformer for Person Re-Identification .................................................. 5831
Bowen Tan, Linfeng Xu, Zihuan Qiu, Qingbo Wu, Fanman Meng

Single-Particle Tracking by Graph Transformer .................................................. 5836
Satoshi Kamiya, Kazuhiro Hotta, Taka-Aki Tsunoyama, Akihiro Kasumi

Large-Scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation .................................................. 5841
Yusong Wu, Ke Chen, Tianyu Zhang, Yuchen Hui, Taylor Berg-Kirkpatrick, Shlomo Dubnov

Active Noise Control Over 3D Space: A Realistic Error Microphone Geometry Design .................................................. 5846
Huiyuan Sun, Prasanga Samarasinghe, Thushara Abhayapala

MTFD: Multi-Teacher Fusion Distillation for Compressed Video Action Recognition .................................................. 5851
Jinxin Guo, Jiaqiang Zhang, Shaojie Li, Xiaoqing Zhang, Ming Ma

Comparison of Soft and Hard Target RNN-T Distillation for Large-Scale ASR .................................................. 5856
Dongseong Hwang, Khe Chai Sim, Yu Zhang, Trevor Strohman

Mask Guided Selective Context Decoding for Handwritten Chinese Text Recognition .................................................. 5861
Tao Li, Shilian Wu, Zengfu Wang

A Multi-Stage Hierarchical Relational Graph Neural Network for Multimodal Sentiment Analysis .................................................. 5866
Peizhu Gong, Jin Liu, Xiliang Zhang, Xingye Li

Towards Practical Edge Inference Attacks Against Graph Neural Networks .................................................. 5871
Kailai Li, Jiawei Sun, Ruoxin Chen, Wei Ding, Kexue Yu, Jie Li, Chentao Wu

Halluaudio: Hallucinate Frequency as Concepts for Few-Shot Audio Classification .................................................. 5876
Zhongjie Yu, Shuyang Wang, Lin Chen, Zhongwei Cheng

AVES: Animal Vocalization Encoder Based on Self-Supervision .................................................. 5881
Masato Hagiwara

Boosting Signal Modulation Few-Shot Learning with Pre-Transformation .................................................. 5886
Peng Sun, Jie Su, Zhenyu Wen, Yejian Zhou, Zhen Hong, Shanqing Yu, Huaji Zhou
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leveraging Heteroscedastic Uncertainty in Learning Complex Spectral Mapping for Single-Channel Speech Enhancement</td>
<td>5891</td>
</tr>
<tr>
<td>Kuan-Lin Chen, Daniel D. E. Wong, Ke Tan, Buye Xu, Anurag Kumar, Vamsi Krishna Ithapu</td>
<td></td>
</tr>
<tr>
<td>Metric Learning for User-Defined Keyword Spotting</td>
<td>5896</td>
</tr>
<tr>
<td>Jaemin Jung, Youkyum Kim, Jihwan Park, Youshin Lim, Byeong-Yeol Kim, Youngjoon Jang, Joon Son Chung</td>
<td></td>
</tr>
<tr>
<td>Multi-View Graph Regularized Deep Autoencoder-Like NMF Framework</td>
<td>5901</td>
</tr>
<tr>
<td>Liang Zhao, Zihao Wang, Ziyue Wang, Zhikui Chen</td>
<td></td>
</tr>
<tr>
<td>A Spatial-Temporal ECG Emotion Recognition Model Based on Dynamic Feature Fusion</td>
<td>5906</td>
</tr>
<tr>
<td>Shuo Xiao, Xiaojing Qiu, Chaogang Tang, Zhenzhen Huang</td>
<td></td>
</tr>
<tr>
<td>A Flow-Guided Non-Local Alignment Network for Video Compressive Sensing Reconstruction</td>
<td>5911</td>
</tr>
<tr>
<td>Chao Zhou, Can Chen, Dengyin Zhang</td>
<td></td>
</tr>
<tr>
<td>Compressed-Sensing-Based 3D Localization with Distributed Passive Reconfigurable Intelligent Surfaces</td>
<td>5916</td>
</tr>
<tr>
<td>Jiguang He, Aymen Fakhreddine, Henk Wyneersch, George C. Alexandropoulos</td>
<td></td>
</tr>
<tr>
<td>Hierarchical Pronunciation Assessment with Multi-Aspect Attention</td>
<td>5921</td>
</tr>
<tr>
<td>Heejin Do, Yunsu Kim, Gary Geunbae Lee</td>
<td></td>
</tr>
<tr>
<td>Generic Dependency Modeling for Multi-Party Conversation</td>
<td>5926</td>
</tr>
<tr>
<td>Weizhou Shen, Xiaojian Quan, Ke Yang</td>
<td></td>
</tr>
<tr>
<td>Certified Robustness of Quantum Classifiers Against Adversarial Examples Through Quantum Noise</td>
<td>5931</td>
</tr>
<tr>
<td>Jhih-Cing Huang, Yu-Lin Tsai, Chao-Han Huck Yang, Cheng-Fang Su, Chia-Mu Yu, Pin-Yu Chen, Sy-Yen Kuo</td>
<td></td>
</tr>
<tr>
<td>Statistical Analysis of Speech Disorder Specific Features to Characterise Dysarthria Severity Level</td>
<td>5936</td>
</tr>
<tr>
<td>Amlu Anna Joshy, P. N. Parameswaran, Siddharth R. Nair, Rajeev Rajan</td>
<td></td>
</tr>
<tr>
<td>Binary Sequence Set Optimization for CDMA Applications Via Mixed-Integer Quadratic Programming</td>
<td>5941</td>
</tr>
<tr>
<td>Alan Yang, Tara Mina, Grace Gao</td>
<td></td>
</tr>
<tr>
<td>Zero-Shot Anomalous Sound Detection in Domestic Environments Using Large-Scale Pretrained Audio Pattern Recognition Models</td>
<td>5946</td>
</tr>
<tr>
<td>Alessandro Ilic Mezza, Giulio Zanetti, Maximo Cobos, Fabio Antonacci</td>
<td></td>
</tr>
<tr>
<td>Compressive Sensing with Tensorized Autoencoder</td>
<td>5951</td>
</tr>
<tr>
<td>Rakib Hyder, M. Salman Asif</td>
<td></td>
</tr>
<tr>
<td>SRTNET: Time Domain Speech Enhancement Via Stochastic Refinement</td>
<td>5956</td>
</tr>
<tr>
<td>Zhibin Qiu, Mengfan Fu, Yineng Yu, Lili Yin, Fuchun Sun, Hao Huang</td>
<td></td>
</tr>
<tr>
<td>Symbol Level Precoding in the RF Domain for Low Hardware Complexity RIS-Assisted MU-MISO Systems</td>
<td>5961</td>
</tr>
<tr>
<td>Christos G. Tsinos, Theodoros A. Tsiftisis, Robert Schober</td>
<td></td>
</tr>
<tr>
<td>Hint-Dynamic Knowledge Distillation</td>
<td>5966</td>
</tr>
<tr>
<td>Yiyang Liu, Chenxin Li, Xiaotong Tu, Xinghao Ding, Yue Huang</td>
<td></td>
</tr>
</tbody>
</table>
A Benchmark for Evaluating Robustness of Spoken Language Understanding Models in Slot Filling ................................................................. 5971
Meizheng Peng, Xu Jia, Min Peng

Maskdul: Data Uncertainty Learning in Masked Face Recognition ................................................................. 5976
Libo Zhang, Weiming Xiong, Ku Zhao, Kehan Chen, Mingyang Zhong

Topological Signal Processing Over Weighted Simplicial Complexes ................................................................. 5981
Claudio Battiloro, Stefania Sardellitti, Sergio Barbarossa, Paolo Di Lorenzo

Selecting Language Models Features VIA Software-Hardware Co-Design ................................................................. 5986
Vlad Pandelea, Edoardo Ragusa, Paolo Gastaldo, Erik Cambria

A Perceptual Neural Audio Coder with a Mean-Scale Hyperprior ................................................................. 5991
Joon Byun, Seungmin Shin, Youngcheol Park, Jongmo Sung, Seungkwon Beach

Single Domain Dynamic Generalization for Iris Presentation Attack Detection ................................................................. 5996
Yachun Li, Jingjing Wang, Yuhui Chen, Di Xie, Shiliang Pu

Token2vec: A Joint Self-Supervised Pre-Training Framework Using Unpaired Speech and Text ........................................... 6001
Xianghu Yue, Junyi Ao, Xiaoxue Gao, Haizhou Li

Convergence Analysis of Graphical Game-Based Nash Q-Learning Using the Interaction Detection Signal of N-Step Return ................................................................. 6006
Yunkai Zhuang, Shangdong Yang, Wenbin Li, Yang Gao

Convex Optimization of Deep Polynomial and ReLU Activation Neural Networks ................................................................. 6011
Burak Bartan, Mert Pilanci

Joint Generative-Contrastive Representation Learning for Anomalous Sound Detection ................................................................. 6016
Xiao-Min Zeng, Yan Song, Zhu Zhuo, Yu Zhou, Yu-Hong Li, Hui Xue, Li-Rong Dai, Ian McLoughlin

Tg-Critic: A Timbre-Guided Model for Reference-Independent Singing Evaluation ................................................................. 6021
Xiaoheng Sun, Yuejie Gao, Hanyao Lin, Huaping Liu

Self-Supervised Learning of Audio Representations Using Angular Contrastive Loss ................................................................. 6026
Shanshan Wang, Soumya Tripathy, Annamaria Mesaros

Learning Expressive and Generalizable Motion Features for Face Forgery Detection ................................................................. 6031
Jingyi Zhang, Peng Zhang, Jingjing Wang, Di Xie, Shiliang Pu

NSV-TTS: Non-Speech Vocalization Modeling and Transfer in Emotional Text-To-Speech ................................................................. 6036
Haitong Zhang, Xinyuan Yu, Yue Lin

On Adversarial Robustness of Audio Classifiers ................................................................. 6041
Kangkang Lu, Manh Cuong Nguyen, Xun Xu, Chuan Sheng Foo

Mitigating Domain Dependency for Improved Speech Enhancement Via SNR Loss Boosting ................................................................. 6046
Lili Yin, Di Wu, Zhibin Qiu, Hao Huang

An End-To-End Framework for Partial View-Aligned Clustering with Graph Structure ................................................................. 6051
Liang Zhao, Qiongjie Xie, Sontao Wu, Shubin Ma

A Topic-Enhanced Approach for Emotion Distribution Forecasting in Conversations ................................................................. 6056
Xin Lu, Weixiang Zhao, Yanyan Zhao, Bing Qin, Zhentao Zhang, Junjie Wen
Learning 3D Human Pose and Shape Estimation Using Uncertainty-Aware Body Part Segmentation .......... 6061
Ziming Wang, Han Yu, Xiaoguang Zhu, Zengwen Li, Changxue Chen, Liang Song

Vitasd: Robust Vision Transformer Baselines for Autism Spectrum Disorder Facial Diagnosis .................. 6066
Xu Cao, Wengian Ye, Elena Sizikova, Xue Bai, Megan Coffee, Hongwu Zeng, Jianguo Cao

Real-Time Target Sound Extraction ........................................................................................................... 6071
Bandhav Veluri, Justin Chan, Malek Itani, Tuochao Chen, Takuya Yoshioka, Shyamnath Gollakota

Underwater Image Restoration with Light-Aware Progressive Network .................................................. 6076
Jian Yang, Chen Li, Xuelong Li

Modaldrop: Modality-Aware Regularization for Temporal-Spectral Fusion in Human Activity Recognition ........................................................................................................................................ 6081
Xin Zeng, Yiqiang Chen, Benfeng Xu, Tengxiang Zhang

A Simulation-Based Framework for Urban Traffic Accident Detection .................................................. 6086
Haohan Luo, Feng Wang

Context-Aware Fine-Tuning of Self-Supervised Speech Models .................................................................. 6091
Suwon Shon, Felix Wu, Kwangyoun Kim, Prashant Sridhar, Karen Livescu, Shinji Watanabe

DAIS: The Delft Database of EEG Recordings of Dutch Articulated and Imagined Speech ....................... 6096
Bo Dekker, Alfred C. Schouten, Odette Scharenborg

Diversifying Message Aggregation in Multi-Agent Communication Via Normalized Tensor Nuclear Norm Regularization ....................................................................................................................... 6101
Yuanzhao Zhai, Kele Xu, Bo Ding, Dawei Feng, Zijian Gao, Huaimin Wang

Learning to Auto-Correct for High-Quality Spectrograms ....................................................................... 6106
Zhiyang Zhou, Shihui Liu

Saliency-Driven Hierarchical Learned Image Coding for Machines ......................................................... 6111
Kristian Fischer, Fabian Brand, Christian Blum, André Kaup

Time-Varying Signals Recovery Via Graph Neural Networks .................................................................... 6116
Jhon A. Castro-Correa, Jhony H. Giraldo, Anindya Mondal, Mohsen Badiey, Thierry Bouwmans, Fragkiskos D. Malliaros

MHLAT: Multi-Hop Label-Wise Attention Model for Automatic ICD Coding .......................................... 6121
Junwen Duan, Han Jiang, Ying Yu

Improving Phase-Vocoder-Based Time Stretching by Time-Directional Spectrogram Squeezing ................ 6126
Natsuki Akaishi, Kohei Yatabe, Yasuhiro Oikawa

Dual-Uncertainty Guided Curriculum Learning and Part-Aware Feature Refinement for Domain Adaptive Person Re-Identification ........................................................................................................... 6131
Zhangping Liu, Bin Liu, Zhiwei Zhao, Qi Chu, Nenghai Yu

Weight Averaging: A Simple Yet Effective Method to Overcome Catastrophic Forgetting in Automatic Speech Recognition .................................................................................................................. 6136
Steven Vander Eeckt, Hugo Van Hamme

Change Point Detection with Neural Online Density-Ratio Estimator ..................................................... 6141
Xiuheng Wang, Ricardo Augusto Borsoi, Cédric Richard, Jie Chen
Finding Optimal Numerical Format for Sub-8-Bit Post-Training Quantization of Vision
Janghwan Lee, Youngdeok Hwang, Jungwook Choi

Audio-Visual Inpainting: Reconstructing Missing Visual Information with Sound
Valentina Sanguineti, Sanket Thakur, Pietro Morerio, Alessio Del Bue, Vittorio Murino

Achievable Error Exponents for Almost Fixed-Length M-Ary Hypothesis Testing
Jun Diao, Lin Zhou, Lin Bai

A Memory-Free Evolving Bipolar Neural Network for Efficient Multi-Label Stream Learning
Sourav Mishra, Suresh Sundaram

Transwnet: Integrating Transformers into CNNs Via Row and Column Attention for Abdominal
Multi-Organ Segmentation
Yazhen Xie, Yanglin Huang, Yuan Zhang, Xuanya Li, Xiongjun Ye, Kai Hu

UAV Local Path Planning Based on Improved Proximal Policy Optimization Algorithm
Jiahao Xu, Xufeng Yan, Cui Peng, Xinquan Wu, Lipeng Gu, Yanbiao Niu

Tracking Objects and Activities with Attention for Temporal Sentence Grounding
Zeyu Xiong, Daizong Liu, Pan Zhou, Jiahao Zhu

Does Your Model Think Like an Engineer? Explainable AI for Bearing Fault Detection with Deep
Learning
Thomas Decker, Michael Lebacher, Volker Tresp

Incorporating Reliability in Graph Information Propagation by Fluid Dynamics Diffusion: A Case
of Multimodal Semisupervised Deep Learning
Andrea Marinoni, Marine Mercier, Qian Shi, Sivasakthy Selvakumaran, Mark Girolami

Fast Yet Effective Speech Emotion Recognition with Self-Distillation
Zhao Ren, Thanh Tam Nguyen, Yi Chang, Björn W. Schuller

Heart Rate Estimation and Performance Analysis Using MIMO Radar with Dispersed Antennas
Peichao Wang, Qian He

Multi-Layer Seasonal Perception Network for Time Series Forecasting
Ruoshu Wang, Shengfa Miao, Di Liu, Xin Jin, Weisheng Zhang

Measure and Countermeasure of the Capsulation Attack Against Backdoor-Based Deep Neural
Network Watermarks
Fang-Qi Li, Shi-Lin Wang, Yun Zhu

A Context-Aware Computational Approach for Measuring Vocal Entrainment in Dyadic
Conversations
Rimita Lahiri, Md Nasir, Catherine Lord, So Hyun Kim, Shrikanth Narayanan

Federated Intelligent Terminals Facilitate Stuttering Monitoring
Yongzi Yu, Wanyong Qiu, Chen Quan, Kun Qian, Zhihua Wang, Yu Ma, Bin Hu, Björn W.
Schuller, Yoshiharu Yamamoto

Automatic Segmentation of Nasopharyngeal Carcinoma in CT Images Using Dual Attention and
Edge Detection
Qizhi Wang, Wei Huang, Yuan Zhang, Xuanya Li, Xiongjun Ye, Kai Hu
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Study of Audio Mixing Methods for Piano Transcription in Violin-Piano Ensembles</td>
<td>6226</td>
</tr>
<tr>
<td>Intent Does Matter! Propagating High-Order Relations for Exploring Interest Preferences</td>
<td>6231</td>
</tr>
<tr>
<td>Learning Scene Flow from 3D Point Clouds with Cross-Transformer and Global Motion Cues</td>
<td>6236</td>
</tr>
<tr>
<td>Bit Error and Block Error Rate Training for ML-Assisted Communication</td>
<td>6241</td>
</tr>
<tr>
<td>Disentangled and Robust Representation Learning for Bragging Classification in Social Media</td>
<td>6246</td>
</tr>
<tr>
<td>Whether Contribution of Features Differ Between Video-Mediated and In-Person Meetings in Important Utterance Estimation</td>
<td>6251</td>
</tr>
<tr>
<td>Spike-Based Optical Flow Estimation Via Contrastive Learning</td>
<td>6256</td>
</tr>
<tr>
<td>LABANet: Lead-Assisting Backbone Attention Network for Oral Multi-Pathology Segmentation</td>
<td>6261</td>
</tr>
<tr>
<td>Hybrid Neural Network with Cross-And Self-Module Attention Pooling for Text-Independent Speaker Verification</td>
<td>6266</td>
</tr>
<tr>
<td>Feature-Rich Audio Model Inversion for Data-Free Knowledge Distillation Towards General Sound Classification</td>
<td>6271</td>
</tr>
<tr>
<td>Cross-Modal Optical Flow Estimation Via Modality Compensation and Alignment</td>
<td>6276</td>
</tr>
<tr>
<td>Joint Data Association, NLOS Mitigation, and Clutter Suppression for Networked Device-Free Sensing in 6G Cellular Network</td>
<td>6281</td>
</tr>
<tr>
<td>Product Graph Learning from Multi-Attribute Graph Signals with Inter-Layer Coupling</td>
<td>6286</td>
</tr>
<tr>
<td>Commdre: Document-Level Relation Extraction with Self-Supervised Commonsense Learning</td>
<td>6291</td>
</tr>
<tr>
<td>TeAw: Text-Aware Few-Shot Remote Sensing Image Scene Classification</td>
<td>6296</td>
</tr>
<tr>
<td>A Bidirectional Joint Model for Spoken Language Understanding</td>
<td>6301</td>
</tr>
<tr>
<td>Learning from Label Proportion with Online Pseudo-Label Decision by Regret Minimization</td>
<td>6306</td>
</tr>
<tr>
<td>Blind Estimation of Audio Processing Graph</td>
<td>6311</td>
</tr>
</tbody>
</table>
Show Me the Instruments: Musical Instrument Retrieval from Mixture Audio ................................................. 6316
Kyungsu Kim, Minju Park, Haesun Joung, Yunkee Chae, Yeongbeom Hong, Seonghyeon Go, Kyogu Lee

Movienet-PS: A Large-Scale Person Search Dataset in the Wild ................................................................. 6321
Jie Qin, Peng Zheng, Yichao Yan, Rong Quan, Xiaogang Cheng, Bingbing Ni

LEAPT: Learning Adaptive Prefix-To-Prefix Translation for Simultaneous Machine Translation .................... 6326
Lei Lin, Shuangtao Li, Xiaodong Shi

Graph Signal Processing for Narrowband Direction of Arrival Estimation ..................................................... 6331
Disheng Li, Wei Liu, Yuriy Zakharov, Paul D Mitchell

Discriminative Speaker Representation Via Contrastive Learning with Class-Aware Attention in Angular Space .................................................................................................................. 6336
Zhe Li, Man-Wai Mak, Helen Mei-Ling Peng

SDTN: Speaker Dynamics Tracking Network for Emotion Recognition in Conversation .................................. 6341
Jiawei Chen, Peijie Huang, Guotai Huang, Qianer Li, Yuhong Xu

Long-Tailed Image Recognition with Dynamic Re-Weighting ........................................................................... 6346
Xinyuan Li, Yu Wang, Jien Kato

Adaptive Simulated Annealing Through Alternating Rényi Divergence Minimization .................................. 6351
Thomas Guilmeau, Emilie Chouzenoux, Victor Elvira

Neural Diarization with Non-Autoregressive Intermediate Attractors ......................................................... 6356
Yusuke Fujita, Tatsuya Komatsu, Robin Scheibler, Yusuke Kida, Tetsuji Ogawa

Monocular 3D Human Pose Estimation Based on Global Temporal-Attentive and Joints-Attention in Video ...................................................................................................................... 6361
Ruhan He, Shanshan Xiang, Peng Tao, Yongsheng Yu

Event-Based Visual Microphone ...................................................................................................................... 6366
Matthew Howard, Keigo Hirakawa

Domain and Language Adaptation Using Heterogeneous Datasets for Wav2vec2.0-Based Speech Recognition of Low-Resource Language .............................................................................. 6371
Kak Soky, Sheng Li, Chenhui Chu, Tatsuya Kawahara

Visual Graph Reasoning Network .................................................................................................................... 6376
Dingbang Li, Xin Lin, Haibin Cai, Wenzhou Chen

A Dynamic Cross-Scale Transformer with Dual-Compound Representation for 3D Medical Image Segmentation .......................................................................................................................... 6381
Ruixia Zhang, Zhiqiong Wang, Zhongyang Wang, Junchang Xin

Multilevel Transformer for Multimodal Emotion Recognition ....................................................................... 6386
Junyi He, Meimei Wu, Meng Li, Xiaobo Zhu, Feng Ye

Preformer: Predictive Transformer with Multi-Scale Segment-Wise Correlations for Long-Term Time Series Forecasting ................................................................................................................. 6391
Dazhao Du, Bing Su, Zhewei Wei

Spatial Cross-Attention for Transformer-Based Image Captioning .................................................................. 6396
Khoa Anh Ngo, Kyuhong Shim, Byonghyo Shim
Speaker-Aware Hierarchical Transformer for Personality Recognition in Multiparty Dialogues ........................................... 6401
Wenjing Han, Yirong Chen, Xiaofen Xing, Guohua Zhou, Xiangmin Xu

Ensemble Knowledge Distillation of Self-Supervised Speech Models ................................................................. 6406
Kuan -Po Huang, Tzu-Hsun Feng, Yu-Kuan Fu, Tsu-Yuan Hsu, Po-Chieh Yen, Wei-Cheng Tseng, Kai-Wei Chang, Hung-Yi Lee

Enhancing and Adversarial: Improve ASR with Speaker Labels ............................................................... 6411
Wei Zhou, Haoqian Wu, Jingjing Xu, Mohammad Zeineldeen, Christoph Lüscher, Ralf Schlüter, Hermann Ney

LDTSF: A Label-Decoupling Teacher-Student Framework for Semi-Supervised Echocardiography
Segmentation ................................................................................................................................................ 6416
Jiapeng Zhang, Yongxiong Wang, Zhiqun Pan, Zhenhui Tang, Lijun Chen, Jinlong Liu

Lexicon-Injected Semantic Parsing for Task-Oriented Dialog ........................................................................... 6421
Xiaojun Meng, Wenlin Dai, Yasheng Wang, Baojun Wang, Zhiyong Wu, Xin Jiang, Qun Liu

MASKED-AP: Attention Pyramid Convolutional Neural Network with Mask for Cervical Cell
Classification ........................................................................................................................................ 6426
Yu Jin, Juan Liu, Hua Chen, Wensi Duan, Dehua Cao, Baochuan Pang

On the Value of Stochastic Side Information in Online Learning .......................................................... 6431
Junzhang Jia, Xuetong Wu, Jamie Evans, Jingge Zhu

Frequency-Selective Hybrid Beamforming for Mmwave Full-Duplex ..................................................... 6436
Andrea Guamo-Morocho, Roberto López-Valcarce

Towards Hyperbolic Regularizers for Point Cloud Part Segmentation .............................................. 6441
Antonio Montanaro, Diego Valsesia, Enrico Magli

Self-Sufficient Framework for Continuous Sign Language Recognition ........................................... 6446
Youngjoon Jang, Youngtaek Oh, Jae Won Cho, Myungchul Kim, Dong-Jin Kim, In So Kweon, Joon Son Chung

Leveraging Large Text Corpora for End-To-End Speech Summarization ....................................... 6451
Kohei Matsuura, Takanori Ashihara, Takafumi Mori, Tomohiro Tanaka, Atsunori Ogawa, Marc Delcroix, Ryo Masumura

Topology Uncertainty Modeling for Imbalanced Node Classification on Graphs ................................................. 6456
Jiayi Gao, Jiaxing Li, Ke Zhang, Youyong Kong

A Dataset for Audio-Visual Sound Event Detection in Movies .......................................................... 6461
Rajat Hebbar, Digbalay Bose, Krishna Somandepalli, Veena Vijai, Shrikanth Narayanan

DDN: Dynamic Aggregation Enhanced Dual-Stream Network for Medical Image Classification .................. 6466
Lang Wang, Juan Liu, Peng Jiang, Dehua Cao, Baochuan Pang

Enhancing Speech-To-Speech Translation with Multiple TTS Targets .............................................. 6471
Jiatong Shi, Yunfang, Ann Lee, Hirofumi Inaguma, Changhan Wang, Juan Pino, Shinji Watanabe

CFFMixer: Multi-Dimensional Feature Fusion for Object Detection .................................................. 6476
Hao Xie, Weizhe Yuan, Bin Kang, Songlin Du
Real-Time Multichannel Speech Separation and Enhancement Using a Beamspace-Domain-Based Lightweight CNN

Marco Olivieri, Luca Comanducci, Mirco Pezzoli, Davide Balsarri, Luca Menescardi, Michele Buccoli, Simone Pecorino, Antonio Grosso, Fabio Antonacci, Augusto Sarti

Neural Mode Estimation

Peng Sun, Zhenyu Wen, Yejian Zhou, Zhen Hong, Tao Lin

Dynamic TF-TDNN: Dynamic Time Delay Neural Network Based on Temporal-Frequency Attention for Dialect Recognition

Chao Liao, Jinwen Huang, Huan Yuan, Peng Yao, Jianchao Tan, Dawei Zhang, Feng Deng, Xiaorui Wang, Chengru Song

Multi-Resolution Convolutional Dictionary Learning for Riverbed Dynamics Modeling

E. Kobayashi, H. Yasuda, K. Hayasaka, Y. Otake, S. Ono, S. Muramatsu

Grad-CAM-Inspired Interpretation of Nearfield Acoustic Holography Using Physics-Informed Explainable Neural Network

Hagar Kafri, Marco Olivieri, Fabio Antonacci, Mordehay Moradi, Augusto Sarti, Sharon Gannot

Double Compression Detection Based on the De-Blocking Filtering of HEVC Videos

Xiangui Kang, Pengcheng Su, Zisheng Huang, Yifang Chen, Jie Wang

Contrastive Learning of Functionality-Aware Code Embeddings

Yiyang Li, Hongqiu Wu, Hai Zhao

Smoothing Point Adjustment-Based Evaluation of Time Series Anomaly Detection

Mingyu Liu, Yijie Wang, Hongzuo Xu, Xiaohui Zhou, Bin Li, Yongjun Wang

Label-Efficient and Robust Learning from Multiple Experts

Bojan Kolosnjaji, Apostolis Zarras

General Or Specific? Investigating Effective Privacy Protection in Federated Learning for Speech Emotion Recognition

Chao Tan, Yang Cao, Sheng Li, Masatoshi Yoshikawa

Brain Network Features Differentiate Intentions from Different Emotional Expressions of the Same Text

Zhongjie Li, Bin Zhao, Gaoyan Zhang, Jianwu Dang

Diffusion-Based Sound Source Localization Using Networks of Planar Microphone Arrays

Davide Albertini, Gioele Greco, Alberto Bernardini, Augusto Sarti

Robust Knowledge Distillation from RNN-T Models with Noisy Training Labels Using Full-Sum Loss

Mohammad Zeineldeen, Kartik Audhkhasi, Murali Karthick Baskar, Bhuvana Ramabhadran

MGAT: Multi-Granularity Attention Based Transformers for Multi-Modal Emotion Recognition

Weiquan Fan, Xiaofen Xing, Bolun Cai, Xiangmin Xu

From English to More Languages: Parameter-Efficient Model Reprogramming for Cross-Lingual Speech Recognition

Chao-Han Huck Yang, Bo Li, Yu Zhang, Nanxin Chen, Rohit Prabhavalkar, Tara N. Sainath, Trevor Strohman
Wireless Sensing for Simultaneous Human Vocal Sound and Heart Sound Recognition
Yu Rong, Kumar Vijay Mishra, Daniel W. Bliss

Adaptive Gaussian Nested Filter for Parameter Estimation and State Tracking in Dynamical Systems
Sara Pérez-Vieites, Victor Elvira

-Complexity Low-Rank Approximation SVD for Massive Matrix in Tensor Train Format
Jung-Chun Chi, Chiao-En Chen, Yuan-Hao Huang

Fast-U2++: Fast and Accurate End-to-End Speech Recognition in Joint CTC/Attention Frames
Chengdong Liang, Xiao-Lei Zhang, Binbin Zhang, Di Wu, Shengqiang Li, Xingchen Song, Zhendong Peng, Fuping Pan

Wekws: A Production First Small-Footprint End-to-End Keyword Spotting Toolkit
Jie Wang, Menglong Xu, Jingyong Hou, Binbin Zhang, Xiao-Lei Zhang, Lei Xie, Fuping Pan

Image Inpainting with Semantic-Aware Transformer
Shiyu Chen, Wenzin Yu, Qi Wang, Jun Gong, Peng Chen

Vararray Meets T-Sot: Advancing the State of the Art of Streaming Distant Conversational Speech Recognition
Naoyuki Kanda, Jian Wu, Xiaofei Wang, Zhuo Chen, Jinyu Li, Takaya Yoshioka

Meta Learning with Adaptive Loss Weight for Low-Resource Speech Recognition
Qiulin Wang, Wenzuan Hu, Lin Li, Qingyang Hong

An Asynchronous Updating Reinforcement Learning Framework for Task-Oriented Dialog System
Sai Zhang, Yawei Hu, Xiaojie Wang, Caixia Yuan

Wave-U-Net Discriminator: Fast and Lightweight Discriminator for Generative Adversarial Network-Based Speech Synthesis
Takuhiro Kaneko, Hirokazu Kameoka, Kou Tanaka, Shogo Seki

An Application of Quantum Mechanics to Attention Methods in Computer Vision
Juntao Zhang, Yihao Luo, Peng Cheng, Zehan Li, Hao Wu, Kun Yu, Wenbo An, Jun Zhou

Robust Dominant Periodicity Detection for Time Series with Missing Data
Qingsong Wen, Linxiao Yang, Liang Sun

An Empirical Study of Backdoor Attacks on Masked Auto Encoders
Shuli Zhuang, Pengfei Xia, Bin Li

Joint Training of Hierarchical GANs and Semantic Segmentation for Expression Translation
Rumeysa Bodur, Binod Bhattarai, Tae-Kyun Kim

Deep Reinforcement Learning for Green UAV-Assisted Data Collection
Abhishek Mondal, Deepak Mishra, Ganesh Prasad, Ashraf Hossain

Delay-Aware Backpressure Routing Using Graph Neural Networks
Zhongyuan Zhao, Bojan Radijicic, Gunjan Verma, Ananthram Swami, Santiago Segarra

Deep Feature Aggregation for Lightweight Single Image Super-Resolution
Yanchun Li, Xinan He, Shujuan Tian, Zhetao Li, Saiqin Long

Class-Guided Triple Head Prediction Network for Long-Tail Object Detection
Xuyang Liu, Yuan Zheng
Sparse Bayesian Learning Based Three-Dimensional Imaging for Antenna Array Radar ........................................ 6646
Yuhan Li, Jesper Rindom Jensen, Maozhong Fu, Zhenmiao Deng, Mads Græsbøll Christensen

MCNET:Fuse Multiple Cues for Multichannel Speech Enhancement ................................................................. 6651
Yujie Yang, Changsheng Quan, Xiaofei Li

Structural Optimization of Factor Graphs for Symbol Detection Via Continuous Clustering and Machine Learning .......................................................................................................................... 6656
Lukas Rapp, Luca Schmid, Andrey Rode, Laurent Schmalen

TOLD: A Novel Two-Stage Overlap-Aware Framework for Speaker Diarization .................................................... 6661
Jiaming Wang, Zhihao Du, Shiliang Zhang

Domain Adaptation Without Catastrophic Forgetting on a Small-Scale Partially-Labeled Corpus for Speech Emotion Recognition .............................................................................................................. 6666
Zhi Zhu, Yoshinao Sato

Codes Correcting Burst and Arbitrary Erasures for Reliable and Low-Latency Communication ......................... 6671
Serge Kas Hanna, Zhiyuan Tan, Wen Xu, Antonia Wachter-Zeh

Logo-Former: Local-Global Spatio-Temporal Transformer for Dynamic Facial Expression Recognition ................................................................. 6676
Fuyan Ma, Bin Sun, Shutao Li

A Dual-Path Transformer Network for Scene Text Detection ..................................................................................... 6681
Jingyu Lin, Yan Yan, Hanzi Wang

Capacity Maximization for Active RIS Assisted Outdoor-To-Indoor Communication System .......................... 6686
Chen He, Weisheng Gong, Yangrui Dong, Xie Xie, Z. Jane Wang

A Multi-Channel Aggregation Framework for Object Detection in Large-Scale SAR Image ............................. 6691
Chule Yang, Chao Zhang, Zunlin Fan, Zeting Yu, Qianchong Sun, Mengyuan Dai

Speech-Based Emotion Recognition with Self-Supervised Models Using Attentive Channel-Wise Correlations and Label Smoothing ........................................................................................................ 6696
Sofoklis Kakouros, Themos Stafylakis, Ladislav Mošner, Lukáš Burget

JSV-VC: Jointly Trained Speaker Verification and Voice Conversion Models ...................................................... 6701
Shogo Seki, Hirokazu Kameoka, Kou Tanaka, Takuhiro Kaneko

How to Push the Fastest Model 50x Faster: Streaming Non-Autoregressive Speech Synthesis on Resource-Limited Devices .............................................................................................................. 6706
Van-Thinh Nguyen, Hung-Cuong Pham, Dang-Khoa Mac

Hierarchical Spatiotemporal Feature Fusion Network for Video Saliency Prediction ........................................... 6711
Yunzuo Zhang, Tian Zhang, Cunyu Wu, Yuxin Zheng

Weavspeech: Data Augmentation Strategy for Automatic Speech Recognition Via Semantic-Aware Weaving ............................................................................................................................................. 6716
Kyusung Seo, Joonhyung Park, Jaeyun Song, Eunho Yang

An Automotive Radar Dataset for Object Classification .......................................................................................... 6721
Akshad Shyam, Kusum Komalavally, Monika Gautam, Yamshikrishna Kancharla, Vennela Gudisa, Virendra Patil, Aanandh Balasubramanian, Sumohana Channappayya
Joint Microstrip Selection and Beamforming Design for MmWave Systems with Dynamic Metasurface Antennas ................................................................. 6726
  Wei Huang, Haiyang Zhang, Nir Shlezinger, Yonina C. Eldar

Subspace Modeling Enabled High-Sensitivity X-Ray Chemical Imaging ........................................... 6731
  Jizhou Li, Bin Chen, Guibin Zan, Guanman Qian, Piero Pianetta, Yijin Liu

Multi-Carrier Wideband OCDM-Based THZ Automotive Radar .................................................. 6736
  Sangeeta Bhattacharjee, Kumar Vijay Mishra, Ramesh Annavajjala, Chandra R. Murthy

Body Prior Guided Graph Convolutional Neural Network for Skeleton-Based Action Recognition ...... 6741
  Qianshuo Hu, Hong Liu, Hua-Qiu Wang, Mengyuan Liu

Real-Time MRI Video Synthesis from Time Aligned Phonemes with Sequence-to-Sequence Networks .......................................................... 6746
  Sathvik Udupa, Prasanta Kumar Ghosh

A Highly Interpretable Deep Equilibrium Network for Hyperspectral Image Deconvolution ........ 6751
  Alexandros Gkillas, Dimitris Ampeliotis, Kostas Berberidis

Improved Acoustic-To-Articulatory Inversion Using Representations from Pretrained Self-Supervised Learning Models .................................................. 6756
  Sathvik Udupa, Siddarth C, Prasanta Kumar Ghosh

MPE4G : Multimodal Pretrained Encoder for Co-Speech Gesture Generation ................................ 6761
  Gwantae Kim, Seonghyeok Noh, Insung Ham, Hanseok Ko

Dynamic Split Computing for Efficient Deep EDGE Intelligence ................................................. 6766
  Arian Bakhtiarnia, Nemanja Mišević, Qi Zhang, Dragana Bajović, Alexandros Iosifidis

Multi-Agent Adversarial Training Using Diffusion Learning ......................................................... 6771
  Ying Cao, Elsa Rizk, Stefan Vlaski, Ali H. Sayed

Efficient Super-Resolution for Compression of Gaming Videos ................................................... 6776
  Yifan Wang, Luka Murn, Luis Herranz, Fei Yang, Marta Mrak, Wei Zhang, Shuai Wan, Marc Gorriz Blanch

Peak-First CTC: Reducing the Peak Latency of CTC Models by Applying Peak-First Regularization .................................................. 6781
  Zhengkun Tian, Hongyu Xiang, Min Li, Feifei Lin, Ke Ding, Guanglu Wan

From Easy to Hard: Two-Stage Selector and Reader for Multi-Hop Question Answering ................. 6786
  Xin-Yi Li, Wei-Jun Lei, Yu-Bin Yang

Sparse Asynchronous Samples from Networks of Tems for Reconstruction of Classes of Non-Bandlimited Signals .................................................. 6791
  Marek Hilton, Pier Luigi Dragotti

Time-Weighted Frequency Domain Audio Representation with GMM Estimator for Anomalous Sound Detection .................................................. 6796
  Jian Guan, Youde Liu, Qiaoxi Zhu, Tieran Zheng, Jiqing Han, Wenwu Wang

DyLiteRADHAR: Dynamic Lightweight Slowfast Network for Human Activity Recognition Using MMWA VE Radar .................................................. 6801
  Biyun Sheng, Yan Bao, Fu Xiao, Lingqing Gu
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereoscopic Video Retargeting Based on Camera Motion Classification</td>
<td>6806</td>
</tr>
<tr>
<td>Linghui Cai, Zhenhua Tang</td>
<td></td>
</tr>
<tr>
<td>Pseudo Multi-Source Domain Extension and Selective Pseudo-Labeling for Unsupervised Domain Adaptive Medical Image Segmentation</td>
<td>6811</td>
</tr>
<tr>
<td>Xiaokang Liu, Zhiqiang Wang, Kai Hu, Xieping Gao</td>
<td></td>
</tr>
<tr>
<td>Joint Unsupervised and Supervised Learning for Context-Aware Language Identification</td>
<td>6816</td>
</tr>
<tr>
<td>Jinseok Park, Hyung Yong Kim, Jihwan Park, Byeong-Yeol Kim, Shukjae Choi, Yunkyu Lim</td>
<td></td>
</tr>
<tr>
<td>Adapted Multimodal Bert with Layer-Wise Fusion for Sentiment Analysis</td>
<td>6821</td>
</tr>
<tr>
<td>Odysseas S. Chlapanis, Georgios Paraskevopoulos, Alexandros Potamianos</td>
<td></td>
</tr>
<tr>
<td>Knowledge-Augmented Frame Semantic Parsing with Hybrid Prompt-Tuning</td>
<td>6826</td>
</tr>
<tr>
<td>Rui Zhang, Yajing Sun, Jingyuan Yang, Wei Peng</td>
<td></td>
</tr>
<tr>
<td>Energy Efficiency Maximization in RIS-Aided Networks with Global Reflection Constraints</td>
<td>6831</td>
</tr>
<tr>
<td>Robert K. Fotock, Alessio Zappone, Marco Di Renzo</td>
<td></td>
</tr>
<tr>
<td>Joint Robust Representation and Generalization Enhancement for Cross-Modality Person Re-Identification</td>
<td>6836</td>
</tr>
<tr>
<td>Heqing Cheng, Yong Feng, Mingliang Zhou, Xiancai Xiong, Yongheng Wang, Baohua Qiang</td>
<td></td>
</tr>
<tr>
<td>Subspace-Based Detector for Distributed Mmwave Mimo Radar Sensors</td>
<td>6841</td>
</tr>
<tr>
<td>Moein Ahmadi, Mohammad Alae-Kerahroodi, Bhavani Shankar M. R., Björn Ottersten</td>
<td></td>
</tr>
<tr>
<td>Unsupervised Feature Selection with self-Weighted and \ell_{2,0}-Norm Constraint</td>
<td>6846</td>
</tr>
<tr>
<td>Yongjin Yuan, Zheng Wang, Feiping Nie, Xuelong Li</td>
<td></td>
</tr>
<tr>
<td>Revisit Sampling Theory of Bandlimited Graph Signals: One Bridge Between GSP and DSP</td>
<td>6851</td>
</tr>
<tr>
<td>Fen Wang, Taihao Li, Xue Zhang</td>
<td></td>
</tr>
<tr>
<td>Multistage Spatial Context Models for Learned Image Compression</td>
<td>6856</td>
</tr>
<tr>
<td>Fangzheng Lin, Heming Sun, Jimming Liu, Jiro Katto</td>
<td></td>
</tr>
<tr>
<td>Electric Network Frequency Detection Using Least Absolute Deviations</td>
<td>6861</td>
</tr>
<tr>
<td>Christos Korgialas, Constantine Kotropoulos</td>
<td></td>
</tr>
<tr>
<td>D-3DLD: Depth-Aware Voxel Space Mapping for Monocular 3D Lane Detection with Uncertainty</td>
<td>6866</td>
</tr>
<tr>
<td>Nayeon Kim, Moonsub Byeon, Daehyun Ji, Dokwan Oh</td>
<td></td>
</tr>
<tr>
<td>Unsupervised Pre-Training for Data-Efficient Text-to-Speech on Low Resource Languages</td>
<td>6871</td>
</tr>
<tr>
<td>Seongyeon Park, Myungsseo Song, Bohyung Kim, Tae-Hyun Oh</td>
<td></td>
</tr>
<tr>
<td>Sensor Selection for Angle of Arrival Estimation Based on the Two-Target Cramér-Rao Bound</td>
<td>6876</td>
</tr>
<tr>
<td>Costas A. Kokke, Mario Coutino, Laura Anitori, Richard Heusdens, Geert Leus</td>
<td></td>
</tr>
<tr>
<td>Deep Autoencoding One-Class Time Series Anomaly Detection</td>
<td>6881</td>
</tr>
<tr>
<td>Xudong Mou, Rui Wang, Tiejun Wang, Jie Sun, Bo Li, Tianyu Wo, Xudong Liu</td>
<td></td>
</tr>
<tr>
<td>Local-Global Siamese Network with Efficient Inter-Scale Feature Learning for Change Detection in VHR Remote Sensing Images</td>
<td>6886</td>
</tr>
<tr>
<td>Yue Zhang, Tao Lei, Shaoxiong Han, Yetong Xu, Asoke K. Nandi</td>
<td></td>
</tr>
<tr>
<td>Signal Reconstruction for FMCW Radar Interference Mitigation Using Deep Unfolding</td>
<td>6891</td>
</tr>
</tbody>
</table>
Shuffleaugment: A Data Augmentation Method Using Time Shuffling ................................. 6896
Yoshinao Sato, Narumitsu Ikeda, Hirokazu Takahashi

F0 Estimation from Telephone Speech Using Deep Feature Loss ........................................ 6901
Supritha M Shetty, Shraddha Revankar, Nalini C Iyer, K T Deepak

Multiresolution Signal Processing of Financial Market Objects ....................................... 6906
Ioana Boier

Learning Hypergraphs from Signals with Dual Smoothness Prior ...................................... 6911
Bohan Tang, Siheng Chen, Xiaowen Dong

Wavsyncswap: End-To-End Portrait-Customized Audio-Driven Talking Face Generation ........ 6916
Weihong Bao, Liyang Chen, Chaoyong Zhou, Sicheng Yang, Zhiyong Wu

Duration-Aware Pause Insertion Using Pre-Trained Language Model for Multi-Speaker Text-To-
Speech ............................................................................................................................... 6921
Dong Yang, Tomoki Koriyama, Yuki Saito, Takaaki Saeki, Detai Xin, Hiroshi Saruwatari

WUDA: Unsupervised Domain Adaptation Based on Weak Source Domain Labels ............... 6926
Shengjie Liu, Chuang Zhu, Yuan Li, Wenqi Tang

Fine-Grained Private Knowledge Distillation ....................................................................... 6931
Yuntong Li, Shaowei Wang, Yingying Wang, Jin Li, Yuqiu Qian, Bangzhou Xin, Wei Yang

Two-Stage UNet with Multi-Axis Gated Multilayer Perceptron for Monaural Noisy-Reverberant
Speech Enhancement ........................................................................................................... 6936
Zehua Zhang, Shiyun Xu, Xiuyi Zhuang, Lianyu Zhou, Heng Li, Mingjiang Wang

Frequency and Scale Perspectives of Feature Extraction ................................................... 6941
Liangqi Zhang, Yihao Luo, Xiang Cao, Haibo Shen, Tianjiang Wang

Temporal Modeling Matters: A Novel Temporal Emotional Modeling Approach for Speech
Emotion Recognition ......................................................................................................... 6946
Jiaxin Ye, Xin-Cheng Wen, Yujie Wei, Yong Xu, Kunhong Liu, Hongming Shan

Once-For-All Sequence Compression for Self-Supervised Speech Models ......................... 6951
Hsuan-Jui Chen, Yen Meng, Hung-Yi Lee

Multitrack Music Transformer .............................................................................................. 6956
Hao-Wen Dong, Ke Chen, Shlomo Dubnov, Julian McAuley, Taylor Berg-Kirkpatrick

Receptive Field Reliant Zero-Cost Proxies for Neural Architecture Search ......................... 6961
Prateek Keserwani, Srinivas Soumithri Miriyala, Vikram N. Rajendiran, Pradeep N.
Shivamurthappa

SMCL: Saliency Masked Contrastive Learning for Long-Tailed Visual Recognition ............... 6966
Sanglee Park, Seung-Won Hwang, Jungmin So

Multi-Dimensional and Multi-Scale Modeling for Speech Separation Optimized by Discriminative
Learning .............................................................................................................................. 6971
Zhaoxi Mu, Xinyu Yang, Wenjing Zhu

RIS Reflection and Placement Optimisation for Underlay D2D Communications in Cognitive
Cellular Networks ............................................................................................................... 6976
Sarbani Ghose, Deepak Mishra, Santi P. Maity, George C. Alexandropoulos
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autonomous Navigation of a Robotic Swarm in Space Exploration Missions</td>
<td>6981</td>
</tr>
<tr>
<td>Siwei Zhang, Tobias Baumgartner, Emanuel Staudinger, Robert Pöhlmann, Fabio Broghammer, Armin Dammann</td>
<td></td>
</tr>
<tr>
<td>Prosody-Controllable Spontaneous TTS with Neural HMMS</td>
<td>6986</td>
</tr>
<tr>
<td>Harm Lameris, Shivam Mehta, Gustav Eje Henter, Joakim Gustafson, Éva Székely</td>
<td></td>
</tr>
<tr>
<td>Misspecified Cramér-Rao Bound of RIS-Aided Localization Under Geometry Mismatch</td>
<td>6991</td>
</tr>
<tr>
<td>Pinjun Zheng, Hui Chen, Tariq Ballal, Henk Wyneersch, Tareq Y. Al-Naffouri</td>
<td></td>
</tr>
<tr>
<td>Noise-Disentanglement Metric Learning for Robust Speaker Verification</td>
<td>6996</td>
</tr>
<tr>
<td>Yao Sun, Hanyi Zhang, Longbiao Wang, Kong Aik Lee, Meng Liu, Jianwu Dang</td>
<td></td>
</tr>
<tr>
<td>Gated Enhanced RPN and Hybrid-View for Few-Shot Object Detection</td>
<td>7001</td>
</tr>
<tr>
<td>Xujun Wei, Zechu Zhou, Pinxue Guo, Wenhui Zhang</td>
<td></td>
</tr>
<tr>
<td>Self-Supervised Audio-Visual Speaker Representation with Co-Meta Learning</td>
<td>7006</td>
</tr>
<tr>
<td>Hui Chen, Hanyi Zhang, Longbiao Wang, Kong Aik Lee, Meng Liu, Jianwu Dang</td>
<td></td>
</tr>
<tr>
<td>Training Sound Event Detection with Soft Labels from Crowdsourced Annotations</td>
<td>7011</td>
</tr>
<tr>
<td>Irene Martín-Morató, Manu Harju, Paul Ahokas, Annamarie Mesaros</td>
<td></td>
</tr>
<tr>
<td>DialogMI: A Dialogue Model Based on Enhancing Dialogue Mutual Information</td>
<td>7016</td>
</tr>
<tr>
<td>Yibo Zhang, Ping Gong, Zelin Wang, Zhe Li, Xuanzhu Yang</td>
<td></td>
</tr>
<tr>
<td>Inter-Scale Sure-Let Denoise with Structured Deep Image Prior: Interpretable Self-Supervised Learning</td>
<td>7021</td>
</tr>
<tr>
<td>Jikai Li, Shogo Muramatsu</td>
<td></td>
</tr>
<tr>
<td>Multichannel Time-Encoding of Finite-Rate-of-Innovation Signals</td>
<td>7026</td>
</tr>
<tr>
<td>Abijith Jagannath Kamath, Chandra Sekhar Seelamantula</td>
<td></td>
</tr>
<tr>
<td>Autovocoder: Fast Waveform Generation from a Learned Speech Representation Using Differentiable Digital Signal Processing</td>
<td>7031</td>
</tr>
<tr>
<td>Jacob J Webber, Cassia Valentini-Botinhao, Evelyn Williams, Gustav Eje Henter, Simon King</td>
<td></td>
</tr>
<tr>
<td>An Experimental Study on Sound Event Localization and Detection Under Realistic Testing Conditions</td>
<td>7036</td>
</tr>
<tr>
<td>Shutong Niu, Jun Du, Qing Wang, Li Chai, Huaxin Wu, Zhaoxu Nian, Lei Sun, Yi Fang, Jia Pan, Chin-Hui Lee</td>
<td></td>
</tr>
<tr>
<td>Feature Selection and Text Embedding for Detecting Dementia from Spontaneous Cantonese</td>
<td>7041</td>
</tr>
<tr>
<td>Xiaoquan Ke, Man-Wai Mak, Helen M. Meng</td>
<td></td>
</tr>
<tr>
<td>Transmit Energy Focusing for Parameter Estimation in Transmit Beamspace Slow-Time MIMO Radar</td>
<td>7046</td>
</tr>
<tr>
<td>Tingting Zhang, Feng Xu, Sergiy A. Vorobyov</td>
<td></td>
</tr>
<tr>
<td>Vehicle View Synthesis by Generative Adversarial Network</td>
<td>7051</td>
</tr>
<tr>
<td>Chan-Shuo Hu, Sung-Wei Tseng, Xin-Yun Fan, Chen-Kuo Chiang</td>
<td></td>
</tr>
<tr>
<td>Meeting Action Item Detection with Regularized Context Modeling</td>
<td>7056</td>
</tr>
<tr>
<td>Jiaqing Liu, Chong Deng, Qinglin Zhang, Qian Chen, Wen Wang</td>
<td></td>
</tr>
<tr>
<td>A Momentum Two-Gradient Direction Algorithm with Variable Step Size Applied to Solve Practical Output Constraint Issue for Active Noise Control</td>
<td>7061</td>
</tr>
<tr>
<td>Xiaoyi Shen, Dongyan Shi, Zhengding Luo, Junwei Ji, Woon-Seng Gan</td>
<td></td>
</tr>
</tbody>
</table>
General Category Network: Handwritten Mathematical Expression Recognition with Coarse-Grained Recognition Task ................................................................. 7066
Xinyu Zhang, Han Ying, Ye Tao, Youlu Xing, Guihuan Feng

Attention Mixup: An Accurate Mixup Scheme Based on Interpretable Attention Mechanism for Multi-Label Audio Classification ................................................. 7071
Wuyang Liu, Yanzhen Ren, Jingru Wang

MCNeT: Measurement-Consistent Networks Via a Deep Implicit Layer for Solving Inverse Problems .................................................................................. 7076
Rahul Mourya, João F. C. Mota

Efficient Privacy Preserving Graph Neural Network for Node Classification ................................................................. 7081
Xinjun Pei, Xiaoheng Deng, Shengwei Tian, Kaiping Xue

Subgradient Descent Learning with Over-The-Air Computation .................................................................................. 7086
Tamir L. S. Gez, Kobi Cohen

Speech and Noise Dual-Stream Spectrogram Refine Network with Speech Distortion Loss for Robust Speech Recognition ........................................................................ 7091
Haoyu Lu, Nan Li, Tongtong Song, Longbiao Wang, Jianwu Dang, Xiaobao Wang, Shiliang Zhang

Joint Unmixing and Demosaicing Methods for Snapshot Spectral Images ........................................................................ 7096
Kinan Abbas, Matthieu Puigt, Gilles Delmaire, Gilles Roussel

SR-Init: An Interpretable Layer Pruning Method ........................................................................................................ 7101
Hai Tang, Yao Lu, Qi Xuan

On Cross-Layer Alignment for Model Fusion of Heterogeneous Neural Networks ................................................................. 7106
Dang Nguyen, Trang Nguyen, Khai Nguyen, Dinh Phung, Hung Bui, Nhat Ho

Zone Plate Virtual Lenses for Memory-Constrained NLOS Imaging .................................................................................. 7111
Pablo Luesia-Lahoz, Diego Gutierrez, Adolfo Muñoz

Sparse Non-Contact Multiple People Localization and Vital Signs Monitoring Via FMCW Radar ........................................ 7116
Yonathan Eder, Zhuoyang Liu, Yonina C. Eldar

Knowledge Transfer for on-Device Speech Emotion Recognition with Neural Structured Learning ........................................ 7121
Yi Chang, Zhao Ren, Thanh Tam Nguyen, Kun Qian, Björn W. Schuller

Convolutional Recurrent Neural Networks for the Classification of Cetacean Bioacoustic Patterns ........................................ 7126
Dimitris N. Makropoulos, Antigoni Tsiami, Aristides Prospathopoulos, Dimitris Kassis, Alexandros Frantzi, Emmanuel Skarsoulis, George Piperakis, Petros Maragos

Joint Discriminator and Transfer Based Fast Domain Adaptation for End-To-End Speech Recognition .................................................................................. 7131
Hang Shao, Tian Tan, Wei Wang, Xun Gong, Yanmin Qian

Neural Source Coding for Bandwidth-Efficient Brain-Computer Interfacing with Wireless Neuro-Sensor Networks ........................................................................ 7136
Thomas Strypsteen, Alexander Bertrand

Clustered Greedy Algorithm for Large-Scale Sensor Selection ................................................................................................. 7141
Kaushani Majumder, Sibi Raj B. Pillai, Satish Mulleti
End-To-End Classification of Cell-Cycle Stages with Center-Cell Focus Tracker Using Recurrent Neural Networks................................................................. 7146
Abin Jose, Rijo Roy, Dennis Eschweiler, Ina Laube, Reza Azad, Daniel Moreno-Andrés, Johannes Stegmaier

Frequency Bin-Wise Single Channel Speech Presence Probability Estimation Using Multiple DNNS........................................................................................................ 7151
Shuai Tao, Himavanth Reddy, Jesper Rindom Jensen, Mads Grøsbøll Christensen

SYNTACC : Synthesizing Multi-Accent Speech by Weight Factorization................................................................. 7156
Tuan-Nam Nguyen, Ngoc-Quan Pham, Alexander Waibel

Mutually Guided Few-Shot Learning for Relational Triple Extraction........................................................................... 7161
Chengmei Yang, Shuai Jiang, Bowei He, Chen Ma, Lianghua He

Graph Contrastive Learning with Learnable Graph Augmentation..................................................................................... 7166
Xinyan Pu, Ke Zhang, Huazhong Shu, Jean Louis Coatrieux, Youyong Kong

Graph Signal Processing for Neuroimaging to Reveal Dynamics of Brain Structure-Function Coupling................................................................. 7171
Maria Giulia Preti, Thomas A. W. Bolton, Alessandra Griffa, Dimitri Van De Ville

Modelling Low-Resource Accents Without Accent-Specific TTS Frontend........................................................................... 7176
Georgi Tinchev, Marta Czarnowska, Kamil Deja, Kayoko Yanagisawa, Marius Cotescu

DO-FAM: Disentangled Non-Linear Latent Navigation for Facial Attribute Manipulation ................................................................. 7181
Yifan Yuan, Sieng Ma, Hongming Shan, Junping Zhang

Newton-Based Trainable Learning Rate................................................................................................................................. 7186
George Retsinas, Giorgos Sfikas, Panagiotis Paraskevas Filntisis, Petros Maragos

Stochastic Super-Resolution for Gaussian Textures ..................................................................................................................... 7191
Émile Pierret, Bruno Galerne

Semantically-Informed Deep Neural Networks for Sound Recognition ........................................................................... 7196
Michele Esposito, Giancarlo Valente, Yenisel Plasencia-Calaña, Michel Dumontier, Bruno L. Giordano, Elia Formisano

F-PABEE: Flexible-Patience-Based Early Exiting for Single-Label and Multi-Label Text Classification Tasks................................................................. 7201
Xiangxiang Gao, Wei Zhu, Jiasheng Gao, Congrui Yin

Hierarchical Hypergraph Recurrent Attention Network for Temporal Knowledge Graph Reasoning ........... 7206
Jiayan Guo, Meiqi Chen, Yan Zhang, Jianqiang Huang, Zhiwei Liu

Topgformer: Topological-Based Graph Transformer for Mapping Brain Structural Connectivity to Functional Connectivity ............................................................................ 7211
Dalu Guo, Ke Zhang, Jiaxing Li, Youyong Kong

MoLE : Mixture of Language Experts for Multi-Lingual Automatic Speech Recognition........................................ 7216
Yoohwan Kwon, Soo-Whan Chung

Source-Filter HiFi-GAN: Fast and Pitch Controllable High-Fidelity Neural Vocoder.............................................. 7221
Reo Toneyama, Yi-Chiao Wu, Tomoki Toda

Analysis of Noisy-Target Training for Dnn-Based Speech Enhancement........................................................................ 7226
Takuya Fujimura, Tomoki Toda
Constrained Non-Negative PARAFAC2 for Electromyogram Separation ........................................ 7231
Abilé Magbonde, Franck Quaine, Bertrand Rivet

Rethinking the Reasonability of the Test Set for Simultaneous Machine Translation .......................... 7236
Mengge Liu, Wen Zhang, Xiang Li, Jian Luan, Bin Wang, Yuhang Guo, Shuoying Chen

Streaming Joint Speech Recognition and Disfluency Detection ...................................................... 7241
Hayato Futami, Emiru Tsunoo, Kentaro Shibata, Yosuke Kashiwagi, Takao Okuda, Siddhant Arora, Shinji Watanabe

Efficient Practices for Profile-To-Frontal Face Synthesis and Recognition ....................................... 7246
Huijiao Wang, Xulei Yang

Distance-Based Online Label Inference Attacks Against Split Learning ........................................... 7251
Junlin Liu, Xinchen Lyu

Dereverberation in Acoustic Sensor Networks Using Weighted Prediction Error with Microphone-Dependent Prediction Delays ................................................................. 7256
Anselm Lohmann, Toon Van Waterschoot, Joerg Bitzer, Simon Doclo

Towards Adversarially Robust Continual Learning ................................................................. 7261
Tao Bai, Chen Chen, Lingjuan Lyu, Jun Zhao, Bihan Wen

Learning Causal Representations for Generalizable Face Anti Spoofing ......................................... 7266
Guanghao Zheng, Yuchen Liu, Wenrui Dai, Chenglin Li, Junni Zou, Hongkai Xiong

Speech Emotion Recognition Based on Low-Level Auto-Extracted Time-Frequency Features .......... 7271
Ke Liu, Jingzhao Hu, Jun Feng

A Principled Approach to Model Validation in Domain Generalization ........................................... 7276
Boyang Lyu, Thuan Nguyen, Matthias Scheutz, Prakash Ishwar, Shuchin Aeron

X-SEPFORMER: End-To-End Speaker Extraction Network with Explicit Optimization on Speaker Confusion ................................................................. 7281
Kai Liu, Ziqing Du, Xucheng Wan, Huan Zhou

Speakeraugment: Data Augmentation for Generalizable Source Separation Via Speaker Parameter Manipulation ................................................................. 7286
Kai Wang, Yuhang Yang, Hao Huang, Ying Hu, Sheng Li

SW-WAVENET: Learning Representation from Spectrogram and Wavegram Using Wavenet for Anomalous Sound Detection ................................................................. 7291
Haihui Chen, Likai Ran, Xixia Sun, Chao Cai

Scale-Adaptive Tiny Object Detection Enhanced by Across-Scale and Shape-Preserved Semantic Location .............................................................................. 7296
Yuting He, Renjie Huang, Yangguang Shi, Guoqiang Xiao, Bin Yang, Yuqi Li

Enlightening the Student in Knowledge Distillation ........................................................................... 7301
Yujie Zheng, Chong Wang, Yi Chen, Jiangbo Qian, Jun Wang, Jiafei Wu

D-CONFORMER: Deformable Sparse Transformer Augmented Convolution for Voxel-Based 3D Object Detection .............................................................................. 7306
Xiao Zhao, Luiwen Su, Xukun Zhang, Dingkang Yang, Mingyang Sun, Shunli Wang, Peng Zhai, Lihua Zhang
Adaptive Non-Local Generative Adversarial Networks for Low-Dose CT Image Denoising .......................... 7311
Linlin Yang, Hongying Liu, Fanhua Shang, Yuanyuan Liu

Robust Binaural Sound Localisation with Temporal Attention .................................................. 7316
Qi Hu, Ning Ma, Guy J. Brown

Near-Field Localization with Dynamic Metasurface Antennas .................................................. 7321
Qianyu Yang, Anna Guerra, Francesco Guidi, Nir Shlezinger, Haiyang Zhang, Davide Dardari, Baoyun Wang, Yonina C. Eldar

NRTSI: Non-Recurrent Time Series Imputation ............................................................................. 7326
Siyuan Shan, Yang Li, Junier B. Oliva

Hierarchical Interactive Reconstruction Network for Video Compressive Sensing ......................... 7331
Tong Zhang, Wenxue Cui, Chen Hui, Feng Jiang

Time-Frequency Awareness Network for Human Mesh Recovery from Videos ............................ 7336
Boyang Zhang, Suping Wu, Meining Jia

Cardiac Disease Diagnosis on Imbalanced Electrocardiography Data Through Optimal Transport Augmentation .................................................................................................................. 7341
Jielin Qiu, Jiacheng Zhu, Mengdi Xu, Peide Huang, Michael Rosenberg, Douglas Weber, Emerson Liu, Ding Zhao

Estimating Uncertainty on Video Quality Metrics .......................................................................... 7346
Pierre David, Patrick Le Callet, Suiyi Ling, Haixiong Wang, Ioannis Katsavounidis, Zafar Shahid, Cosmin Stejerean

Hiding Speaker’s Sex in Speech Using Zero-Evidence Speaker Representation in an Analysis/Synthesis Pipeline .................................................................................................................. 7351
Paul-Gauthier Noé, Xiaoxiao Miao, Xin Wang, Junichi Yamagishi, Jean-François Bonastre, Driss Matrouf

Predictive Skim: Contrastive Predictive Coding for Low-Latency Online Speech Separation ............. 7356
Chenda Li, Yifei Wu, Yanmin Qian

Topological Slepians: Maximally Localized Representations of Signals Over Simplicial Complexes .... 7361
Claudio Battiloro, Paolo Di Lorenzo, Sergio Barbarossa

Quaternion Orthogonal Transformer for Facial Expression Recognition in the Wild ......................... 7366
Yu Zhou, Liyuan Guo, Lianghai Jin

Integrating the Sensing and Radio Communications Channel Modelling from Radar Mutual Interference ............................................................................................................................ 7371
Narcis Cardona, J. Samuel Romero, Wenfei Yang, Jian Li

HyperSteg: Hyperbolic Learning for Deep Steganography .............................................................. 7376
Shivam Agarwal, Ritesh Soun, Rahul Shivani, Vishnu Varanasi, Navroop Gill, Ramit Sawhney

Gaze Pre-Train for Improving Disparity Estimation Networks ......................................................... 7381
Ron M. Hecht, Ohad Rahamin, Shaul Oron, Andrea Forgacs, Gershon Celniker, Dan Levi, Omer Tsimhoni

Color Guided Depth Map Super-Resolution with Nonlocla Autoregres-Sive Modeling .................... 7386
Wei Xu, Na Qi, Qing Zhu, Jingzhong Qi, Longlu Huang, Kun Cao, Yuxin Bao, Qianwen Wang
On Parametric Misspecified Bayesian Cramér-Rao Bound: An Application to Linear/Gaussian Systems................................................................. 7391
  Shuo Tang, Gerald Lamontain, Tales Imbiriba, Pau Closas

A Novel Efficient Multi-View Traffic-Related Object Detection Framework................................................................. 7396
  Kun Yang, Jing Liu, Dingkang Yang, Hanqi Wang, Peng Sun, Yanni Zhang, Yan Liu, Liang Song

Pseudo-Inverted Bottleneck Convolution for Darts Search Space ................................................................. 7401
  Arash Ahmadian, Louis S. P. Liu, Yue Fei, Konstantinos N. Plataniotis, Mahdi S. Hosseini

Multi-Head Feature Pyramid Networks for Breast Mass Detection................................................................. 7406
  Hexiang Zhang, Zhenghua Xu, Dan Yao, Shuo Zhang, Junyang Chen, Thomas Lukasiewicz

On Neural Architectures for Deep Learning-Based Source Separation of Co-Channel OFDM Signals ................................................................. 7411
  Gary C. F. Lee, Amir Weiss, Alejandro Lancho, Yury Polyanskiy, Gregory W. Wornell

A Sidecar Separator Can Convert a Single-Talker Speech Recognition System to a Multi-Talker One ................................................................. 7416
  Lingwei Meng, Jiawen Kang, Mingyu Cui, Yuejiao Wang, Xixin Wu, Helen Meng

Learning from Single-Expert Annotated Labels for Automatic Sleep Staging................................................................. 7421
  Zhiheng Luan, Yanzhen Ren, Li Peng, Xiong Chen, Xiuping Yang, Weiping Tu, Yuhong Yang

Label-Guided Contrastive Learning for Out-of-Domain Detection ................................................................. 7426
  Shun Zhang, Tongliang Li, Jiaqi Bai, Zhoujun Li

Ontology-Aware Network for Zero-Shot Sketch-Based Image Retrieval ................................................................. 7431
  Haoxiang Zhang, He Jiang, Ziqiang Wang, Deqiang Cheng

Fast Multiscale 3D Reconstruction Using Single-Photon Lidar Data ................................................................. 7436
  Sandor Plosz, Istvan Gyongy, Jonathan Leach, Stephen McLaughlin, Gerald S. Buller, Abderrahim Halimi

Quantile Online Learning for Semiconductor Failure Analysis................................................................. 7441
  Bangjian Zhou, Pan Jieming, Maheswari Sivan, Aaron Voon-Yew Thean, J. Senthilnath

Distributed Signal Processing for Out-Of-System Interference Suppression in Cell-Free Massive MIMO ................................................................. 7446
  Zakir Hussain Shaik, Erik G. Larsson

An Auto-Encoder Based Method for Camera Fingerprint Compression ................................................................. 7451
  Kaixuan Zhang, Zihan Liu, Jiashang Hu, Shilin Wang

Combining Loss Reweighting and Sample Resampling for Long-Tailed Instance Segmentation ................................................................. 7456
  Yaochi Zhao, Sen Chen, Qiong Chen, Zhuhua Hu

Geometry-Aware DOA Estimation Using a Deep Neural Network with Mixed-Data Input Features ................................................................. 7461
  Ulrik Kowalk, Simon Doclo, Joerg Bitzer

Target Speaker Extraction with Ultra-Short Reference Speech by VE-VE Framework ................................................................. 7466
  Lei Yang, Wei Liu, Lufen Tan, Jaemo Yang, Han-Gil Moon

MarginNCE: Robust Sound Localization with a Negative Margin ................................................................. 7471
  Sooyoung Park, Arda Senocak, Joon Son Chung
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Could the BubbleView Metaphor Be Used to Infer Visual Attention on 3D Graphical Content?</td>
<td>7476</td>
</tr>
<tr>
<td>Alexandre Bruckert, Mona Abid, Matthieu Perreira Da Silva, Patrick Le Callet</td>
<td></td>
</tr>
<tr>
<td>An Interpretable Model Using Evidence Information for Multi-Hop Question Answering Over Long Texts</td>
<td>7481</td>
</tr>
<tr>
<td>Yanyi Chen, Ruifang Liu, Xiyun Liu, Yidong Shi, Ge Bai</td>
<td></td>
</tr>
<tr>
<td>Self-Convolution for Automatic Speech Recognition</td>
<td>7486</td>
</tr>
<tr>
<td>Tian-Hao Zhang, Qi Liu, Xinyuan Qian, Song-Lu Chen, Feng Chen, Xu-Cheng Yin</td>
<td></td>
</tr>
<tr>
<td>Multi-Dimensional Signal Recovery Using Low-Rank Deconvolution</td>
<td>7491</td>
</tr>
<tr>
<td>David Reixach</td>
<td></td>
</tr>
<tr>
<td>Room Impulse Response Reconstruction Based on Spatio-Temporal-Spectral Features Learned from a Spherical Microphone Array Measurement</td>
<td>7496</td>
</tr>
<tr>
<td>Amy Bastine, Thushara D. Abhayapala, Jihui Aimee Zhang</td>
<td></td>
</tr>
<tr>
<td>Estimating Normalized Graph Laplacians in Financial Markets</td>
<td>7501</td>
</tr>
<tr>
<td>José Vinicius De M. Cardoso, Jiaxi Ying, Sandeep Kumar, Daniel P. Palomar</td>
<td></td>
</tr>
<tr>
<td>Radio Map Based UAV Target Localization</td>
<td>7506</td>
</tr>
<tr>
<td>Chen He, Weisheng Gong, Yangrui Dong, Xie Xie, Z. Jane Wang</td>
<td></td>
</tr>
<tr>
<td>Joint Estimation of Clustered User Activity and Correlated Channels with Unknown Covariance in mMTC</td>
<td>7511</td>
</tr>
<tr>
<td>Hamza Djelouat, Markus Leinonen, Markku Juntti</td>
<td></td>
</tr>
<tr>
<td>ICEL: Learning with Inconsistent Explanations</td>
<td>7516</td>
</tr>
<tr>
<td>Biao Liu, Xiaoyu Wu, Bo Yuan</td>
<td></td>
</tr>
<tr>
<td>ICStega: Image Captioning-Based Semantically Controllable Linguistic Steganography</td>
<td>7521</td>
</tr>
<tr>
<td>Xilong Wang, Yaofei Wang, Kejiang Chen, Jinyang Ding, Weiming Zhang, Nenghai Yu</td>
<td></td>
</tr>
<tr>
<td>VF-Taco2: Towards Fast and Lightweight Synthesis for Autoregressive Models with Variation Autoencoder and Feature Distillation</td>
<td>7526</td>
</tr>
<tr>
<td>Yuhao Liu, Cheng Gong, Longbiao Wang, Xixin Wu, Qiuyu Liu, Jianwu Dang</td>
<td></td>
</tr>
<tr>
<td>Evaluating Parameter-Efficient Transfer Learning Approaches on SURE Benchmark for Speech Understanding</td>
<td>7531</td>
</tr>
<tr>
<td>Yingting Li, Ambuj Mehrish, Rishabh Bhardwaj, Navonil Majumder, Bo Cheng, Shuai Zhao, Amir Zadeh, Rada Mihalcea, Soujanya Poria</td>
<td></td>
</tr>
<tr>
<td>Multi-User Data Detection in Massive MIMO with 1-Bit ADCS</td>
<td>7536</td>
</tr>
<tr>
<td>Amin Radbord, Italo Atzeni, Antti Tölli</td>
<td></td>
</tr>
<tr>
<td>Full-Band General Audio Synthesis with Score-Based Diffusion</td>
<td>7541</td>
</tr>
<tr>
<td>Santiago Pascual, Gautam Bhattacharya, Chunghsin Yeh, Jordi Pons, Joan Serrà</td>
<td></td>
</tr>
<tr>
<td>Phase-Aware Spoof Speech Detection Based on Res2net with Phase Network</td>
<td>7546</td>
</tr>
<tr>
<td>Juntae Kim, Sung Min Ban</td>
<td></td>
</tr>
<tr>
<td>Subband Dependency Modeling for Sound Event Detection</td>
<td>7551</td>
</tr>
<tr>
<td>Yadong Guan, Guibin Zheng, Jiqing Han, Huanliang Wang</td>
<td></td>
</tr>
<tr>
<td>Radio Sensing with Large Intelligent Surface for 6G</td>
<td>7556</td>
</tr>
<tr>
<td>Cristian J. Vaca-Rubio, Pablo Ramirez-Espinosa, Kimmo Kansanen, Zheng-Hua Tan, Elisabeth De Carvalho</td>
<td></td>
</tr>
</tbody>
</table>
Av-Sepformer: Cross-Attention Sepformer for Audio-Visual Target Speaker Extraction
Jiaxin Lin, Xinyu Cai, Heinrich Dinkel, Jun Chen, Zhiyong Yan, Yongqing Wang, Junbo Zhang, Zhiyong Wu, Yujun Wang, Helen Meng

7561

ZO-DARTS: Differentiable Architecture Search with Zeroth-Order Approximation
Lunchen Xie, Kaiyu Huang, Fan Xu, Qingjiang Shi

7566

Gaussian Prior Reinforcement Learning for Nested Named Entity Recognition
Yawen Yang, Xuming Hu, Fukun Ma, Shu’Ang Li, Aiwei Liu, Lijie Wen, Philip S. Yu

7571

Single-Shot Fractional Fourier Phase Retrieval
Yixiao Yang, Ran Tao

7576

A Deep Temporal Factor Analysis Method for Large Scale Financial Portfolio Selection
Yao Zhou, Ruidan Su, Shikui Tu, Lei Xu

7581

Role of Lexical Boundary Information in Chunk-Level Segmentation for Speech Emotion Recognition
Wei-Cheng Lin, Carlos Busso

7586

Towards Zero-Shot Code-Switched Speech Recognition
Brian Yan, Matthew Wiesner, Ondrej Klejch, Preethi Jyothi, Shinji Watanabe

7591

U-Shiftformer: Brain Tumor Segmentation Using a Shifted Attention Mechanism
Chih-Wei Lin, Zhongsheng Chen

7596

Test Your Samples Jointly: Pseudo-Reference for Image Quality Evaluation
Marcelin Tworski, Stéphane Lathuilière

7601

Exploring Binary Classification Loss for Speaker Verification
Bing Han, Zhengyang Chen, Yanmin Qian

7606

Understanding Shared Speech-Text Representations
Gary Wang, Kyle Kastner, Ankur Bapna, Zhehui Chen, Andrew Rosenberg, Bhuvana Ramabhadran, Yu Zhang

7611

Dual-Feature Enhancement for Weakly Supervised Temporal Action Localization
Siying Liu, Qiankun Liu, Qi Chu, Bin Liu, Nenghai Yu

7616

Semantic Centralized Contrastive Learning for Unsupervised Hashing
Fengming Liang, Changlin Fan, Bo Xiao, Kongming Liang

7621

Suffix Retrieval-Augmented Language Modeling
Zecheng Wang, Yuk-Cheung Tam

7626

Information and Sensing Beamforming Optimization for Multi-User Multi-Target MIMO ISAC Systems
Minghe Zhu, Lei Li, Shuqiang Xia, Tsung-Hui Chang

7631

Multiple Signed Graph Learning for Gene Regulatory Network Inference
Abdullah Karaaslanli, Satabdi Saha, Tapabrata Maiti, Selin Aviyente

7636

Low-Latency Electrolaryngeal Speech Enhancement Based on Fastspeech2-Based Voice Conversion and Self-Supervised Speech Representation
Kazuhiro Kobayashi, Tomoki Hayashi, Tomoki Toda

7641
CORS Od: Class-Oriented Relational Self Distillation ....................................................... 7646
Muzhou Yu, Sia Huat Tan, Kailu Wu, Runpei Dong, Linfeng Zhang, Karsheng Ma

Evopose: A Recursive Transformer for 3D Human Pose Estimation with Kinematic Structure Priors .......... 7651
Yaqi Zhang, Yan Lu, Bin Liu, Zhiwei Zhao, Qi Chu, Nenghai Yu

Improved WiFi-Based Respiration Tracking Via Contrast Enhancement ........................................ 7656
Wei-Hsiang Wang, Xiaolu Zeng, Beibei Wang, Yexin Cao, K. J. Ray Liu

Investigating SINDy as a Tool for Causal Discovery in Time Series Signals .................................. 7661
Andrew O'Brien, Rosina Weber, Edward Kim

Align, Write, Re-Order: Explainable End-to-End Speech Translation Via Operation Sequence
Generation ................................................................................................................................. 7666
Motoi Omachi, Brian Yan, Siddharth Dalmia, Yuya Fujita, Shinji Watanabe

Faster than Fast: Accelerating the Griffin-Lim Algorithm .......................................................... 7671
Rossen Nenov, Dang-Khoa Nguyen, Peter Balazs

Deep Learning-Based Compressive Sampling Optimization in Massive MIMO Systems .................... 7676
Saidur R. Pavel, Yimin D. Zhang, Maria S. Greco, Fulvio Gini

MPS-AMS: Masked Patches Selection and Adaptive Masking Strategy Based Self-Supervised
Medical Image Segmentation .................................................................................................... 7681
Xiangtao Wang, Ruizhi Wang, Biao Tian, Jiaojiao Zhang, Shuo Zhang, Junyang Chen,
Thomas Lukasiewicz, Zhenghua Xu

Robust Watermarking Scheme in Encrypted Domain Based on Integer Lifting Wavelet Transform
and Compressed Sensing ....................................................................................................... 7686
Di Xiao, Qin Tang, Aozhu Zhao, Min Li

Multi-Resolution Sequence Aggregation and Model-Agnostic Framework for Time-Series
Forecasting ............................................................................................................................... 7691
Juhyun Lyu, Jinseok Yang, Junghee Kim, Woohyung Lim, Wonbin Ahn, Dongwan Kang,
Minjae Kim, Nam Soo Kim

Dual-Graph Co-representation Learning for knowledge-Graph Enhanced Recommendation ............ 7696
Xinbiao Liu, Bin Liang, Junyu Niu, Chaofeng Sha, Dong Wu

Polarized Signal Singular Spectrum Analysis with Complex SSA ............................................... 7701
Sébastien Journé, Nicolas Le Bihan, Florent Chatelain, Julien Flamant

A Token-Level Contrastive Framework for Sign Language Translation ........................................ 7706
Biao Fu, Peigen Ye, Liang Zhang, Pei Yu, Cong Hu, Xiaodong Shi, Yidong Chen

Ultra Real-Time Portrait Matting Via Parallel Semantic Guidance ............................................. 7711
Xin Huang, Jiake Xie, Bo Xu, Han Huang, Ziweng Li, Cheng Lu, Yandong Guo, Yong Tang

Knowledge-Aware Bayesian Co-Attention for Multimodal Emotion Recognition ......................... 7716
Zihan Zhao, Yu Wang, Yanfeng Wang

Densitytoken: Weakly-Supervised Crowd Counting with Density Classification .......................... 7721
Zaiyi Hu, Bingli Wang, Xuelong Li

Contrastive Representation Learning for Acoustic Parameter Estimation .................................... 7726
Philipp Götz, Cagdas Tuna, Andreas Walther, Emanuël A. P. Habets
Compressive Estimation of Near Field Channels for Ultra Massive-Mimo Wideband THz Systems
Simon Tarboush, Anum Ali, Tareq Y. Al-Naffouri

Improving the out-Of-Distribution Generalization Capability of Language Models:
Counterfactually-Augmented Data is Not Enough
Caoyun Fan, Wenqing Chen, Jidong Tian, Yitian Li, Hao He, Yaohui Jin

Towards Realizing the Value of Labeled Target Samples: A Two-Stage Approach for Semi-Supervised Domain Adaptation
Mengqun Jin, Kai Li, Shuyan Li, Chunming He, Xi Li

Audio-Visual Speech Enhancement with a Deep Kalman Filter Generative Model
Ali Golmakani, Mostafa Sadeghi, Romain Serizel

WL-MSR: Watch and Listen for Multimodal Subtitle Recognition
Jiawei Liu, Hao Wang, Weining Wang, Xingjian He, Jing Liu

Deep Unfolding-Enabled Hybrid Beamforming Design for mmWave Massive MIMO Systems
Nhan Nguyen, Mengyuan Ma, Nir Shlezinger, Yonina C. Eldar, A. L. Swindlehurst, Markku Juntti

End-To-End Word-Level Disfluency Detection and Classification in Children’s Reading Assessment
Lavanya Venkatasubramaniam, Vishal Sunder, Eric Fosler-Lussier

Tell Model Where to Attend: Improving Interpretability of Aspect-Based Sentiment Classification Via Small Explanation Annotations
Zhenxiao Cheng, Jie Zhou, Wen Wu, Qin Chen, Liang He

Exploiting Multi-Decision and Deep Refinement for Ultrasound Image Segmentation
Wenjing Liu, Xuanya Li, Kai Hu, Xieping Gao

Learning from Positive and Unlabeled Data Using Observer-GAN
Omar Zamzam, Haleh Akrami, Richard Leahy

Transformer-Based Multi-Prototype Approach for Diabetic Macular Edema Analysis in OCT Images
Plácido L. Vidal, Joaquim De Moura, Jorge Novo, Marcos Ortega, Jaime S. Cardoso

On the Design and Training Strategies for Rnn-Based Online Neural Speech Separation Systems
Kai Li, Yi Luo

Learning How to Learn Domain-Invariant Parameters for Domain Generalization
Feng Hou, Yao Zhang, Yang Liu, Jin Yuan, Cheng Zhong, Yang Zhang, Zhongchao Shi, Jianping Fan, Zhiqiang He

Wespeaker: A Research and Production Oriented Speaker Embedding Learning Toolkit
Hongji Wang, Chengdong Liang, Shuai Wang, Zhengyang Chen, Binbin Zhang, Xu Xiang, Yanlei Deng, Yanmin Qian

Multi-Label Temporal Evidential Neural Networks for Early Event Detection
Xujiang Zhao, Xuchao Zhang, Chen Zhao, Jin-Hee Cho, Lance Kaplan, Dong Hyun Jeong, Audun Jøsang, Haifeng Chen, Feng Chen

Is Quality Enough? Integrating Energy Consumption in a Large-Scale Evaluation of Neural Audio Synthesis Models
Constance Douwes, Giovanni Bindi, Antoine Caillon, Philippe Esling, Jean-Pierre Briot
On Crowdsourcing-Design with Comparison Category Rating for Evaluating Speech Enhancement

Angélica S. Z. Suárez, Clément Laroche, Line H. Clemmensen, Sneha Das

A Prototypical Semantic Decoupling Method Via Joint Contrastive Learning for Few-Shot Named Entity Recognition

Guanting Dong, Zechen Wang, Liwen Wang, Daichi Guo, Dayuan Fu, Yuxiang Wu, Chen Zeng, Xuefeng Li, Tingfeng Hui, Keqing He, Xinyue Cui, Qixiang Gao, Weiran Xu

Revisit Out-Of-Vocabulary Problem for Slot Filling: A Unified Contrastive Framework with Multi-Level Data Augmentations

Daichi Guo, Guanting Dong, Dayuan Fu, Yuxiang Wu, Chen Zeng, Tingfeng Hui, Liwen Wang, Xuefeng Li, Zechen Wang, Keqing He, Xinyue Cui, Weiran Xu

Trinet: Stabilizing Self-Supervised Learning from Complete Or Slow Collapse

Lixin Cao, Jun Wang, Ben Yang, Dan Su, Dong Yu

An Adapter Based Multi-Label Pre-Training for Speech Separation and Enhancement

Tianrui Wang, Xie Chen, Zhuo Chen, Shu Yu, Weibin Zhu

Large-Scale Language Model Rescoring on Long-Form Data

Tongzhou Chen, Cyril Allauzen, Yinghui Huang, Daniel Park, David Rybach, W. Ronny Huang, Rodrigo Cabrera, Kartik Audhkhasi, Bhuvana Ramabhadran, Pedro J. Moreno, Michael Riley

Contrastive Speech Mixup for Low-Resource Keyword Spotting

Dianwen Ng, Ruixi Zhang, Jia Qi Yip, Chong Zhang, Yukun Ma, Trung Hieu Nguyen, Chongjia Ni, Eng Siong Chng, Bin Ma

Low-Complexity Acoustic Echo Cancellation with Neural Kalman Filtering

Dong Yang, Fei Jiang, Wei Wu, Xuefei Fang, Muyong Cao

Learning on Entropy Coded Images with CNN

Rémi Piau, Thomas Maugé, Aline Roumy

Asymmetric Polynomial Loss for Multi-Label Classification

Yusheng Huang, Jiexing Qi, Xinbing Wang, Zhouhan Lin

MAID: A Conditional Diffusion Model for Long Music Audio Inpainting

Kaiyang Liu, Wendong Gan, Chenchen Yuan

Rumor Detection Via Assessing the Spreading Propensity of Users

Peng Zheng, Zhen Huang, Yong Dou, Yeqing Yan

Deep Network Series for Large-Scale High-Dynamic Range Imaging

Amir Aghabiglou, Matthieu Terris, Adrian Jackson, Yves Wiaux

Towards Improved Sonar Performance Using Environment-Informed Sparse Sub-Array Processing

Alexandre L’Her, Angélique Drémeau, Florent Le Courtois, Gaultier Real, Xavier Cristol, Yann Stéphan

Learning Properties of Holomorphic Neural Networks of Dual Variables

Dmitry Kozlov, Mikhail Bakulin, Stanislav Pavlov, Aleksandr Zuev, Mariya Krylova, Igor Kharchikov
An Effective Anomalous Sound Detection Method Based on Representation Learning with Simulated Anomalies

Han Chen, Yan Song, Zhe Zhuo, Yu Zhou, Yu-Hong Li, Hui Xue, Ian McLoughlin

Hearing and Seeing Abnormality: Self-Supervised Audio-Visual Mutual Learning for Deepfake Detection

Chang-Sung Sung, Jun-Cheng Chen, Chu-Song Chen

Sample-Efficient Robust MMV Recovery Algorithm

Yuvraj Singh, Jahvi Singh Rohela, Satish Mulleti

Joint Channel and Direction Estimation for Ground-To-UAV Communications Enabled by a Simultaneous Reflecting and Sensing RIS

Jiguang He, Aymen Fakhreddine, George C. Alexandropoulos

Towards Low-Power Heart Rate Estimation Based on User’s Demographics and Activity Level for Wearables


Covariance Regularization for Probabilistic Linear Discriminant Analysis

Zhiyuan Peng, Mingjie Shao, Xuanji He, Xi Li, Tan Lee, Ke Ding, Guanglu Wan

I-Tuning: Tuning Frozen Language Models with Image for Lightweight Image Captioning

Ziyang Luo, Zhipeng Hu, Yadong Xi, Rongsheng Zhang, Jing Ma

Graph-Based Spectro-Temporal Dependency Modeling for Anti-Spoofing

Feng Chen, Shiwen Deng, Tieran Zheng, Yongjun He, Jiqing Han

Mixer: DNN Watermarking Using Image Mixup

Kassem Kallas, Teddy Furon

Leveraging Positional-Related Local-Global Dependency for Synthetic Speech Detection

Xiaohui Liu, Meng Liu, Longbiao Wang, Kong Aik Lee, Hanyi Zhang, Jianwu Dang

Improving Disfluency Detection with Multi-Scale Self Attention and Contrastive Learning

Peiying Wang, Chaqun Duan, Meng Chen, Xiaodong He

Improving Massively Multilingual ASR with Auxiliary CTC Objectives

William Chen, Brian Yan, Jiatong Shi, Yifan Peng, Soumi Maiti, Shinji Watanabe

Two-Branch Multi-Scale Deep Neural Network for Generalized Document Recapture Attack Detection

Jiaxing Li, Chenqi Kong, Shiqi Wang, Haoliang Li

Resource Allocation for UAV-Enabled Integrated Sensing and Communication (ISAC) Via Multi-Objective Optimization

Omid Rezaei, Mohammad Mahdi Naghsh, Seyed Mohammad Karbasi, Mohammad Mahdi Nayebi

Disentangling the Horowitz Factor: Learning Content and Style from Expressive Piano Performance

Huan Zhang, Simon Dixon

Personalizing Federated Learning with Over-The-Air Computations

Zihan Chen, Zeshen Li, Howard H. Yang, Tony Q. S. Quek
Achieving Fair Speech Emotion Recognition Via Perceptual Fairness ................................................................. 7966
       Woan-Shiuan Chien, Chi-Chun Lee

Soft Label Coding for End-To-end Sound Source Localization with Ad-hoc Microphone Arrays ......................... 7971
       Linfeng Feng, Yijun Gong, Xiao-Lei Zhang

Towards Reliable Image Outpainting: Learning Structure-Aware Multimodal Fusion with Depth Guidance................................. 7976
       Lei Zhang, Chunyu Lin, Kang Liao, Yao Zhao

SafeDeep: A Scalable Robustness Verification Framework for Deep Neural Networks ........................................ 7981
       Anahita Baninajjar, Kamran Hosseini, Ahmed Rezine, Amir Aminifar

Exploiting Modality-Invariant Feature for Robust Multimodal Emotion Recognition with Missing Modalities ................................................................ 7986
       Haolin Zuo, Rui Liu, Jinming Zhao, Guanglai Gao, Haizhou Li

Front-End Adapter: Adapting Front-End Input of Speech Based Self-Supervised Learning for Speech Recognition ........................................................................................................... 7991
       Xie Chen, Ziyang Ma, Changli Tang, Yujin Wang, Zhisheng Zheng

Target-Speaker Voice Activity Detection Via Sequence-to-Sequence Prediction .............................................................. 7996
       Ming Cheng, Weiqing Wang, Yucong Zhang, Xiaoyi Qin, Ming Li

Knowledge-Graph Augmented Music Representation for Genre Classification ................................................................. 8001
       Han Ding, Wenjing Song, Cui Zhao, Fei Wang, Ge Wang, Wei Xi, Jizhong Zhao

Pre-Training Strategies Using Contrastive Learning and Playlist Information for Music Classification and Similarity .................................................................................................................. 8006
       Pablo Alonso-Jiménez, Xavier Favory, Hadrien Foroughmand, Grigoris Bourdalas, Xavier Serra, Thomas Lidy, Dmitry Bogdanov

Assisted RTF-Vector-Based Binaural Direction of Arrival Estimation Exploiting a Calibrated External Microphone Array ........................................................................................................... 8011
       Daniel Fejgin, Simon Doclo

Synthetic Pseudo Anomalies for Unsupervised Video Anomaly Detection: A Simple Yet Efficient Framework Based on Masked Autoencoder ........................................................................... 8016
       Xiangyu Huang, Caidan Zhao, Chenxing Gao, Lvdong Chen, Zhiqiang Wu

Using Modified Adult Speech as Data Augmentation for Child Speech Recognition ..................................................... 8021
       Zijian Fan, Xinwei Cao, Giampiero Salvi, Torbjørn Svendsen

Phonetic Anchor-Based Transfer Learning to Facilitate Unsupervised Cross-Lingual Speech Emotion Recognition .......................................................................................................................... 8026
       Shreya G. Upadhyay, Luz Martinez-Lucas, Bo-Hao Su, Wei-Cheng Lin, Woan-Shiuan Chien, Ya-Tse Wu, William Katz, Carlos Busso, Chi-Chun Lee

Improving Retrieval-Based Dialogue System Via Syntax-Informed Attention ................................................................. 8031
       Tengtao Song, Nuo Chen, Ji Jiang, Zhihong Zhu, Yuexian Zou

Weakly-Supervised Scene-Specific Crowd Counting Using Real-Synthetic Hybrid Data .............................................. 8036
       Yaowu Fan, Jia Wan, Yuan Yuan, Qi Wang

Phonation Mode Detection in Singing: A Singer Adapted Model ................................................................................. 8041
       Yixin Wang, Wei Wei, Ye Wang
Improving the Stochastic Gradient Descent’s Test Accuracy by Manipulating the \( \infty \) Norm of Its Gradient Approximation

Paul Rodriguez

Semi-Supervised Domain Generalization with Graph-Based Classifier

Minxiang Ye, Yifei Zhang, Shiqiang Zhu, Anhuan Xie, Senwei Xiang

Zero-Shot Speech Emotion Recognition Using Generative Learning with Reconstructed Prototypes

Xinzhou Xu, Jun Deng, Zixing Zhang, Zhen Yang, Björn W. Schuller

Disambiguation of Cognitive Impairment Diagnosis with EEG-Based Dual-Contrastive Learning

Zhenxi Song, Zian Pei, Huixia Ren, Lin Zhu, Yi Guo, Zhiguo Zhang

CN-CVS: A Mandarin Audio-Visual Dataset for Large Vocabulary Continuous Visual to Speech Synthesis

Chen Chen, Dong Wang, Thomas Fang Zheng

Hierarchical Graph Learning for Stock Market Prediction Via a Domain-Aware Graph Pooling Operator

Arie N. Arya, Yao Lei Xu, L. Stankovic, Danilo P. Mandic

Flowgrad: Using Motion for Visual Sound Source Localization

Rajsuryan Singh, Pablo Zinemanas, Xavier Serra, Juan Pablo Bello, Magdalena Fuentes

A Generalized Subspace Distribution Adaptation Framework for Cross-Corpus Speech Emotion Recognition

Shaokai Li, Peng Song, Liang Ji, Yun Jin, Wenming Zheng

Simplicial Vector Autoregressive Model for Streaming Edge Flows

Joshin Krishnan, Rohan Money, Baltasar Beferull-Lozano, Elvin Isufi

A Radar-Jammer Zero-Sum Repeated Bayesian Game

Sofia Suvorova, Ali Pezeshki, Ross Kyprianou, Bill Moran

A Frequency-Weighted Leaky Fxlms Algorithm with Application to Feedback Active Noise Control Systems

Yu Tang, Hongwei Zhang

Speech Modeling with a Hierarchical Transformer Dynamical VAE

Xiaoyu Lin, Xiaoyu Bie, Simon Leglaive, Laurent Girin, Xavier Alameda-Pineda

Deep Unfolded Tensor Robust PCA with Self-Supervised Learning

Harry Dong, Megna Shah, Sean Donegan, Yuejie Chi

Improved Training of Mixture-Of-Experts Language GANs

Yekun Chai, Qiyue Yin, Junge Zhang

Dynamic Distributed Convex Optimization "Over-The-Air" in Decentralized Wireless Networks

Navneet Agrawal, Renato L. G. Cavalcante, Slawomir Stanczak

Study and Design of Robust Personal Sound Zones with Vast Using Low Rank Rirs

Sankha Subhra Bhattacharjee, Liming Shi, Guoli Ping, Xiaoxiang Shen, Mads Græsbøll Christensen

Online Binaural Speech Separation of Moving Speakers with a Wavesplit Network

Cong Han, Nima Mesgarani
Image Segmentation for Improved Lossless Screen Content Compression ................................................................. 8131
Shabhrish Reddy Uddehal, Tilo Strutz, Hannah Och, André Kaup

Target Sound Extraction with Variable Cross-Modality Clues .................................................................................. 8136
Chenda Li, Yao Qian, Zhuo Chen, Dongmei Wang, Takuya Yoshioka, Shujie Liu, Yanmin Qian, Michael Zeng

Dynamic Chunk Convolution for Unified Streaming and Non-Streaming Conformer ASR ........................................ 8141
Xilai Li, Goeric Huybrechts, Srikanth Ronanki, Jeff Farriss, Sravan Bodapati

Deep-Unfolded Adaptive Projected Subgradient Method for Mimo Detection ....................................................... 8146
Jochen Fink, Renato L. G. Cavalcante, Zoran Utkovski, Slawomir Stanczak

Predicting Brain Age Using Transferable Covariance Neural Networks ............................................................... 8151
Saurabh Sihaq, Gonzalo Mateos, Corey McMillan, Alejandro Ribeiro

Retinal Biomarkers for Detecting Diabetic Retinopaty Using Smartphone-Based Deep Learning Frameworks .......................................................................................................................... 8156
Mahmut Karakaya, Ramazan S. Aygun

Study of Manifold Geometry Using Multiscale Non-Negative Kernel Graphs ............................................................. 8161
Carlos Hurtado, Sarah Shekkizhar, Javier Ruiz-Hidalgo, Antonio Ortega

Understandable Relu Neural Network for Signal Classification .................................................................................. 8166
Marie Guyomard, Susana Barbosa, Lionel Fillatre

A Video Anomaly Detection Framework Based on Appearance-Motion Semantics Representation Consistency .......................................................................................................................... 8171
Xiangyu Huang, Caidan Zhao, Zhiqiang Wu

Syngen: A Syntactic Plug-And-Play Module for Generative Aspect-Based Sentiment Analysis ................................ 8176
Chengze Yu, Taiqiang Wu, Jiayi Li, Xingyu Bai, Yujie Yang

TinyOOD: Effective out-Of-Distribution Detection for TinyML .................................................................................. 8181
Yongchang Li, Juncheng Jia, Yan Zuo, Weipeng Zhu

Radar Clutter Covariance Estimation: A Nonlinear Spectral Shrinkage Approach ................................................ 8186
Shashwat Jain, Vikram Krishnamurthy, Muralidhar Rangaswamy, Bosung Kang, Sandeep Gogineni

Gated Contextual Adapters for Selective Contextual Biasing in Neural Transducers .............................................. 8191
Anastasios Alexandridis, Kanthashree Mysore Sathyendra, Grant P. Strimel, Feng-Ju Chang, Ariya Rastrow, Nathan Susanj, Athanasios Mouchtaris

LightGrad: Lightweight Diffusion Probabilistic Model for Text-To-Speech ............................................................ 8196
Jie Chen, Xingchen Song, Zhendong Peng, Binbin Zhang, Fuping Pan, Zhiyong Wu

VAN-ICP: GPU-Accelerated Approximate Nearest Neighbor Search for ICP Registration Via Voxel Dilation .................. 8201
Weimin Wang, Qiong Chang

Distance-Based Weight Transfer for Fine-Tuning from Near-Field to Far-Field Speaker Verification ....................... 8206
Li Zhang, Qing Wang, Hongji Wang, Yue Li, Wei Rao, Yannan Wang, Lei Xie

Utilizing Wav2Vec in Database-Independent Voice Disorder Detection ............................................................... 8211
Saska Tirronen, Farhad Javanmardi, Manila Kodali, Sudarsana Reddy Kadiri, Paavo Alku
CAT: Causal Audio Transformer for Audio Classification ................................................................. 8216
Xiaoyu Liu, Hanlin Lu, Jianbo Yuan, Xinyu Li

Water Leak Detection and Localization Using Convolutional Autoencoders .............................................. 8221
Daniele Ugo Leonzio, Paolo Bestagini, Marco Marcon, Gian Paolo Quarta, Stefano Tubaro

Multimodal Emotion Recognition Based on Deep Temporal Features Using Cross-Modal Transformer and Self-Attention ............................................. 8226
Bubai Maji, Monorama Swain, Rajlakshmi Guha, Aurobinda Routray

Reliability Estimation for Synthetic Speech Detection .................................................................................. 8231
Davide Salvi, Paolo Bestagini, Stefano Tubaro

VPPT: Visual Pre-Trained Prompt Tuning Framework for Few-Shot Image Classification ............................... 8236
Zhao Song, Ke Yang, Naiyang Guan, Junjie Zhu, Peng Qiao, Qingyong Hu

LINK: Linguistic Steganalysis Framework with External Knowledge .............................................................. 8241
Jinshuai Yang, Zhongliang Yang, Xinrui Ge, Jiajun Zou, Yue Gao, Yongfeng Huang

Rate Splitting and Precoding Strategies for Multi-User MIMO Broadcast Channels with Common and Private Streams ........................................... 8246
Liana Khamidullina, André L. F. De Almeida, Martin Haardt

Sparse Bayesian Learning Assisted Decision Fusion in Millimeter Wave Massive MIMO Sensor Networks ............................................................................. 8251
Apoorva Chawla, Domenico Ciuonzo, Pierluigi Salvo Rossi

Robust Adaptive Beamforming with Proximal Method ................................................................................. 8256
Ruifu Li, Danijela Cabric

Elliptical Wishart Distribution: Maximum Likelihood Estimator from Information Geometry ......................... 8261
Imen Ayadi, Florent Bouchard, Frédéric Pascal

Range-ISL Minimization and Spectral Shaping in MIMO Radar Systems Via Waveform Design ..................... 8266
Ehsan Raei, Mohammad Alaee-Kerahroodi, Bhavani Shankar, Björn Ottersten

Self-Attention for Enhanced OAMP Detection in MIMO Systems .............................................................. 8271
Alexander Fuchs, Christian Knoll, Nima N. Moghadam, Alexey Pak, Jinliang Huang, Erik Leitinger, Franz Perknopf

Modeling Global Latent Semantic in Multi-Turn Conversations with Random Context Reconstruction ........... 8276
Chengwen Zhang, Danqin Wu

Optimizing Vision Transformers for Medical Image Segmentation ............................................................. 8281
Qianying Liu, Chaitanya Kaul, Jun Wang, Christos Anagnostopoulos, Roderick Murray-Smith, Fani Deligianni

Multimodal Dyadic Impression Recognition Via Listener Adaptive Cross-Domain Fusion ............................. 8286
Yuanchao Li, Peter Bell, Catherine Lai

Universal Speaker Recognition Encoders for Different Speech Segments Duration ...................................... 8291
Sergey Novoselov, Vladimir Volokhov, Galina Lavrentyeva

Unsupervised Speaker Verification Using Pre-Trained Model and Label Correction .................................... 8296
Zhicong Chen, Jie Wang, Wexuan Hu, Lin Li, Qingyang Hong
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-Resolution Embedding Extractor for Speaker Diarisation</td>
<td>8301</td>
</tr>
<tr>
<td>Hee-Soo Heo, Youngki Kwon, Bong-Jin Lee, You Jin Kim, Jee-Weon Jung</td>
<td></td>
</tr>
<tr>
<td>Targeted Adversarial Attacks Against Neural Machine Translation</td>
<td>8306</td>
</tr>
<tr>
<td>Sahar Sadrizadeh, Amirhossein Dabiri Aghdam, Ljiljana Dolamic, Pascal Frossard</td>
<td></td>
</tr>
<tr>
<td>Deep Learning Sparse Array Design Using Binary Switching Configurations</td>
<td>8311</td>
</tr>
<tr>
<td>Syed A. Hamza, Kyle Juretus, Moeness G. Amin, Fauzia Ahmad</td>
<td></td>
</tr>
<tr>
<td>Sparsity-Driven Joint Blind Deconvolution-Demodulation with Application to Motor Fault Detection</td>
<td>8316</td>
</tr>
<tr>
<td>Varun A. Kelkar, Dehong Liu, Hiroshi Inoue, Makoto Kanemaru</td>
<td></td>
</tr>
<tr>
<td>Papez: Resource-Efficient Speech Separation with Auditory Working Memory</td>
<td>8321</td>
</tr>
<tr>
<td>Hyunseok Oh, Juheon Yi, Youngki Lee</td>
<td></td>
</tr>
<tr>
<td>Cassia Valenti-Botinhao, Andrea Lorena Aldana Blanco, Ondrej Klejch, Peter Bell</td>
<td></td>
</tr>
<tr>
<td>A Meta-Gnn Approach to Personalized Seizure Detection and Classification</td>
<td>8331</td>
</tr>
<tr>
<td>Abdellah Rahmani, Arun Venkitaraman, Pascal Frossard</td>
<td></td>
</tr>
<tr>
<td>Community Detection Graph Convolutional Network for Overlap-Aware Speaker Diarization</td>
<td>8336</td>
</tr>
<tr>
<td>Jie Wang, Zhicong Chen, Haodong Zhou, Lin Li, Qingyang Hong</td>
<td></td>
</tr>
<tr>
<td>Smart Split-Federated Learning Over Noisy Channels for Embryo Image Segmentation</td>
<td>8341</td>
</tr>
<tr>
<td>Zahra Hafezi Kafshgari, Ivan V. Bajic, Parvaneh Saeedi</td>
<td></td>
</tr>
<tr>
<td>Variational Inference Aided Estimation of Time Varying Channels</td>
<td>8346</td>
</tr>
<tr>
<td>Benedikt Böck, Michael Baur, Valentina Rizzello, Wolfgang Utschick</td>
<td></td>
</tr>
<tr>
<td>Online Learning-Based Waveform Selection for Improved Vehicle Recognition in Automotive Radar</td>
<td>8351</td>
</tr>
<tr>
<td>Charles E. Thornton, William W. Howard, R. Michael Buehrer</td>
<td></td>
</tr>
<tr>
<td>Gator: Graph-Aware Transformer with Motion-Disentangled Regression for Human Mesh Recovery from a 2D Pose</td>
<td>8356</td>
</tr>
<tr>
<td>Yingxuan You, Hong Liu, Xia Li, Wenhao Li, Ti Wang, Runwei Ding</td>
<td></td>
</tr>
<tr>
<td>Bayesian Network Modeling and Prediction of Transitions Within the Homelessness System</td>
<td>8361</td>
</tr>
<tr>
<td>Khandker Sadia Rahman, Daphney-Stavroula Zois, Charalampos Chelmis</td>
<td></td>
</tr>
<tr>
<td>Distributed Adaptive Norm Estimation for Blind System Identification in Wireless Sensor Networks</td>
<td>8366</td>
</tr>
<tr>
<td>M. Blochberger, F. Elvander, R. Ali, J. Østergaard, J. Jensen, M. Moonen, T. Van Waterschoot</td>
<td></td>
</tr>
<tr>
<td>Diffusion-Based Generative Speech Source Separation</td>
<td>8371</td>
</tr>
<tr>
<td>Robin Scheibler, Younga Ji, Soo-Whan Chung, Jaeuk Byun, Soyeon Choe, Min-Seok Choi</td>
<td></td>
</tr>
<tr>
<td>Lost in Translation: Generating Adversarial Examples Robust to Round-Trip Translation</td>
<td>8376</td>
</tr>
<tr>
<td>Neel Bhandari, Pin-Yu Chen</td>
<td></td>
</tr>
<tr>
<td>Relapse Prediction from Long-Term Wearable Data Using Self-Supervised Learning and Survival Analysis</td>
<td>8381</td>
</tr>
<tr>
<td>E. Fekas, A. Zlatintsi, P. P. Filntisis, C. Garoufis, N. Efthymiou, P. Maragos</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Authors</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>On the Relevance of the Differences Between HRTF Measurement Setups for Machine Learning</td>
<td>Johan Pauwels, Lorenzo Picinali</td>
</tr>
<tr>
<td>IR-ECG: Invertible Reconstruction of ECG</td>
<td>Peng Wang, Xi Huang, Li Cui</td>
</tr>
<tr>
<td>Dynamic Fair Node Representation Learning</td>
<td>Oyku Deniz Kose, Yanning Shen</td>
</tr>
<tr>
<td>Speech Intelligibility Classifiers from 550k Disordered Speech Samples</td>
<td>Subhashini Venugopalan, Jimmy Tobin, Samuel J. Yang, Katie Seaver, Richard J. N. Cave, Pan-Pan Jiang, Neil Zeghidour, Rus Heywood, Jordan Green, Michael P. Brenner</td>
</tr>
<tr>
<td>Disentangled Feature Learning for Real-Time Neural Speech Coding</td>
<td>Xue Jiang, Xiulian Peng, Yuan Zhang, Yan Lu</td>
</tr>
<tr>
<td>Collaborative Audio-Visual Event Localization Based on Sequential Decision and Cross-Modal Consistency</td>
<td>Yuqian Kuang, Xiaopeng Fan</td>
</tr>
<tr>
<td>Accelerated Massive MIMO Detector Based on Annealed Underdamped Langevin Dynamics</td>
<td>Nicolas Zilberstein, Chris Dick, Rahman Doost-Mohammady, Ashutosh Sabharwal, Santiago Segarra</td>
</tr>
<tr>
<td>Repetition Counting from Compressed Videos Using Sparse Residual Similarity</td>
<td>Rishabh Khurana, Jayesh Rajkumar Vachhani, Sourabh Vasant Gothe, Pranay Kashyap</td>
</tr>
<tr>
<td>Clean Sample Guided Self-Knowledge Distillation for Image Classification</td>
<td>Jiuye Wang, Yanxiong Li, Qianhua He, Wei Xie</td>
</tr>
<tr>
<td>LGVT: Local-Global Vision Transformer for Breast Cancer Histopathological Image Classification</td>
<td>Lang Wang, Juan Liu, Peng Jiang, Dehua Cao, Baochuan Pang</td>
</tr>
<tr>
<td>Robust Angle Estimation for Hybrid mmWave Systems</td>
<td>Yuan-Pei Lin, Ting-Ming Yang</td>
</tr>
<tr>
<td>HiSSNet: Sound Event Detection and Speaker Identification Via Hierarchical Prototypical Networks for Low-Resource Headphones</td>
<td>N Shashaank, Berker Banar, Mohammad Rasool Izadi, Jeremy Kemmerer, Shuo Zhang, Chuan-Che Jeff Huang</td>
</tr>
<tr>
<td>Bayesian Cramér-Rao Bound Estimation with Score-Based Models</td>
<td>Evan Scope Crafts, Bo Zhao</td>
</tr>
<tr>
<td>Enabling Large-Scale Image Search with Co-Attention Mechanism</td>
<td>Zechao Hu, Adrian G. Bors</td>
</tr>
</tbody>
</table>
Dialog Act Guided Contextual Adapter for Personalized Speech Recognition ................................................................. 8466
Feng-Ju Chang, Thejaswi Muniyappa, Kanthashree Mysore Sathyendra, Kai Wei, Grant P. Strimel, Ross McGowan

LMCodec: A Low Bitrate Speech Codec with Causal Transformer Models ................................................................. 8471
Teerapat Jenrungrot, Michael Chinen, W. Bastiaan Kleijn, Jan Skoglund, Zalán Borsos, Neil Zeghidour, Marco Tagliasacchi

SEPDIFF: Speech Separation Based on Denoising Diffusion Model ............................................................................ 8476
Bo Chen, Chao Wu, Wenbin Zhao

Few but Informative Local Hash Code Matching for Image Retrieval ........................................................................ 8481
Zechao Hu, Adrian G. Bors

Vision, Deduction and Alignment: An Empirical Study on Multi-Modal Knowledge Graph
Alignment ........................................................................................................................................................................... 8486
Yangning Li, Jiaoayan Chen, Yinghui Li, Yuejia Xiang, Xi Chen, Hai-Tao Zheng

Low-Resource Music Genre Classification with Cross-Modal Neural Model Reprogramming .................................. 8491
Yun-Xing Hung, Chao-Han Huck Yang, Pin-Yu Chen, Alexander Lerch

MAST: Multiscale Audio Spectrogram Transformers ................................................................................................. 8496
Sreyan Ghosh, Ashish Seth, S. Umesh, Dinesh Manocha

Dynamic Scalable Self-Attention Ensemble for Task-Free Continual Learning ........................................................ 8501
Fei Ye, Adrian G. Bors

Stuart: Individualized Classroom Observation of Students with Automatic Behavior Recognition and Tracking ........ 8506
Huayi Zhou, Fei Jiang, Jiaxin Si, Lili Xiong, Hongtao Lu

Reducing the Communication and Computational Cost of Random Fourier Features Kernel LMS in Diffusion Networks ......................................................................................................................... 8511
Daniel G. Tiglea, Renato Candido, Luis A. Azpicueta-Ruiz, Magno T. M. Silva

SLICER: Learning Universal Audio Representations Using Low-Resource Self-Supervised Pre-Training ................ 8516
Ashish Seth, Sreyan Ghosh, S. Umesh, Dinesh Manocha

Deep Proximal Gradient Method for Learned Convex Regularizers ........................................................................ 8521
Aaron Berk, Yanting Ma, Petros Boufounos, Pu Wang, Hassan Mansour

Analysis and Transformation of Voice Level in Singing Voice .................................................................................... 8526
Frederik Bous, Axel Roebel

Data-Driven Graph Convolutional Neural Networks for Power System Contingency Analysis .................................. 8531
Valentin Bolz, Johannes Rueß, Andreas Zell

Robust Online Multiband Drift Estimation in Electrophysiology Data ......................................................................... 8536

Consistent Estimators of a New Class of Covariance Matrix Distances in the Large Dimensional Regime ....................... 8541
Roberto Pereira, Xavier Mestre, David Gregoratti
Make More of Your Data: Minimal Effort Data Augmentation for Automatic Speech Recognition and Translation .............................................................. 8546
  Tsz Kin Lam, Shigehiko Schamoni, Stefan Riezler

Compressing Cross-Domain Representation Via Lifelong Knowledge Distillation .............................................................. 8551
  Fei Ye, Adrian G. Bors

Phase Unwrapping in Correlated Noise for FMCW Lidar Depth Estimation .............................................................. 8556
  A. Ulvog, J. Rapp, T. Koike-Akino, H. Mansour, P. Boufounos, K. Parsons

A Bayesian Perspective on Noise2Noise: Theory and Extensions .............................................................. 8561
  Sarah Miller, Christina Karam, Achour Idoughi, Kodai Kikuchi, Keigo Hirakawa

Self-Similarity is All You Need for Fast and Light-Weight Generic Event Boundary Detection .............................................................. 8566
  Sourabh Vasant Gothe, Jayesh Rajkumar Vachhani, Rishabh Khurana, Pranay Kashyap

On Batching Variable Size Inputs for Training End-To-End Speech Enhancement Systems .............................................................. 8571
  Philippe Gonzalez, Tommy Sonne Alstrøm, Tobias May

Learning to Reconnect Interrupted Trajectories for Weakly Supervised Multi-Object Tracking .............................................................. 8576
  Yu-Lei Li, Yang Lu, Jie Li, Hanzi Wang

Textless Direct Speech-To-Speech Translation with Discrete Speech Representation .............................................................. 8581
  Xinjian Li, Ye Jia, Chung-Cheng Chiu

Improving Self-Supervised Learning for Audio Representations by Feature Diversity and Decorrelation .............................................................. 8586
  Bac Nguyen, Stefan Uhlich, Fabien Cardinaux

Asymptotically Optimal Nonparametric Classification Rules for Spike Train Data .............................................................. 8591
  Miroslaw Pawlak, Mateusz Pabian, Dominik Rzepka

Fully Complex-Valued Deep Learning Model for Visual Perception .............................................................. 8596
  Aniruddh Sikdar, Sumanth Udupa, Suresh Sundaram

Beamformer-Guided Target Speaker Extraction .............................................................. 8601
  Mohamed Elminshawi, Srikanth Raj Chetupalli, Emanuël A. P. Habets

Context-Aware Coherent Speaking Style Prediction with Hierarchical Transformers for Audiobook Speech Synthesis .............................................................. 8606
  Shun Lei, Yixuan Zhou, Liyang Chen, Zhiyong Wu, Shiyin Kang, Helen Meng

Less is More: A Unified Architecture for Device-Directed Speech Detection with Multiple Invocation Types .............................................................. 8611

Lasso-Based Fast Residual Recovery for Modulo Sampling .............................................................. 8616
  Shaik Basheeruddin Shah, Satish Mulleti, Yonina C. Eldar

Ensemble Prosody Prediction for Expressive Speech Synthesis .............................................................. 8621
  Tian Huey Teh, Vivian Hu, Devang S Ram Mohan, Zack Hodari, Christopher G. R. Wallis, Tomás Gómez-Ibarrondo, Alexandra Torresquinterno, James Leoni, Mark Gales, Simon King

Massively Multilingual Shallow Fusion with Large Language Models .............................................................. 8626
  Ke Hu, Tara N. Sainath, Bo Li, Nan Du, Yanping Huang, Andrew M. Dai, Yu Zhang, Rodrigo Cabrera, Zhifeng Chen, Trevor Strohman
Identifying Coordination in a Cognitive Radar Network - A Multi-Objective Inverse Reinforcement Learning Approach ................................................................. 8631
Luke Snow, Vikram Krishnamurthy, Brian M. Sadler

Dual-Based Online Learning of Dynamic Network Topologies ................................................................. 8636
Seyed Saman Saboksayr, Gonzalo Mateos

Downlink Covariance Estimation in URA FDD Massive MIMO Systems .................................................. 8641
Salime Bameri, Khalid Almahrog, Ramy H. Gohary, Amr El-Keyi, Yahia Ahmed

Pyramid Dynamic Inference: Encouraging Faster Inference Via Early Exit Boosting ............................. 8646
Ershad Banijamali, Pegah Kharazmi, Sepehr Eghbali, Jixuan Wang, Clement Chung, Samridhi Choudhary

Improving Automatic Sleep Staging Via Temporal Smoothness Regularization ..................................... 8651
Huy Phan, Elisabeth Heremans, Oliver Y. Chen, Philipp Koch, Alfred Mertins, Maarten De Vos

Spatially Selective Deep Non-Linear Filters for Speaker Extraction ....................................................... 8656
Kristina Tesch, Timo Gerkmann

Egocentric Audio-Visual Noise Suppression .................................................................................................. 8661
Roshan Sharma, Weipeng He, Ju Lin, Egor Lakomkin, Yang Liu, Kaustubh Kalgaonkar

HRTF Field: Unifying Measured HRTF Magnitude Representation with Neural Fields ............................ 8666
You Zhang, Yuxiang Wang, Zhiyao Duan

Toward Privacy-Enhancing Ambulatory-Based Well-Being Monitoring: Investigating User Re-Identification Risk in Multimodal Data ................................................................. 8671
Ravi Pranjal, Ranjana Seshadri, Rakesh Kumar Sanath Kumar Kadaba, Tiantian Feng, Shrikanth S. Narayanan, Theodora Chaspari

A Knowledge-Driven Vowel-Based Approach of Depression Classification from Speech Using Data Augmentation ........................................................................................................ 8676
Kexin Feng, Theodora Chaspari

InfoShape: Task-Based Neural Data Shaping Via Mutual Information ..................................................... 8681
Homa Esfahanizadeh, William Wu, Manya Ghobadi, Regina Barzilay, Muriel Médard

Conversational Text-To-SQL: An Odyssey into State-of-The-Art and Challenges Ahead ........................ 8686
Sree Hari Krishnan Parthasarathi, Lu Zeng, Dilek Hakkani-Tür

Interpolation of Spatial Room Impulse Responses Using Partial Optimal Transport ................................ 8691
Aaron Geldert, Nils Meyer-Kahlen, Sebastian J. Schlecht

Sign Language Recognition Via Deformable 3D Convolutions and Modulated Graph Convolutional Networks .................................................................................................................. 8696
Katerina Papadimitriou, Gerasimos Potamianos

Efficient Feature Extraction for Non-Maximum Suppression in Visual Person Detection ....................... 8701
Charalampos Symeonidis, Ioannis Mademlis, Ioannis Pitas, Nikos Nikolaidis

Deformable Temporal Convolutional Networks for Monaural Noisy Reverberant Speech Separation ...... 8706
William Ravenscroft, Stefan Goetz, Thomas Hain

Speech Privacy Leakage from Shared Gradients in Distributed Learning ................................................. 8711
Zhuohang Li, Jiaxin Zhang, Jian Liu
VarietySound: Timbre-Controllable Video to Sound Generation Via Unsupervised Information Disentanglement
Chenyang Cui, Zhou Zhao, Yi Ren, Jinglin Liu, Rongjie Huang, Feiyang Chen, Zhefeng Wang, Baoxing Huai, Fei Wu

Neural Networks with Quantization Constraints
 Ignacio Hounie, Juan Elenter, Alejandro Ribeiro

Wireless Power Transfer Using Chirp Waveforms
Arijit Roy, Constantinos Psomas, Ioannis Krikidis

Classifying Non-Individual Head-Related Transfer Functions with a Computational Auditory Model: Calibration and Metrics
Rapolas Daugintis, Roberto Barumerli, Lorenzo Picinali, Michele Geronazzo

Cov Loss: Covariance-Based Loss for Deep Face Recognition
Ibrahim Alkanhal, Abdullah Almansour, Lamia Alsalloom, Raed Alfaadawy, Marios Savvides

Improved Small Sample Hypothesis Testing Using the Uncertain Likelihood Ratio
James Zachary Hare, Lance M. Kaplan

S. S. Krishna Chaitanya Bulusu, Nuatti Tervo, Praneeth Susarla, Mikko J. Sillanpää, Olli Silvén, Markku Juntti, Aarno Pärssinen

End-To-End Spoken Language Understanding with Tree-Constrained Pointer Generator
Guangzhi Sun, Chao Zhang, Philip C. Woodland

Recursive/Iterative Unique Projection-Aggregation Decoding of Reed-Muller Codes
Marzieh Hashemipour-Nazari, Renate Debets, Kees Goossens, Alexios Balatsoukas-Stimming

WIFI-Based Robust Child Presence Detection for Smart Cars
Sakila S. Jayaweera, Beibei Wang, Xiaolu Zeng, Wei-Hsiang Wang, K. J. Ray Liu

Exploring Sequence-To-Sequence Transformer-Transducer Models for Keyword Spotting
Beltrán Labrador, Guanlong Zhao, Ignacio López Moreno, Angelo Scorza Scarpati, Liam Fowl, Quan Wang

Performance Comparison of TTS Models for Brazilian Portuguese to Establish a Baseline
Wilmer Lobato, Felipe Farias, William Cruz, Marcellus Amadeus

Low-Rank Tensor Decompositions for Quaternion Multiway Arrays
Osimone Imhogiemhe, Julien Flamant, Xavier Luciani, Yassine Zniyed, Sebastian Miron

Received Power Maximization with Practical Phase-Dependent Amplitude Response in RIS-Aided OFDM Wireless Communications
Dimitris Kompostiotis, Dimitris Vordonis, Vassilis Paliouras

The 2nd Clarity Enhancement Challenge for Hearing Aid Speech Intelligibility Enhancement: Overview and Outcomes
Michael A. Akeroyd, Will Bailey, Jon Barker, Trevor J. Cox, John F. Culling, Simone Graetzer, Graham Naylor, Zuzanna Podwinska, Zehai Tu

Internal Language Model Estimation Based Adaptive Language Model Fusion for Domain Adaptation
Rao Ma, Xiaobo Wu, Jin Qiu, Yanan Qin, Haihua Xu, Peihao Wu, Zejun Ma
Cochlear Decomposition: A Novel Bio-Inspired Multiscale Analysis Framework ........................................ 8796
Hessa Alfalahi, Ahsan Khandoker, Ghada Alhussein, Leontios Hadjileontiadis

Motion Matters: A Novel Motion Modeling for Cross-View Gait Feature Learning .................................. 8801
Jingqi Li, Jiaqi Gao, Yuzhen Zhang, Hongming Shan, Junping Zhang

Last: Scalable Lattice-Based Speech Modelling in Jax ............................................................................. 8806
Ke Wu, Ehsan Variani, Tom Bagby, Michael Riley

Sparsity Constraint Implementation for the Joint Eigenvalue Decomposition of Matrices .................. 8811
Rémi André, Xavier Luciani

Client Selection for Generalization in Accelerated Federated Learning: A Bandit Approach ............ 8816
Dan Ben Ami, Kobi Cohen, Qing Zhao

Multilingual End-To-End Spoken Language Understanding for Ultra-Low Footprint Applications ...... 8821
Markus Müller, Anastasios Alexandridis, Zach Trozenski, Joel Whiteman, Grant Strimel, Nathan Susanj, Athanasios Mouchtaris, Siegfried Kunzmann

Robust Monocular Localization of Drones by Adapting Domain Maps to Depth Prediction
Inaccuracies .............................................................................................................................................. 8826
Priyesh Shukla, S. Sureshkumar, Alex C. Stutts, Sathya Ravi, Theja Tulabandhula, Amit R. Trivedi

Acoustic Source Localization in the Spherical Harmonics Domain Exploiting Low-Rank
Approximations ........................................................................................................................................ 8831
Maximo Cobos, Mirco Pezzoli, Fabio Antonacci, Augusto Sarti

Investigating Content-Aware Neural Text-to-Speech MOS Prediction Using Prosodic and
Linguistic Features .................................................................................................................................... 8836
Alexandra Vioni, Georgia Maniati, Nikolaos Ellinas, June Sig Sung, Inchul Hwang, Aimilios Chalamandaris, Pirros Tsiakoulis

Joint Angle and Respiration Estimation for Passive and Device-Free Respiration Monitoring ............ 8841
Gerrit Maus, Dieter Brückmann

The Role of Initial Entanglement in Adaptive Gibbs State Preparation on Quantum Computers .......... 8846
Sophia E. Economou, Ada Warren, Edwin Barnes

Learning Dependencies of Discrete Speech Representations with Neural Hidden Markov Models ........ 8851
Sung-Lin Yeh, Hao Tang

Supervised Contrastive Learning as Multi-Objective Optimization for Fine-Tuning Large Pre-
Trained Language Models ......................................................................................................................... 8856
Youness Moukafih, Mounir Ghogho, Kamel Smaili

Unsupervised Domain Adaptation for Preference Learning Based Speech Emotion Recognition ....... 8861
Abinay Reddy Naini, Mary A. Kohler, Carlos Busso

An Augmented Gaussian Sum Filter Through a Mixture Decomposition ............................................. 8866
Kostas Tsampourakis, Victor Elvira

EfficientSpeech: An On-Device Text to Speech Model ............................................................................. 8871
Rowel Atienza

Contrastive Self-Supervised Learning for Automated Multi-Modal Dance Performance Assessment .... 8876
Yun Zhong, Fan Zhang, Yiannis Demiris
SG-VD: Stochastic Gates Based Speech Activity Detection
Jonathan Svirsky, Ofir Lindenbaum

Automatic Error Detection in Integrated Circuits Image Segmentation: A Data-Driven Approach
Zhikang Zhang, Bruno Machado Trindade, Michael Green, Zifan Yu, Christopher Pawlowicz, Fengbo Ren

End-To-End Spoken Language Understanding Using Joint CTC Loss and Self-Supervised, Pretrained Acoustic Encoders
Jixuan Wang, Martin Radfar, Kai Wei, Clement Chung

Anomalous Signal Detection for Cyber-Physical Systems Using Interpretable Causal Neural Network
Shuo Zhang, Jing Liu

On the Primal and Dual Formulations of the Discrete Mumford-Shah Functional
Nelly Pustelnik

Abstract Representation for Multi-Intent Spoken Language Understanding
Rim Abrougui, Géraldine Damnati, Johannes Heinecke, Frédéric Béchet

CSM in Motion Vector Steganalysis: The Effect of Coders on Motion Vectors in H.264 Video Encoding
Verena Lachner, Katharina Schaar, Ralf Zimmermann

Distributed Gaussian Process Hyperparameter Optimization for Multi-Agent Systems
Peiyuan Zhai, Raj Thilak Rajan

Regularized EM Algorithm
Pierre Houdouin, Esa Ollila, Frédéric Pascal

Waveform Boundary Detection for Partially Spoofed Audio
Zexin Cai, Weiqing Wang, Ming Li

Simultaneous Acoustic Echo Sorting and 3-D Room Geometry Inference
Kathleen Macwilliam, Filip Elvander, Toon Van Waterschoot

Simultaneously Learning Robust Audio Embeddings and Balanced Hash Codes for Query-By-Example
Anup Singh, Kris Demuynck, Vipul Arora

Row Conditional-TGAN for Generating Synthetic Relational Databases
Mohamed Gueye, Yazid Attabi, Maxime Dumas

Distill-Quantize-Tune - Leveraging Large Teachers for Low-Footprint Efficient Multilingual NLU on Edge
Pegah Kharazi, Zhewei Zhao, Clement Chung, Samridhi Choudhary

Comprehensive Complexity Assessment of Emerging Learned Image Compression on CPU and GPU
Farhad Pakdaman, Moncef Gabbouj

Convolutional Filtering on Sampled Manifolds
Zhiyang Wang, Luana Ruiz, Alejandro Ribeiro

Detail-Aware Uncalibrated Photometric Stereo
Antonio Agudo
Optimal Condition Training for Target Source Separation ................................................................. 8966
Efthymios Tzinis, Gordon Wichern, Paris Smaragdis, Jonathan Le Roux

Torchaudio-Squim: Reference-Less Speech Quality and Intelligibility Measures in Torchaudio ........ 8971
Anurag Kumar, Ke Tan, Zhaoheng Ni, Pranay Manocha, Xiaohui Zhang, Ethan Henderson, Buye Xu

PRIME: 3D Human Pose and Body Shape Recovery with Perspective Projection ............................. 8976
Baobei Xu, Shukai Fang, Zhaoyang Li, Shicai Yang, Di Xie, Shiliang Pu

Spice+: Evaluation of Automatic Audio Captioning Systems with Pre-Trained Language Models .... 8981
Félix Gontier, Romain Serizel, Christophe Cerisara

Tangent Bundle Filters and Neural Networks: from Manifolds to Cellular Sheaves and Back .......... 8986
C. Battiloro, Z. Wang, H. Riess, P. Di Lorenzo, A. Ribeiro

Neural-AFC: Learning-Based Step-Size Control for Adaptive Feedback Cancellation with Closed-Loop Model Training ........................................................................................................ 8991
Behrad Soleimani, Henning Schepker, Majid Mirbagheri

Continuous Descriptor-Based Control for Deep Audio Synthesis ...................................................... 8996
Ninon Devis, Nils Demerlé, Sarah Nabi, David Genova, Philippe Esling

Native Multi-Band Audio Coding Within Hyper-Autoencoded Reconstruction Propagation Networks ................................................................. 9001
Darius Petermann, Inseon Jang, Minje Kim

Contrastive Learning-Based Audio to Lyrics Alignment for Multiple Languages ......................... 9006
Simon Durand, Daniel Stoller, Sebastian Ewert

Improved Projection Learning for Lower Dimensional Feature Maps ............................................ 9011
Ilan Price, Jared Tanner

Counterfactual Explanation for Multivariate Times Series Using a Contrastive Variational Autoencoder .......................................................................................................................... 9016
William Todo, Merwann Selmani, Béatrice Laurent, Jean-Michel Loubes

Gaitmixer: Skeleton-Based Gait Representation Learning Via Wide-Spectrum Multi-Axial Mixer 9021
Ekkasit Pinyoamuntapong, Ayman Ali, Pu Wang, Minwoo Lee, Chen Chen

Applying Independent Vector Analysis on EEG-Based Motor Imagery Classification .................. 9026
Caroline P. A. Moraes, Bruno Aristimunha, Lucas Heck Dos Santos, Walter Hugo Lopez Pinaya, Raphael Yokoingawa De Camargo, Denis G. Fantinato, Aline Neves

Prosody is Not Identity: A Speaker Anonymization Approach Using Prosody Cloning .............. 9031
Sarina Meyer, Florian Lux, Julia Koch, Pavel Denisov, Pascal Tilli, Ngoc Thang Vu

ViT-Cat: Parallel Vision Transformers with Cross Attention Fusion for Popularity Prediction in MEC Networks ........................................................................................................... 9036
Zohreh Hajiaekhondi-Meybodi, Arash Mohammadi, Ming Hou, Jamshid Abouei, Konstantinos N. Plataniotis

BreathIE: Estimating Breathing Inhale Exhale Ratio Using Motion Sensor Data from Consumer Earbuds ..................................................................................................................... 9041
Nafiul Rashid, Md Mahbubur Rahman, Tousif Ahmed, Jilong Kuang, Jun Alex Gao
A Point is a Wave: Point-Wave Network for Place Recognition ................................................................. 9046
Ge Li, Ruonan Zhang

Exploiting Speaker Embeddings for Improved Microphone Clustering and Speech Separation in Ad-Hoc Microphone Arrays ........................................................................................................... 9051
Stijn Kindt, Jenthe Thienpondt, Niles Madhu

Neural Architecture Search with Multimodal Fusion Methods for Diagnosing Dementia ........................................ 9056
Michail Chatzianastasis, Loukas Ilias, Dimitris Askounis, Michalis Vazirgiannis

Explainable Audio Classification of Playing Techniques with Layer-Wise Relevance Propagation ........................... 9061
Changhong Wang, Vincent Lostanlen, Mathieu Lagrange

An Edge Alignment-Based Orientation Selection Method for Neutron Tomography ............................................ 9066
Diyu Yang, Shimin Tang, Singanallur V. Venkatakrishnan, Mohammad S. N. Chowdhury, Yuxuan Zhang, Hassina Z. Bilheux, Gregory T. Buzzard, Charles A. Bouman

Enhance Transferability of Adversarial Examples with Model Architecture .......................................................... 9071
Mingyuan Fan, Wenzhong Guo, Zuobin Ying, Ximeng Liu

Beyond Neural-On-Neural Approaches to Speaker Gender Protection ............................................................... 9076
Loes Van Bemmel, Zhuan Liu, Nik Vaessen, Martha Larson

Masked Autoencoders Are Articulatory Learners .................................................................................................. 9081
Ahmed Adel Attia, Carol Y. Espy-Wilson

Super Dilated Nested Arrays with Ideal Critical Weights and Increased Degrees of Freedom ............................... 9086
Ahmed M. A. Shaalan, Jun Du

Multimodal Facial Action Unit Detection with Physiological Signals ................................................................. 9091
Zhihua Li, Lijun Yin

Estimating and Analyzing Neural Information Flow Using Signal Processing on Graphs ........................................ 9096
Felix Schwock, Julien Bloch, Les Atlas, Shima Abadi, Azadeh Yazdan-Shahmorad

Matrix Low-Rank Approximation for Policy Gradient Methods .............................................................................. 9101
Sergio Rozada, Antonio G. Marques

Soft 2D-To-3D Delivery Using Deep Graph Neural Networks for Holographic-Type Communication ..................... 9106
Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe

Exploring Subgroup Performance in End-To-End Speech Models ......................................................................... 9111
Alkis Koudounas, Eliana Pastor, Giuseppe Attanasio, Vittorio Mazzia, Manuel Giollo, Thomas Gueudre, Luca Caglierio, Luca De Alfaro, Elena Baralis, Daniele Amberti

Perceptual–Neural–Physical Sound Matching .................................................................................................... 9116
Han Han, Vincent Lostanlen, Mathieu Lagrange

An Implicit Gradient Method for Constrained Bilevel Problems Using Barrier Approximation ............................ 9121
Ioannis Tsaknakis, Prashant Khanduri, Mingyi Hong

Benchmarking Convolutional Neural Network Inference on Low-Power Edge Devices ........................................... 9126
Oscar Ferraz, Helder Araujo, Vitor Silva, Gabriel Falcao

Signal Analysis-Synthesis Using the Quantum Fourier Transform ........................................................................ 9131
Aradhita Sharma, Glen Uehara, Vivek Narayanaswamy, Leslie Miller, Andreas Spanias
Blind Acoustic Room Parameter Estimation Using Phase Features ....................................................... 9136
Christopher Ick, Adib Mehrabi, Wenyu Jin

Exploring Wav2vec 2.0 Fine Tuning for Improved Speech Emotion Recognition .................................................. 9141
Li-Wei Chen, Alexander Rudnicky

BER-Aware Dynamic Resource Management for Edge-Assisted Goal-Oriented Communications .............. 9146
Francesco Binucci, Paolo Banelli

Adaptable End-To-End ASR Models Using Replaceable Internal LMs and Residual Softmax ......................... 9151
Keqi Deng, Philip C. Woodland

Co-Operative CNN for Visual Saliency Prediction on WCE Images ................................................................. 9156
George Dimas, Anastasios Koulaouzidis, Dimitris K. Iakovidis

Outside Knowledge Visual Question Answering Version 2.0 ................................................................. 9161
Benjamin Z. Reichman, Anirudh Sundar, Christopher Richardson, Tamara Zubiati, Prithwijit Chowdhury, Aaryan Shah, Jack Truxal, Micah Grimes, Dristi Shah, Woo Ju Chee, Saif Punjwani, Atishay Jain, Larry Heck

Learning Environmental Structure Using Acoustic Probes with a Deep Neural Network ............................ 9166
Toros Arikan, Amir Weiss, Hari Vishnu, Grant Deane, Andrew Singer, Gregory Wornell

Single-Shot Domain Adaptation Via Target-Aware Generative Augmentations ........................................... 9171
Rakshith Subramanyam, Kowshik Thopalli, Spring Berman, Pavan Turaga, Jayaraman J. Thiagarajan

A Parallel Attention Mechanism for Image Manipulation Detection and Localization ................................ 9176
Qiang Zeng, Hongxia Wang, Yang Zhou, Rui Zhang, Sijiang Meng

Speech MOS Multi-Task Learning and Rater Bias Correction ................................................................. 9181
Haleh Akrami, Hannes Gamper

Improving Audio Captioning Using Semantic Similarity Metrics ................................................................. 9186
Rehana Mahfuz, Yinyi Guo, Erik Visser

Glacier: Glass-Box Transformer for Interpretable Dynamic Neuroimaging .................................................. 9191
Usman Mahmood, Zening Fu, Vince Calhoun, Sergey Plis

Efficient Speech Translation with Dynamic Latent Perceivers ................................................................. 9196
Ioannis Tsiamas, Gerard I. Gállego, José A. R. Fonollosa, Marta R. Costa-Jussà

Early Detection of Cognitive Decline Using Voice Assistant Commands .................................................. 9201
Eli Kurtz, Youxiang Zhu, Tiffany Driesse, Bang Tran, John A. Batsis, Robert M. Roth, Xiaohui Liang

Boosting the Accuracy of SRAM-Based in-Memory Architectures Via Maximum Likelihood-Based Error Compensation Method ................................................................. 9206
Hyungyo Kim, Naresh Shanbhag

Using Machine Learning to Understand the Relationships Between Audiometric Data, Speech Perception, Temporal Processing, and Cognition ................................................................. 9211
Rana M. Khalil, Alexandra Papanicolaou, Renee Ti Chou, Bobby E. Gibbs, Samira Anderson, Sandra Gordon-Salant, Michael P. Cummings, Matthew J. Goupell
M-SpeechCLIP: Leveraging Large-Scale, Pre-Trained Models for Multilingual Speech to Image Retrieval .............................................................. 9216
    Layne Berry, Yi-Jen Shih, Hsuan-Fu Wang, Heng-Jui Chang, Hung-Yi Lee, David Harwath

Coupled CP Tensor Decomposition with Shared and Distinct Components for Multi-Task Fmri Data Fusion.............................................................. 9221
    R. A. Borsoi, I. Lehmann, M. A. B. S. Akhonda, V. D. Calhoun, K. Usevich, D. Brie, T. Adali

ASSD: Synthetic Speech Detection in the AAC Compressed Domain .............................................................. 9226
    Amit Kumar Singh Yadav, Ziyue Xiang, Emily R. Bartusiak, Paolo Bestagini, Stefano Tubaro, Edward J. Delp

Robust Self-Guided Deep Image Prior ........................................................................................................ 9231
    Evan Bell, Shijun Liang, Qing Qu, Saiprasad Ravishankar

Fully Distributed Federated Learning with Efficient Local Cooperations ...................................................... 9236
    Evangelos Georgatos, Christos Mavrokefalidis, Kostas Berberidis

HYDRA-HGR: A Hybrid Transformer-Based Architecture for Fusion of Macroscopic and Microscopic Neural Drive Information ........................................ 9241
    Mansoorer Montazerin, Elahe Rahimian, Farnoosh Naderkhani, S. Farokh Atashzar, Hamid Alinejad-Rokny, Arash Mohammadi

AERO: Audio Super Resolution in the Spectral Domain .............................................................................. 9246
    Moshe Mandel, Or Tal, Yossi Adi

On the Reduction of Large-Scale Room Acoustic Models ......................................................................... 9251
    Pavlos Stoikos, Olympia Axelou, George Floros, Nestor Evmorfopoulos, George Stamoulis

JPEG Pleno Call for Proposals Responses Quality Assessment .................................................................... 9256
    Joao Prazeres, Zhe Luo, Antonio M. G. Pinheiro, Luis A. Da Silva Cruz, Stuart Perry

Unlimited Sampling Radar: Life Below the Quantization Noise ................................................................ 9261
    Thomas Feuillen, Bhavani Shankar Mr, Ayush Bhandari

UWB Localization-Of-Things Via Soft Information: Network Experimentation in Indoor Environment .............................................................. 9266
    Carlos A. Gómez-Vega, Moe Z. Win, Andrea Conti

Leveraging Label Correlations in a Multi-Label Setting: A Case Study in Emotion ..................................... 9271
    Georgios Chochlakis, Gireesh Mahajan, Sabyasachee Baruah, Keith Burghardt, Kristina Lerman, Shrikanth Narayanan

Analysing the Masked Predictive Coding Training Criterion for Pre-Training a Speech Representation Model ........................................................................ 9276
    Hemant Yadav, Sunayana Sitaram, Rajiv Ratn Shah

Using Emotion Embeddings to Transfer Knowledge Between Emotions, Languages, and Annotation Formats .............................................................................. 9281
    Georgios Chochlakis, Gireesh Mahajan, Sabyasachee Baruah, Keith Burghardt, Kristina Lerman, Shrikanth Narayanan

Jamming Source Localization Using Augmented Physics-Based Model .................................................... 9286
    Andrea Nardin, Tales Imbiriba, Pau Closas

Audio-Visual Speaker Diarization in the Framework of Multi-User Human-Robot Interaction .................... 9291
    Timothée Dhaussy, Bassam Jabaian, Fabrice Lefèvre, Radu Horaud
Personalized Task Load Prediction in Speech Communication

Robert P. Spang, Karl El Hajal, Sebastian Möller, Milos Cernak

Text-To-ECG: 12-Lead Electrocardiogram Synthesis Conditioned on Clinical Text Reports

Hyunseung Chung, Jiho Kim, Joon-Myoung Kwon, Ki-Hyun Jeon, Min Sung Lee, Edward Choi

Fixed-Point Quantization Aware Training for on-Device Keyword-Spotting

Sashank Macha, Om Oza, Alex Escott, Francesco Calivá, Robbie Armitano, Santosh Kumar Cheekatmaalia, Sree Hari Krishnan Parthasarathi, Yuzong Liu

Model-Free Learning of Optimal Beamformers for Passive IRS-Assisted Sumrate Maximization

Hassaan Hashmi, Spyridon Pougkakiotis, Dionysios S. Kalogerias

Self-Supervised Speech Representation Learning for Keyword-Spotting with Light-Weight Transformers

Chenyang Gao, Yue Gu, Francesco Caliva, Yuzong Liu

Towards Dialogue Modeling Beyond Text

Tongzi Wu, Yuhao Zhou, Wang Ling, Hojin Yang, Joana Vêloso, Lin Sun, Ruixin Huang, Norberto Guimarães, Scott Sanner

Variable Rate Allocation for Vector-Quantized Autoencoders

Federico Baldassarre, Alaaeldin El-Nouby, Hervé Jégou

Simultaneous Estimation of Direction of Arrival and Sound Speed Using a Non-Uniform Sensor Array

Ryouichi Nishimura, Kenichi Takizawa

Identifiable Bounded Component Analysis Via Minimum Volume Enclosing Parallelotope

Jingzhou Hu, Kejun Huang

A Statistical Interpretation of the Maximum Subarray Problem

Dennis Wei, Dmitry M. Malioutov

Learning to Personalize Equalization for High-Fidelity Spatial Audio Reproduction

Arjun Gupta, Pablo F. Hoffmann, Sebastian Prepelita, Philip Robinson, Vamsi K. Ithapu, David L. Alon

Deep Learning for Lagrangian Drift Simulation at the Sea Surface

Daria Botvynko, Carlos Granero-Belinchon, Simon Van Gennip, Abdesslam Benzinou, Ronan Fablet

Machine Learning Based Early Debris Detection Using Automotive Low Level Radar Data

Kanishka Tyagi, Shan Zhang, Yihang Zhang, John Kirkwood, Sanling Song, Narbik Manukian

Volume-Regularized Nonnegative Tucker Decomposition with Identifiability Guarantees

Yuchen Sun, Kejun Huang

SAN: A Robust End-To-End ASR Model Architecture

Zeping Min, Qian Ge, Guanhua Huang

Efficient Domain Adaptation for Speech Foundation Models

Bo Li, Dongseong Hwang, Zhouyuan Huo, Junwen Bai, Guru Prakash, Tara N. Sainath, Khe Chai Sim, Yu Zhang, Wei Han, Trevor Strohman, Francoise Beaufays
Hearttoheart: The Arts of Infant Versus Adult-Directed Speech Classification
Najla D. Al Futaisi, Alejandrina Cristia, Björn W. Schuller

Deep AHS: A Deep Learning Approach to Acoustic Howling Suppression
Hao Zhang, Meng Yu, Dong Yu

Powerful and Extensible WFST Framework for Rnn-Transducer Losses
Aleksandr Laptev, Vladimir Bataev, Igor Gitman, Boris Ginsburg

Speaker-Independent Acoustic-to-Articulatory Speech Inversion
Peter Wu, Li-Wei Chen, Cheol Jun Cho, Shinji Watanabe, Louis Goldstein, Alan W Black, Gopala K. Anumanchipalli

Hadamard Layer to Improve Semantic Segmentation
Angello Hoyos, Mariano Rivera

SMUG: Towards Robust MRI Reconstruction by Smoothed Unrolling
Hui Li, Jinghan Jia, Shijun Liang, Yuguang Yao, Saiprasad Ravishankar, Sijia Liu

Repackagingaugment: Overcoming Prediction Error Amplification in Weight-Averaged Speech Recognition Models Subject to Self-Training
Jae-Hong Lee, Dong-Hyun Kim, Joon-Hyuk Chang

Joint Antenna Selection and Beamforming in Integrated Automotive Radar Sensing-Communications with Quantized Double Phase Shifters
Lifan Xu, Shunqiao Sun, Yimin D. Zhang, Athina Petropulu

Exploring Attention Mechanisms for Multimodal Emotion Recognition in an Emergency Call Center Corpus
Theo Deschamps-Berger, Lori Lamel, Laurence Devillers

Evaluation of Categorical Generative Models - Bridging the Gap Between Real and Synthetic Data
Florence Regol, Anja Kroon, Mark Coates

Joint Modelling of Spoken Language Understanding Tasks with Integrated Dialog History
Siddhant Arora, Hayato Futami, Emiru Tsunoo, Brian Yan, Shinji Watanabe

Learn Topological Representation with Flexible Manifold Layer
Ziheng Jiao, Hongyuan Zhang, Xuelong Li

Signal Processing with Optical Quadratic Random Sketches
Rémi Delogne, Vincent Schellekens, Laurent Daudet, Laurent Jacques

SemanticAC: Semantics-Assisted Framework for Audio Classification
Yicheng Xiao, Yue Ma, Shuyan Li, Hantao Zhou, Ran Liao, Xiu Li

DWFoermer: Dynamic Window Transformer for Speech Emotion Recognition
Shuaqi Chen, Xiaofen Xing, Weihin Zhang, Weidong Chen, Xiangmin Xu

Unique Bispectrum Inversion for Signals with Finite Spectral/Temporal Support
Samuel Pinilla, Kumar Vijay Mishra, Brian M. Sadler

Online Model Compression for Federated Learning with Large Models
Tien-Ju Yang, Yonghui Xiao, Giovanni Motta, Françoise Beaufays, Rajiv Mathews, Mingqing Chen
GSWIN: Gated MLP Vision Model with Hierarchical Structure of Shifted Window
Mocho Go, Hideyuki Tachibana ................................................................. 9541

Two-Phase Prototypical Contrastive Domain Generalization for Cross-Subject EEG-Based Emotion Recognition
Honghua Cai, Jiahui Pan ........................................................................... 9546

Factorized Projection-Domain Spatio-Temporal Regularization for Dynamic Tomography
Berk Iskender, Marc L. Klasky, Brian M. Patterson, Yoram Bresler ................. 9551

Mouth Breathing Detection Using Audio Captured Through Earbuds
Tousif Ahmed, Md. Mahbubur Rahman, Ebrahim Nemati, Jilong Kuang, Alex Gao .......................................................... 9556

Deep Implicit Distribution Alignment Networks for cross-Corpus Speech Emotion Recognition
Yan Zhao, Jincen Wang, Yuan Zong, Wenming Zheng, Haijun Lian, Li Zhao .......................................................... 9561

Discriminative Vector Learning with Application to Single Channel Speech Separation
Ha Minh Tan, Kai-Wen Liang, Jia-Ching Wang ........................................... 9566

Self-Supervised Adversarial Training for Contrastive Sentence Embedding
Jen-Tzung Chien, Yuan-An Chen .............................................................. 9571

Efficient Personalized Federated Learning on Selective Model Training
Yeting Guo, Fang Liu, Tongqing Zhou, Zhiping Cai, Nong Xiao ................. 9576

Efficient and Effective Multi-Camera Pose Estimation with Weighted M-Estimate Sample Consensus
Xinyu Lin, Yingjie Zhou, Xin Zhang, Yipeng Liu, Ce Zhu ............................. 9581

Euro: Espnet Unsupervised ASR Open-Source Toolkit
Dongji Gao, Jiatong Shi, Shun-Po Chuang, Leibny Paola Garcia, Hung-Yi Lee, Shinji Watanabe, Sanjeev Khudanpur ................. 9586

HeMPPCAT: Mixtures of Probabilistic Principal Component Analysers for Data with Heteroscedastic Noise
Alec S. Xu, Laura Balzano, Jeffrey A. Fessler ............................................ 9591

Voice-Preserving Zero-Shot Multiple Accent Conversion
Mumin Jin, Prashant Serai, Jilong Wu, Andros Tjandra, Vimal Manohar, Qing He .......................................................... 9596

Navigating and Reaching Therapeutic Goals with Dynamical Systems in Conversation-Based Interventions
Victor Ardulov, Shrikanth Narayanan .......................................................... 9601

Dictionary Learning on Graph Data with Weisfieler-Lehman Sub-Tree Kernel and Ksvd
Kaveen Liyanage, Reese Pearsall, Clemente Izurieta, Bradley M. Whitaker .... 9606

Networked Policy Gradient Play in Markov Potential Games
Sarper Aydin, Ceyhun Eksin ................................................................. 9611

Level-Line Guided Edge Drawing for Robust Line Segment Detection
Xinyu Lin, Yingjie Zhou, Yipeng Liu, Ce Zhu ........................................... 9616

Speech-Text Based Multi-Modal Training with Bidirectional Attention for Improved Speech Recognition
Yuhang Yang, Haihua Xu, Hao Huang, Eng Siong Chng, Sheng Li ................. 9621
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimating Acoustic Direction of Arrival Using a Single Structural</td>
<td>9626</td>
</tr>
<tr>
<td>Sensor on a Resonant Surface</td>
<td></td>
</tr>
<tr>
<td>Tre Dipassio, Michael C. Heilemann, Benjamin Thompson, Mark F. Bocko</td>
<td></td>
</tr>
<tr>
<td>Comparative Study of IRS Assisted Opportunistic Communications Over</td>
<td>9631</td>
</tr>
<tr>
<td>I.i.d. and Los Channels</td>
<td></td>
</tr>
<tr>
<td>L. Yashvanth, Chandra R. Murthy</td>
<td></td>
</tr>
<tr>
<td>Tensor Low Rank Column-Wise Compressive Sensing for Dynamic Imaging</td>
<td>9636</td>
</tr>
<tr>
<td>Silpa Babu, Selin Aviyente, Namrata Vaswani</td>
<td></td>
</tr>
<tr>
<td>Graphmad: Graph Mixup for Data Augmentation Using Data-Driven Convex</td>
<td>9641</td>
</tr>
<tr>
<td>Clustering</td>
<td></td>
</tr>
<tr>
<td>Madeline Navarro, Santiago Segarra</td>
<td></td>
</tr>
<tr>
<td>Neural Feature Predictor and Discriminative Residual Coding for Low-</td>
<td>9646</td>
</tr>
<tr>
<td>Bitrate Speech Coding</td>
<td></td>
</tr>
<tr>
<td>Haici Yang, Wootaeck Lim, Minje Kim</td>
<td></td>
</tr>
<tr>
<td>Towards Bandwidth Estimation for Graph Signal Reconstruction</td>
<td>9651</td>
</tr>
<tr>
<td>Ajinkya Jayawant, Antonio Ortega</td>
<td></td>
</tr>
<tr>
<td>Deep Spectrum Cartography Using Quantized Measurements</td>
<td>9656</td>
</tr>
<tr>
<td>Sabash Timilsina, Sagar Shrestha, Xiao Fu</td>
<td></td>
</tr>
<tr>
<td>M22: Rate-Distortion Inspired Gradient Compression</td>
<td>9661</td>
</tr>
<tr>
<td>Yangyi Liu, Sadaf Salehkalaibar, Stefano Rini, Jun Chen</td>
<td></td>
</tr>
<tr>
<td>Joint Training and Decoding for Multilingual End-To-End Simultaneous</td>
<td>9666</td>
</tr>
<tr>
<td>Speech Translation</td>
<td></td>
</tr>
<tr>
<td>Wuwei Huang, Renren Jin, Wen Zhang, Jian Luan, Bin Wang, Deyi Xiong</td>
<td></td>
</tr>
<tr>
<td>Robust Network Topologies for Distributed Learning</td>
<td>9671</td>
</tr>
<tr>
<td>Chutian Wang, Stefan Vlaski</td>
<td></td>
</tr>
<tr>
<td>TDMA-Based Multi-User Binary Computation Offloading in the Finite-</td>
<td>9676</td>
</tr>
<tr>
<td>Block-Length Regime</td>
<td></td>
</tr>
<tr>
<td>M. Amin Manouchehrpour, Harvinder Lehal, Mahsa Salmani, Timothy N.</td>
<td></td>
</tr>
<tr>
<td>Davidson</td>
<td></td>
</tr>
<tr>
<td>Robustness and Convergence of Mirror Descent for Blind Deconvolution</td>
<td>9681</td>
</tr>
<tr>
<td>Ronak Mehta, Sathya N. Ravi, Vikas Singh</td>
<td></td>
</tr>
<tr>
<td>Generalized Two-Stage Particle Filter for High Dimensions</td>
<td>9686</td>
</tr>
<tr>
<td>Marija Iloska, Mónica F. Bugallo</td>
<td></td>
</tr>
<tr>
<td>A Learnable Spatial Mapping for Decoding the Directional Focus of</td>
<td>9691</td>
</tr>
<tr>
<td>Auditory Attention Using EEG</td>
<td></td>
</tr>
<tr>
<td>Yuanming Zhang, Haoxin Ruan, Ziyun Yuan, Haoliang Du, Xia Gao, Jing Lu</td>
<td></td>
</tr>
<tr>
<td>Multi-Modal Food Classification in a Diet Tracking System with</td>
<td>9696</td>
</tr>
<tr>
<td>Spoken and Visual Inputs</td>
<td></td>
</tr>
<tr>
<td>Shivani Gowda, Yifan Hu, Mandy Korpusik</td>
<td></td>
</tr>
<tr>
<td>A Role Engineering Approach Based on Spectral Clustering Analysis for</td>
<td>9701</td>
</tr>
<tr>
<td>Restful Permissions in Cloud</td>
<td></td>
</tr>
<tr>
<td>Yutang Xia, Yang Luo, Wu Luo, Qingni Shen, Yahui Yang, Zhonghai Wu</td>
<td></td>
</tr>
<tr>
<td>Cross-Modal Matching and Adaptive Graph Attention Network for RGB-D</td>
<td>9706</td>
</tr>
<tr>
<td>Scene Recognition</td>
<td></td>
</tr>
<tr>
<td>Yuhui Guo, Xun Liang, James T. Kwok, Xiangping Zheng, Bo Wu, Yuefeng Ma</td>
<td></td>
</tr>
<tr>
<td>Rapid Audiometric Evaluation for Personalized Headphone Listening</td>
<td>9711</td>
</tr>
<tr>
<td>Matthew J. Goupell, Marjan Davoodian, Sarah Weinstein, David Gadzinski, Dmitry N. Zotkin, Kaushik Sethunath, Ramani Duraiswami</td>
<td></td>
</tr>
</tbody>
</table>
Adaptive Time-Scale Modification for Improving Speech Intelligibility Based on Phoneme Clustering for Streaming Services ................................................................. 9716
Sohee Jang, Jiye Kim, Yeon-Ju Kim, Joon-Hyuk Chang

Unlimited Sampling in Phase Space ........................................................................................................................................ 9721
Peiyu Zhang, Ayush Bhandari

Towards Efficient and Optimal Joint Beamforming and Antenna Selection: A Machine Learning Approach .............................................................. 9726
Sagar Shrestha, Xiao Fu, Mingyi Hong

Exploring Progressive Hybrid-Degraded Image Processing for Homography Estimation ......................................................... 9731
Yijun Lin, Xingzhe Su, Fengge Wu, Junsvuo Zhao

Training Graph Neural Networks on Growing Stochastic Graphs ....................................................................................... 9736
Juan Cerviño, Luana Ruiz, Alejandro Ribeiro

A Fast and Accurate Pitch Estimation Algorithm Based on the Pseudo Wigner-Ville Distribution ........................................ 9741
Yisi Liu, Peter Wu, Alan W Black, Gopala K. Anumanchipalli

Federated Self-Learning with Weak Supervision for Speech Recognition .................................................................................. 9746
Milind Rao, Gopinath Chennumpati, Gautam Tiwari, Anit Kumar Sahu, Anirudh Raju, Ariya Rastrow, Jasha Droppo

Freevc: Towards High-Quality Text-Free One-Shot Voice Conversion .................................................................................. 9751
Jingyi Li, Weiping Tu, Li Xiao

Managing Information Updating with Edge Computing: A Distributed and Learning Approach ........................................ 9756
Junyi He, Di Zhang, Shumeng Liu, Yuezi Zhou, Yaoxue Zhang

Recursive Estimation of User Intent from Noninvasive Electroencephalography Using Discriminative Models ................................................................. 9761
Niklas Smedemark-Margulies, Basak Celik, Tales Imbiriba, Aziz Kocanaogullari, Deniz Erdogmus

Multimodal Microscopy Image Alignment Using Spatial and Shape Information and a Branch-And-Bound Algorithm ......................................................... 9766
Shuonan Chen, Bovey Y Rao, Stephanie Herrlinger, Attila Losonczy, Liam Paninski, Erdem Varol

Learning Cross-Lingual Visual Speech Representations .................................................................................................... 9771
Andreas Zinonos, Alexandros Haliassos, Pingchuan Ma, Stavros Petridis, Maja Pantic

Rethinking Learning-Based Method for Lossless Genome Compression .......................................................................................... 9776
Han Yang, Fei Gu, Jieping Ye

TEFISTA-NET: GTD Parameter Estimation of Low-Frequency Ultra-Wideband Radar Via Model-Based Deep Learning ......................................................... 9781
Rui Li, Xueqian Wang, Gang Li, Xiao-Ping Zhang

Multi-User Methods for Vibrational Radar Backscatter Communications .................................................................................. 9786
Jessica Centers, Jeffrey Krolik

A New Personalized Efficacy Atlas for Pallidal Deep Brain Stimulation .................................................................................. 9791
Xiongbiao Luo
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling the Wave Equation Using Physics-Informed Neural Networks Enhanced with Attention to Loss Weights</td>
<td>9796</td>
</tr>
<tr>
<td>Shaikhah Alkhadhr, Mohamed Almekkawy</td>
<td></td>
</tr>
<tr>
<td>Seri: Sketching-Reasoning-Integrating Progressive Workflow for Empathetic Response Generation</td>
<td>9801</td>
</tr>
<tr>
<td>Guanqun Bi, Yanan Cao, Piji Li, Yuqiang Xie, Fang Fang, Zheng Lin</td>
<td></td>
</tr>
<tr>
<td>On the Minimum Perimeter Criterion for Bounded Component Analysis</td>
<td>9806</td>
</tr>
<tr>
<td>Sergio Cruces</td>
<td></td>
</tr>
<tr>
<td>Weighted Sampling for Masked Language Modeling</td>
<td>9811</td>
</tr>
<tr>
<td>Linhan Zhang, Qian Chen, Wen Wang, Chong Deng, Xin Cao, Kongzhang Hao, Yuxin Jiang, Wei Wang</td>
<td>9816</td>
</tr>
<tr>
<td>Cold Diffusion for Speech Enhancement</td>
<td>9816</td>
</tr>
<tr>
<td>Hao Yen, François G. Germain, Gordon Wichern, Jonathan Le Roux</td>
<td></td>
</tr>
<tr>
<td>Direction Aware Positional and Structural Encoding for Directed Graph Neural Networks</td>
<td>9821</td>
</tr>
<tr>
<td>Yonas Sium, Georgios Kollias, Tsuyoshi Idé, Payel Das, Naoki Abe, Aurélie Lozano, Qi Li</td>
<td>9826</td>
</tr>
<tr>
<td>CryoSWD: Sliced Wasserstein Distance Minimization for 3D Reconstruction in Cryo-Electron Microscopy</td>
<td>9831</td>
</tr>
<tr>
<td>Mona Zehni, Zhizhen Zhao</td>
<td></td>
</tr>
<tr>
<td>Towards a Unified Conformer Structure: from ASR to ASV Task</td>
<td>9836</td>
</tr>
<tr>
<td>Dexin Liao, Tao Jiang, Feng Wang, Lin Li, Qingyang Hong</td>
<td></td>
</tr>
<tr>
<td>Active Learning for Efficient Few-Shot Classification</td>
<td>9836</td>
</tr>
<tr>
<td>Aymane Abdali, Vincent Gripon, Lucas Drumetz, Bartosz Boguslawski</td>
<td></td>
</tr>
<tr>
<td>Robust Spatiotemporal Fusion of Satellite Images Via Convex Optimization</td>
<td>9841</td>
</tr>
<tr>
<td>Ryosuke Isono, Kazuki Naganuma, Shunsuke Ono</td>
<td></td>
</tr>
<tr>
<td>CLIP4VideoCap: Rethinking Clip for Video Captioning with Multiscale Temporal Fusion and Commonsense Knowledge</td>
<td>9846</td>
</tr>
<tr>
<td>Tanvir Mahmud, Feng Liang, Yaling Qing, Diana Marculescu</td>
<td></td>
</tr>
<tr>
<td>Direct Position Determination with One-Bit Signal for Multiple Targets</td>
<td>9851</td>
</tr>
<tr>
<td>Lihua Ni, Di Zhang, Tianyi Xing, Maoyan Ran, Ning Liu, Qun Wan</td>
<td></td>
</tr>
<tr>
<td>Graph-Based Point Cloud Color Denoising with 3-Dimensional Patch-Based Similarity</td>
<td>9856</td>
</tr>
<tr>
<td>Ryosuke Watanabe, Keisuke Nonaka, Eduardo Pavez, Tatsuya Kobayashi, Antonio Ortega</td>
<td></td>
</tr>
<tr>
<td>Music Mixing Style Transfer: A Contrastive Learning Approach to Disentangle Audio Effects</td>
<td>9861</td>
</tr>
<tr>
<td>Junghyun Koo, Marco A. Martinez-Ramirez, Wei-Hsiang Liao, Stefan Uhlich, Kyogu Lee, Yuki Mitsufuji</td>
<td></td>
</tr>
<tr>
<td>FedRPO: Federated Relaxed Pareto Optimization for Acoustic Event Classification</td>
<td>9866</td>
</tr>
<tr>
<td>Meng Feng, Chieh-Chi Kao, Qingming Tang, Amit Solomon, Viktor Rozgic, Chao Wang</td>
<td></td>
</tr>
<tr>
<td>Global and Nodal Mutual Information Maximization in Heterogeneous Graphs</td>
<td>9871</td>
</tr>
<tr>
<td>Costas Mavromatis, George Karypis</td>
<td></td>
</tr>
<tr>
<td>Spherical Sector Harmonics Based Soundfield Radial Extrapolation and Robustness Analysis</td>
<td>9876</td>
</tr>
<tr>
<td>Hanwen Bi, Thushara D. Abhayapala, Fei Ma, Prasanga N. Samarasinghe</td>
<td></td>
</tr>
</tbody>
</table>
Mabnet: Master Assistant Buddy Network with Hybrid Learning for Image Retrieval ................................................. 9881
Rohit Agarwal, Gyanendra Das, Saksham Aggarwal, Alexander Horsch, Dilip K. Prasad

Boundary Cue Guidance and Contextual Feature Mining for Glass Segmentation .................................................. 9886
Qiquan Xiao, Yuan Zhang, Xuanya Li, Kai Hu

Dynamic Local and Global Context Exploration for Small Object Detection ......................................................... 9891
Ziji Zhang, Ping Gong, Haotian Sun, Pingping Wu, Xuanyuan Yang

Distributed Quantum Sensing Network with Geographically Constrained Measurement Strategies ................. 9896
Yingkang Cao, Xiaodi Wu

3D Point Cloud Completion Based on Multi-Scale Degradation ........................................................................... 9901
Jianing Long, Qingmeng Zhu, Hao He, Zhipeng Yu, Qilin Zhang, Zhihong Zhang

Semi-Supervised Graph Ultra-Sparsifier Using Reweighted $\ell_1$ Optimization ..................................................... 9906
Jiayu Li, Tianyun Zhang, Shengmin Jin, Reza Zafarani

Contextually-Rich Human Affect Perception Using Multimodal Scene Information ............................................. 9911
Digbalay Bose, Rajat Hebbar, Krishna Somandepalli, Shrikanth Narayanan

Knowledge-Aware Few Shot Learning for Event Detection from Short Texts ....................................................... 9916
Jinjin Guo, Zhichao Huang, Guangning Xu, Bowen Zhang, Chaoqun Duan

ERSAM: Neural Architecture Search for Energy-Efficient and Real-Time Social Ambiance Measurement ........... 9921
Chaojian Li, Wenwan Chen, Jiayi Yuan, Yingyan Celine Lin, Ashutosh Sabharwal

Conditional Conformer: Improving Speaker Modulation for Single and Multi-User Speech Enhancement ............. 9926
Tom O’Malley, Shaojin Ding, Arun Narayanan, Quan Wang, Rajeev Rikhye, Qiao Liang, Yanzhang He, Ian McGraw

Loss Function Design for DNN-Based Sound Event Localization and Detection on Low-Resource Realistic Data ................................................. 9931
Qing Wang, Jun Du, Zhaoxu Nian, Shutong Niu, Li Chai, Huaxin Wu, Jia Pan, Chin-Hui Lee

EMCLR: Expectation Maximization Contrastive Learning Representations ......................................................... 9936
Meng Liu, Ran Yi, Lizhuang Ma

An Online Algorithm for Contrastive Principal Component Analysis ............................................................... 9941
Siavash Golkar, David Lipshutz, Tiberiu Tesileanu, Dmitri B. Chklovskii

Multi-Object Localization and Irrelevant-Semantic Separation for Nuclei Segmentation in Histopathology Images ......................................................................................... 9946
Ya Tang, Xiongjun Ye, Xuanya Li, Zhineng Chen

Motion-Aware Video Paragraph Captioning Via Exploring Object-Centered Internal Knowledge .................... 9951
Yimin Hu, Guorui Yu, Yuejie Zhang, Rui Feng, Tao Zhang, Xuequan Lu, Shang Gao

Recursive Joint Attention for Audio-Visual Fusion in Regression Based Emotion Recognition .................. 9956
R Gnana Praveen, Eric Granger, Patrick Cardinal

A Reality Check and a Practical Baseline for Semantic Speech Embedding ..................................................... 9961
Guangyu Chen, Yuanyuan Cao
Graph Representation Learning for Stroke Recurrence Prediction .................................................. 9966
Nicholas Glaze, Artun Bayer, Xiaoqian Jiang, Sean Savitz, Santiago Segarra

Toward Auto-Evaluation with Confidence-Based Category Relation-Aware Regression .................. 9971
Jiexin Wang, Jiahao Chen, Bing Su

Fretnet: Continuous-Valued Pitch Contour Streaming for Polyphonic Guitar Tablature Transcription........ 9976
Frank Cwitkowitz, Toni Hirvonen, Anssi Klapuri

Minimising Distortion for GAN-Based Facial Attribute Manipulation............................................. 9981
Mingyu Shao, Li Lu, Ye Ding, Qing Liao

Space-Time Graph Neural Networks with Stochastic Graph Perturbations ........................................ 9986
Samar Hadou, Charilaos I. Kanatsoulis, Alejandro Ribeiro

Interpretation of Neural Networks is Susceptible to Universal Adversarial Perturbations...................... 9991
Haniyeh Ehsani Oskouie, Farzan Farnia

On Designing a 3D Imaging Summer Project for Ontario’s High School Students During Covid-19 Pandemic ......................................................................................................................... 9996
Fengbo Lan, Gene Cheung, Prabhkirat Arora, Deinabo Richard-Koko, Lisa Cole

MUG: A General Meeting Understanding and Generation Benchmark .............................................. 10001
Qinglin Zhang, Chong Deng, Jiaqing Liu, Hai Yu, Qian Chen, Wen Wang, Zhijie Yan, Jinglin Liu, Yi Ren, Zhou Zhao

Active Beam Tracking with Reconfigurable Intelligent Surface ......................................................... 10006
Han Han, Tao Jiang, Wei Yu

Improving Acoustic Echo Cancellation by Mixing Speech Local and Global Features with Transformer ................................................................................................................................. 10011
Yajie Liu, Xinmeng Xu, Weiping Tu, Yuhong Yang, Li Xiao

CNEG-VC: Contrastive Learning Using Hard Negative Example in Non-Parallel Voice Conversion....... 10016
Bima Prihasto, Yi-Xing Lin, Phuong Thi Le, Chien-Lin Huang, Jia-Ching Wang

ITER-SIS: Robust Unlimited Sampling Via Iterative Signal Sieving .................................................. 10021
Ruiming Guo, Ayush Bhandari

Modeling Turn-Taking in Human-To-Human Spoken Dialogue Datasets Using Self-Supervised Features ........................................................................................................................................... 10026
Edmilson Morais, Matheus Damasceno, Hagai Aronowitz, Aharon Satt, Ron Hoory

Higher-Order Link Prediction Via Learnable Maximum Mean Discrepancy .................................... 10031
Georgios V. Karanikolas, Alba Pagès-Zamora, Georgios B. Giannakis

Predicting Multi-Codebook Vector Quantization Indexes for Knowledge Distillation .................... 10036
Liyou Guo, Xiaoyu Yang, Quandong Wang, Yuxiang Kong, Zengwei Yao, Fan Cui, Fangjun Kuang, Wei Kang, Long Lin, Mingshuang Luo, Piotr Zelasko, Daniel Povey

Bert is Robust! a Case Against Word Substitution-Based Adversarial Attacks .............................. 10041
Jens Hauser, Zhao Meng, Damian Pascual, Roger Wattenhofer

When is Mimo Massive in Radar? ........................................................................................................ 10046
Jaimin Shah, Martina Cardone, Alex Dytso, Cynthia Rush
Adapting Self-Supervised Models to Multi-Talker Speech Recognition Using Speaker Embeddings
Zili Huang, Desh Raj, Paola Garcia, Sanjeev Khudanpur

Exploring the Role of Fricatives in Classifying Healthy Subjects and Patients with Amyotrophic Lateral Sclerosis and Parkinson’s Disease
Tanuka Bhattacharjee, Yamini Belur, Atchayaram Nalini, Ravi Yadav, Prasanta Kumar Ghosh

Multiple Access Computation Offloading for the K-User Case
Xiaomeng Liu, Christian Schaible, Timothy N. Davidson

Algebraic Convolutional Filters on Lie Group Algebras
Harshat Kumar, Alejandro Parada-Mayorga, Alejandro Ribeiro

Generalized Invariant Matching Property Via Lasso
Kang Du, Yu Xiang

Exploring Approaches to Multi-Task Automatic Synthesizer Programming
Daniel Faronbi, Iran Roman, Juan Pablo Bello

Adaptive Eccm for Mitigating Smart Jammers
Shashwat Jain, Kunal Pattanayak, Vikram Krishnamurthy, Christopher Berry

Cross-Utterance ASR Rescoring with Graph-Based Label Propagation
Srinath Tankasala, Long Chen, Andreas Stolcke, Anirudh Raju, Qianli Deng, Chander Chandak, Aparna Khare, Roland Maas, Venkatesh Ravichandran

Exploiting Prompt Learning with Pre-Trained Language Models for Alzheimer’s Disease Detection
Yi Wang, Jiajun Deng, Tianzi Wang, Bo Zheng, Shoukang Hu, Xunying Liu, Helen Meng

Representation Learning of Clinical Multivariate Time Series with Random Filter Banks
Alireza Keshavarzian, Hojjat Salehinejad, Shahrokh Valaee

Identifying Source Speakers for Voice Conversion Based Spoofing Attacks on Speaker Verification Systems
Danwei Cai, Zexin Cai, Ming Li

Enhancement of Text-Predicting Style Token with Generative Adversarial Network for Expressive Speech Synthesis
Hiroki Kanagawa, Yusuke Ijima

Static and Dynamic Source and Filter Cues for Classification of Amyotrophic Lateral Sclerosis Patients and Healthy Subjects
Tanuka Bhattacharjee, Chowdam Venkata Thirumala Kumar, Yamini Belur, Atchayaram Nalini, Ravi Yadav, Prasanta Kumar Ghosh

Enhancing Spatio-Spectral Regularization by Structure Tensor Modeling for Hyperspectral Image Denoising
Shingo Takemoto, Shunsuke Ono

Towards Scale Adaptive Underwater Detection Through Refined Pyramid Grid
Xiaoheng Deng, Lirong Liao, Ping Jiang, Yurong Qian

Fast Robust Principle Component Analysis Using Gauss-Newton Iterations
William Chettleburgh, Zhishen Huang, Ming Yan

On Negative Sampling for Contrastive Audio-Text Retrieval
Huang Xie, Okko Räsänen, Tuomas Virtanen
Ideal: Improved Dense Local Contrastive Learning for Semi-Supervised Medical Image Segmentation

Hritam Basak, Soumitri Chattopadhyay, Rohit Kundu, Sayan Nag, Rammohan Mallipeddi

Enhancing Ontology Translation Through Cross-Lingual Agreement

Mingjie Tian, Fausto Giunchiglia, Rui Song, Xing Chen, Hao Xu

Modular Conformer Training for Flexible End-To-End ASR

Kartik Audhkhasi, Brian Farris, Bhuvana Ramabhadran, Pedro J. Moreno

A Speech Representation Anonymization Framework Via Selective Noise Perturbation

Minh Tran, Mohammad Soleymani

Quantum Transfer Learning Using the Large-Scale Unsupervised Pre-Trained Model Wavlm-Large for Synthetic Speech Detection

Ruoyu Wang, Jun Du, Tian Gao

Spammer Detection on Short Video Applications: A New Challenge and Baselines

Muyang Yi, Dong Liang, Rui Wang, Yue Ding, Hongtao Lu

Low Precision Representations for High Dimensional Models

Rajarshi Saha, Mert Pilanci, Andrea J. Goldsmith

Mingling Or Misalignment? Temporal Shift for Speech Emotion Recognition with Pre-Trained Representations

Siyuan Shen, Feng Liu, Aimin Zhou

JaCappella Corpus: A Japanese a Cappella Vocal Ensemble Corpus

Tomohiko Nakamura, Shinnosuke Takamichi, Naoko Tanji, Satoru Fukayama, Hiroshi Sarawatari

Quantum Graph Transformers

Georgios Kollias, Vassilis Kalantzis, Theodoros Salonidis, Shashanka Ubaru

Articulation GAN: Unsupervised Modeling of Articulatory Learning

Gašper Beguš, Alan Zhou, Peter Wu, Gopala K. Anumanchipalli

Active IRS-Assisted MIMO Channel Estimation and Prediction

Mirza A. Haider, Saidur R. Pavel, Yimin D. Zhang, Elias Aboutanios

Efficient Stuttering Event Detection Using Siamese Networks

Payal Mohapatra, Bashima Islam, Md Tamzeed Islam, Ruochen Jiao, Qi Zhu

TOPO-MLP: A Simplicial Network Without Message Passing

Karthikeyan Natesan Ramamurthy, Aldo Guzmán-Sáenz, Mustafa Hajij

Improving Transformer-Based End-to-End Speaker Diarization by Assigning Auxiliary Losses to Attention Heads

Ye-Rin Jeoung, Joon-Young Yang, Jeong-Hwan Choi, Joon-Hyuk Chang

Improving Fairness and Robustness in End-To-End Speech Recognition Through Unsupervised Clustering

Irina-Elena Veliche, Pascale Fung

Progressive Multi-Stage Neural Audio Codec with Psychoacoustic Loss and Discriminator

Byeong Hyeon Kim, Hyungseob Lim, JiHyun Lee, Inseon Jang, Hong-Goo Kang
Medleyvox: An Evaluation Dataset for Multiple Singing Voices Separation ........................................... 10221
  Chang-Bin Jeon, Hyeongi Moon, Keunwoo Choi, Ben Sangbae Chon, Kyogu Lee

On Designing Light-Weight Object Trackers Through Network Pruning: Use CNNS Or Transformers? .................................................. 10226
  Saksham Aggarwal, Taneesh Gupta, Pawan K. Sahu, Arnav Chavan, Rishabh Tiwari, Dilip K. Prasad, Deepak K. Gupta

Improving Non-Autoregressive Speech Recognition with Autoregressive Pretraining.......................... 10231
  Yanjia Li, Lahiru Samarakoon, Ivan Fung

Position-Aware Graph-Based Learning of Whole Slide Images .......................................................... 10236
  Milan Aryal, Nasim Yahya Soltani

DVQVC: An Unsupervised Zero-Shot Voice Conversion Framework .................................................. 10241
  Dayong Li, Xian Li, Xiaofei Li

Enhancing the Accuracy of Resistive In-Memory Architectures Using Adaptive Signal Processing ........ 10246
  Han-Mo Ou, Naresh R. Shanbhag

Learning to Detect Novel and Fine-Grained Acoustic Sequences Using Pretrained Audio Representations ................................................................. 10251
  Vasudha Kowtha, Miquel Espi Marques, Jonathan Huang, Yichi Zhang, Carlos Avendano

MODEFORMER: Modality-Preserving Embedding for Audio-Video Synchronization Using Transformers ..................................................................................... 10256
  Akash Gupta, Rohun Tripathi, Wondong Jang

Compensatory Debiasing for Gender Imbalances in Language Models .................................................. 10261
  Tae-Jin Woo, Woo-Jeoung Nam, Yeong-Joon Ju, Seong-Whan Lee

Federated Semi-Supervised Learning for Object Detection in Autonomous Driving .......................... 10266
  Fangyuan Chi, Yixiao Wang, Panos Nasiopoulos, Victor C. M. Leung, Mahsa T. Pourazad

SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing ....................... 10271
  Siwen Ding, You Zhang, Zhiyao Duan

Choice Fusion as Knowledge for Zero-Shot Dialogue State Tracking ................................................. 10276
  Ruolin Su, Jingfeng Yang, Ting-Wei Wu, Biing-Hwang Juang

Text Classification in the Wild: A Large-Scale Long-Tailed Name Normalization Dataset .................... 10281
  Jiexing Qi, Shuhao Li, Zhixin Guo, Yusheng Huang, Chenghu Zhou, Weinan Zhang, Xinbing Wang, Zhouhan Lin

Reverberation as Supervision for Speech Separation ............................................................................. 10286
  Rohith Aralikatti, Christoph Boeddeker, Gordon Wichern, Aswin Subramanian, Jonathan Le Roux

A Closer Look at Scoring Functions and Generalization Prediction ..................................................... 10291
  Puja Trivedi, Danai Koutra, Jayaraman J. Thiagarajan

Accelerated Distributed Stochastic Non-Convex Optimization Over Time-Varying Directed Networks ........................................................................................................ 10296
  Yiyue Chen, Abolfazl Hashemi, Haris Vikalo
KG-ECO: Knowledge Graph Enhanced Entity Correction for Query Rewriting
Jinglun Cai, Mingda Li, Ziyian Jiang, Eunah Cho, Zheng Chen, Yang Liu, Xing Fan, Chenlei Guo

UML: A Universal Monolingual Output Layer for Multilingual Asr
Chao Zhang, Bo Li, Tara N. Sainath, Trevor Strohman, Shuo-Yin Chang

A Phoneme-Informed Neural Network Model for Note-Level Singing Transcription
Sangeon Yong, Li Su, Juhan Nam

Sub-Band Contrastive Learning-Based Knowledge Distillation for Sound Classification
Achyut Mani Tripathi, Aakansha Mishra

Improving Music Genre Classification from Multi-Modal Properties of Music and Genre Correlations Perspective
Ganghui Ru, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao

Importance of Different Temporal Modulations of Speech: A Tale of Two Perspectives
Samik Sadhu, Hynek Hermansky

A progressive Image Dehazing Framework with Inter and Intra Contrastive Learning
Honglei Xu, Shaohui Liu, Yan Shu, Feng Jiang

Waveform Design to Improve the Estimation of Target Parameters Using the Fourier Transform Method in a MIMO OFDM DFRC System
Satwika Bhogavalli, Eric Grivel, K. V. S. Hari, Vincent Corretja

FNeural Speech Enhancement with Very Low Algorithmic Latency and Complexity Via Integrated Full- And Sub-band Modeling
Zhong-Qiu Wang, Samuele Cornell, Shukjae Choi, Younglo Lee, Byeong-Yeol Kim, Shinji Watanabe

A Gaussian Latent Variable Model for Incomplete Mixed Type Data
Marzieh Ajirak, Petar M. Djuric

DMFormer: Closing the Gap Between CNN and Vision Transformers
Zimian Wei, Hengyue Pan, Lujun Li, Menglong Lu, Xin Niu, Peijie Dong, Dongsheng Li

A Method of Constructing and Automatically Labeling Radio Frequency Signal Training Dataset for UAV
Chao Liu, Ruipeng Ma, Zheng Si, Mingmin Chi

Improving Fast-Slow Encoder Based Transducer with Streaming Deliberation
Ke Li, Jay Mahadeokar, Jinxi Guo, Yangyang Shi, Gil Keren, Ozlem Kalinli, Michael L. Seltzer, Duc Le

Dynamic Alignment Mask CTC: Improved Mask CTC with Aligned Cross Entropy
Xulong Zhang, Haobin Tang, Jianzong Wang, Ning Cheng, Jian Luo, Jing Xiao

Online Caching with Fetching Cost for Arbitrary Demand Pattern: A Drift-Plus-Penalty Approach
Shashank P, B. N. Bharath

Improving EEG-Based Emotion Recognition by Fusing Time-Frequency and Spatial Representations
Kexin Zhu, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao
Avoid Overthinking in Self-Supervised Models for Speech Recognition ......................................................... 10381
  Dan Berrebbi, Brian Yan, Shinji Watanabe

Code-Switching Text Generation and Injection in Mandarin-English ASR .......................................................... 10386
  Haibin Yu, Yixuan Hu, Yao Qian, Ma Jin, Linquan Liu, Shujie Liu, Yu Shi, Yanmin Qian,
  Edward Lin, Michael Zeng

On the Effectiveness of Monoaural Target Source Extraction for Distant End-To-end Automatic
Speech Recognition ......................................................................................................................... 10391
  Catalin Zorila, Rama Doddipatla

Channel Estimation with Tightly-Coupled Antenna Arrays ........................................................................... 10396
  Bamelak Tadele, Volodymyr Shyianov, Faouzi Bellili, Amine Mezghani

AutoGCF: Personalized Aggregation on Neural Graph Collaborative Filtering ........................................ 10401
  Xiaoyu You, Chi Li, Jianwei Xu, Mi Zhang

Dual Collaborative Visual-Semantic Mapping for Multi-Label Zero-Shot Image Recognition ................ 10406
  Yunting Hu, Xuan Jin, Xi Chen, Yin Zhang

Joint Estimation of DOA and Distance in Noisy Reverberant Conditions .................................................. 10411
  Suliang Bu, Tuo Zhao, Yunxin Zhao

Communication-Constrained Exchange of Zeroth-Order Information with Application to
Collaborative Target Tracking ........................................................................................................ 10416
  Ege Can Kaya, Mehmet Berk Sahin, Abolfazl Hashemi

On the Robustness of Non-Intrusive Speech Quality Model by Adversarial Examples .................... 10421
  Hsin-Yi Lin, Huan-Hsin Tseng, Yu Tsao

More Speaking Or More Speakers? .......................................................................................................... 10426
  Dan Berrebbi, Ronan Collobert, Navdeep Jaitly, Tatiana Likhomanenko

Filler Word Detection with Hard Category Mining and Inter-Category Focal Loss .............................. 10431
  Zhiyuan Zhao, Lijun Wu, Chuanxin Tang, Dacheng Yin, Yucheng Zhao, Chong Luo

Unsupervised Video Anomaly Detection for Stereotypical Behaviours in Autism ................................ 10436
  Jiaqi Gao, Xinyang Jiang, Yuqing Yang, Dongsheng Li, Lili Qiu

HPFTN: Hierarchical Progressive Fusion Transformer Network for Video Denoising ........................ 10441
  Shuatao Zhang, Yuan Zhang, Zheng Zhao, Di Xie, Shiliang Pu

Dasformer: Deep Alternating Spectrogram Transformer for Multi/Single-Channel Speech
Separation ......................................................................................................................................... 10446
  Shuo Wang, Xiangyu Kong, Xiulian Peng, Hesam Movassagh, Vinod Prakash, Yan Lu

Real-Time Human Reconstruction Based on Human Pose Prior and Epipolar Refinement ............... 10451
  Kuncheng Luo, Zhiheng Li

Coded Matrix Computations for D2D-Enabled Linearized Federated Learning .................................... 10456
  Anindya Bijoy Das, Aditya Ramamoorthy, David J. Love, Christopher G. Brinton

NAS-DYMC: NAS-Based Dynamic Multi-Scale Convolutional Neural Network for Sound Event
Detection ............................................................................................................................................. 10461
  Jun Wang, Peng Yao, Feng Deng, Jianchao Tan, Chengru Song, Xiaorui Wang
Signal Processing and Quantum State Tomography on Noisy Devices................................................................. 10466
   Wenbo Shi, Robert Malaney

Class-Aware Contextual Information for Semantic Segmentation ................................................................. 10471
   Huadong Tang, Youpeng Zhao, Yingying Jiang, Zhuoxin Gan, Qiang Wu

Dual-Stage Graph Convolution Network with Graph Learning for Traffic Prediction ................................. 10476
   Zilong Li, Qianqian Ren, Long Chen, Jianguo Sun

TFCnet: Time-Frequency Domain Corrector for Speech Separation .............................................................. 10481
   Weinan Tong, Jiaxu Zhu, Jun Chen, Zhiyong Wu, Shiyin Kang, Helen Meng

Conversation-Oriented ASR with Multi-Look-Ahead CBS Architecture ......................................................... 10486
   Huaibo Zhao, Shinya Fujie, Tetsuji Ogawa, Jin Sakuma, Yusuke Kida, Tetsunori Kobayashi

Real-Time Audio-Visual End-To-End Speech Enhancement .............................................................................. 10491
   Zirun Zhu, Hemin Yang, Min Tang, Ziyi Yang, Sefik Emre Eskimez, Huaming Wang

Unsupervised Anomaly Detection and Localization of Machine Audio: A Gan-Based Approach .................. 10496
   Anbai Jiang, Wei-Qiang Zhang, Yufeng Deng, Pingyi Fan, Jia Liu

Information Extraction from Pill Bottle Images Via Text Stitching ................................................................. 10501
   Rahul Kumar Gupta, Shilka Roy, Sujit Jos, Unni V. S., Lauren Lavoie, Frederic Medous, Walter Smith

VQ-CL: Learning Disentangled Speech Representations with Contrastive Learning and Vector
   Quantization ........................................................................................................................................ 10506
   Huaizhen Tang, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao

No Reference Quality Assessment for Screen Content Images Based on Entire and High-Influence
   Regions................................................................................................................................................ 10511
   Zhuoran Xu, Yang Yang, Zhixiang Zhang, Weiming Zhang

Enhanced Dcf Tracker Regularized by Reliable Sample Construction ............................................................. 10516
   Kun Hu, Mingyu Cao, Mengzhu Wang, Long Lan, Wenjing Yang, Huibin Tan

Sparse Delay-Doppler Channel Estimation for OTFS Modulation Using 2D-Music ........................................ 10521
   Akshay S. Bondre, Christ D. Richmond, Ahmed Alkhateeb, Nicolò Michelusi

A Contrastive Knowledge Transfer Framework for Model Compression and Transfer Learning ............. 10526
   Kaiqi Zhao, Yitao Chen, Ming Zhao

Lightweight Feature Encoder for Wake-Up Word Detection Based on Self-Supervised Speech
   Representation ........................................................................................................................................ 10531
   Hyungjun Lim, Younggwan Kim, Kiho Yeom, Eunjoo Seo, Hoodong Lee, Stanley Jungkyu Choi, Honglak Lee

Kernel Ridge Regression for Generalized Graph Signal Processing .......................................................... 10536
   Xingchao Jian, Wee Peng Tay

Adversarial Attacks on Genotype Sequences .................................................................................................... 10541
   Daniel Mas Montserrat, Alexander G. Ioannidis

Learning a Weight Map for Weakly-Supervised Localization .................................................................... 10546
   Tal Shaharabany, Lior Wolf
Transcription Free Filler Word Detection with Neural Semi-CRFs................................................. 10551
  Ge Zhu, Yujia Yan, Juan-Pablo Caceres, Zhiyao Duan

Pyramid Spatial Feature Transform and Shared-Offsets Deformable Alignment Based Convolutional Network for HDR Imaging ................................................................. 10556
  Junda Liao, Qin Liu, Takeshi Ikenaga

Balanced Deep CCA for Bird Vocalization Detection................................................................. 10561
  Sumit Kumar, B. Anshuman, Linus Rüttimann, Richard H. R. Hahnloser, Vipul Arora

Joint Noise Reduction and Listening Enhancement for Full-End Speech Enhancement .................. 10566
  Haoyu Li, Yun Liu, Junichi Yamagishi

Restoration of Time-Varying Graph Signals Using Deep Algorithm Unrolling............................ 10571
  Hayate Kojima, Hikari Noguchi, Koki Yamada, Yuichi Tanaka

Mask the Bias: Improving Domain-Adaptive Generalization of CTC-Based ASR with Internal Language Model Estimation................................................................. 10576
  Nilaksh Das, Monica Sankara, Sranan Bodapati, Jinglun Cai, Devang Kulshreshtha, Jeff Farris, Katrin Kirchhoff

Parallel Sentence-Level Explanation Generation for Real-World Low-Resource Scenarios .............. 10581
  Yan Liu, Xiaokang Chen, Qi Dai

Comparative Layer-Wise Analysis of Self-Supervised Speech Models......................................... 10586
  Ankita Pasad, Bowen Shi, Karen Livescu

DepthFormer: Multimodal Positional Encodings and Cross-Input Attention for Transformer-based Segmentation Networks................................................................. 10591
  Francesco Barbato, Giulia Rizzoli, Pietro Zanuttigh

Improved Belief Propagation Decoding of Turbo Codes ............................................................ 10596
  Yifei Shen, Yuqing Ren, Andreas Toftegaard Kristensen, Xiaohu You, Chuan Zhang, Andreas Burg

Learning Speech Representations with Flexible Hidden Feature Dimensions................................ 10601
  Huaizhen Tang, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao

Personalized Lightweight Text-To-Speech: Voice Cloning with Adaptive Structured Pruning........... 10606
  Sung-Feng Huang, Chia-Ping Chen, Zhi-Sheng Chen, Yu-Pao Tsai, Hung-Yi Lee

QI-TTS: Questioning Intonation Control for Emotional Speech Synthesis..................................... 10611
  Haobin Tang, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao

Using Adapters to Overcome Catastrophic Forgetting in End-To-End Automatic Speech Recognition ................................................................. 10616
  Steven Vander Eeckt, Hugo Van Hamme

Partially Adaptive Multichannel Joint Reduction of Ego-Noise and Environmental Noise............. 10621
  Huajian Fang, Niklas Wittner, Johannes Twiefel, Stefan Wermter, Timo Gerkmann

URM4DMU: An User Representation Model for Darknet Markets Users..................................... 10626
  Hongmeng Liu, Jiapeng Zhao, Yixuan Huo, Yuyan Wang, Chun Liao, Liyan Shen, Shiyao Cui, Jinqiao Shi
A Probabilistic Framework for Pruning Transformers Via a Finite Admixture of Keys
Tan M. Nguyen, Tam Nguyen, Long Bui, Hai Do, Duy Khuong Nguyen, Dung D. Le, Hung Tran-The, Nhat Ho, Stan J. Osher, Richard G. Baraniuk

Asynchronous Federated Learning for Real-Time Multiple Licence Plate Recognition Through Semantic Communication
Renyou Xie, Chaojie Li, Xiaojun Zhou, Zhaoyang Dong

Image Reconstruction Without Explicit Priors
Angela F. Gao, Oscar Leong, He Sun, Katherine L. Bouman

DGN: Descriptor Generation Network for Feature Matching in Monocular Endoscopy 3D Reconstruction
Kaiyun Zhang, Wenkang Fan, Yinran Chen, Xiongbiao Luo

Alternating Phase Langevin Sampling with Implicit Denoiser Priors for Phase Retrieval
Rohun Agrawal, Oscar Leong

Emotion Recognition in Conversation from Variable-Length Context
Mian Zhang, Xiaobing Zhou, Wenliang Chen, Min Zhang

Learning to Regularized Resource Allocation with Budget Constraints
Shaoke Fang, Qingsong Liu, Lei Xu, Wenfei Wu

Solving Audio Inverse Problems with a Diffusion Model
Eloi Moliner, Jaakko Lehtinen, Vesa Välimäki

MRNET: Multi-Refinement Network for Dual-Pixel Images Defocus Deblurring
Dafeng Zhang, Xiaobing Wang, Zhezhu Jin

Articulatory Representation Learning Via Joint Factor Analysis and Neural Matrix Factorization
Jiachen Lian, Alan W Black, Yijing Lu, Louis Goldstein, Shinji Watanabe, Gopala K. Anumanchipalli

Pretraining Conformer with ASR for Speaker Verification
Danwei Cai, Weiqing Wang, Ming Li, Rui Xia, Chuanzeng Huang

Deep Triple-Supervision Learning Unannotated Surgical Endoscopic Video Data for Monocular Dense Depth Estimation
Wenkang Fan, Kaiyun Zhang, Hong Shi, Jianhua Chen, Yinran Chen, Xiongbiao Luo

VE-KWS: Visual Modality Enhanced End-to-End Keyword Spotting
Ao Zhang, He Wang, Pengcheng Guo, Yihui Fu, Lei Xie, Yingying Gao, Shilei Zhang, Junlan Feng

StreamSpeech: Low-Latency Neural Architecture for High-Quality on-Device Speech Synthesis
Georgi Shopov, Stefan Gerdjikov, Stoyan Mihov

Local-Global Progressive U-Transformers for Accurate Hepatic and Portal Veins Segmentation in Abdominal MR Images
Yu Wu, Dongliang Shen, Jiabao Jin, Guanping Xu, Yinran Chen, Xiongbiao Luo

ACF: Aligned Contrastive Finetuning for Language and Vision Tasks
Wei Zhu, Peng Wang, Xiaoqin Wang, Yuan Ni, Guotong Xie

DB-UNet: MLP Based Dual Branch UNet for Accurate Vessel Segmentation in OCTA Images
Chengliang Wang, Haojian Ning, Xinrun Chen, Shiyong Li
Structured Pruning of Self-Supervised Pre-Trained Models for Speech Recognition and Understanding ................................................................. 10716
Yifan Peng, Kwangyoun Kim, Felix Wu, Prashant Sridhar, Shinji Watanabe

Margin-Mixup: A Method for Robust Speaker Verification in Multi-Speaker Audio .................................................. 10721
Jenthe Thienpondt, Nilesh Madhu, Kris Demuynck

EI2SR: Learning an Enhanced Intra-Instance Semantic Relationship for Arbitrary-Shaped Scene Text Detection .................................................. 10726
Yan Shu, Shaohui Liu, Yu Zhou, Honglei Xu, Feng Jiang

Numerical Semantic Modeling for Implicit Discourse Relation Recognition ................................................................. 10731
Chenxu Wang, Ping Jian, Hai Wang

Forensics for Adversarial Machine Learning Through Attack Mapping Identification .................................................. 10736
Allen Yan, Jinsub Kim, Raviv Raich

AMPose: Alternately Mixed Global-Local Attention Model for 3D Human Pose Estimation .................................................. 10741
Hongxin Lin, Yunwei Chiu, Peiyuan Wu

Cross-Modal Mutual Learning for Cued Speech Recognition ................................................................. 10746
Lei Liu, Li Liu

Multi-Dimensional Frequency Dynamic Convolution with Confident Mean Teacher for Sound Event Detection ................................................................. 10751
Shengchang Xiao, Xueshuai Zhang, Pengyuan Zhang

Boosting Face Recognition Performance with Synthetic Data and Limited Real Data .................................................. 10756
Wenqing Wang, Lingqing Zhang, Chi-Man Pun, Jiu-Cheng Xie

The Potential of Neural Speech Synthesis-Based Data Augmentation for Personalized Speech Enhancement ................................................................. 10761
Anastasia Kuznetsova, Aswin Sivaraman, Minje Kim

Deep Neural Mel-Subband Beamformer for in-Car Speech Separation ................................................................. 10766
Vinay Kothapally, Yong Xu, Meng Yu, Shi-Xiong Zhang, Dong Yu

Multi-Task Bias-Variance Trade-Off Through Functional Constraints ................................................................. 10771
Juan Cerviño, Juan Andrés Bazerque, Miguel Calvo-Fullana, Alejandro Ribeiro

M-CTRL: A Continual Representation Learning Framework with Slowly Improving Past Pre-Trained Model ................................................................. 10776
Jin-Seong Choi, Jae-Hong Lee, Chae-Won Lee, Joon-Hyuk Chang

Designing and Evaluating Speech Emotion Recognition Systems: A Reality Check Case Study with IEMOCAP ................................................................. 10781
Nikolaos Antoniou, Athanasios Katsamanis, Theodoros Giannakopoulos, Shrikanth Narayanan

A Deep Fusion Rule for Infrared and Visible Image Fusion: Feature Communication for Importance Assessment ................................................................. 10786
Xuran Lv, Jinyong Cheng, Guohua Lv, Zhonghe Wei

Optimal Kernel for Real-Time Arbitrary-Shaped Text Detection ................................................................. 10791
Haozhao Ma, Chuang Yang, Yuan Yuan, Qi Wang
Particle Flow Gaussian Sum Particle Filter ................................................................. 10796
Karthik Comandur, Yunpeng Li, Santosh Nannuru

Self-Supervised Accent Learning for Under-Resourced Accents Using Native Language Data .......... 10801
Mehul Kumar, Jiyeon Kim, Dhananjaya Gowda, Abhinav Garg, Chanwook Kim

Learning to Locate the Text Forgery in Smartphone Screenshots ........................................ 10806
Zeqin Yu, Bin Li, Yuzhen Lin, Jinhua Zeng, Jishen Zeng

Noncoherent Multiuser Grassmannian Constellations for the Mimo Multiple Access Channel ........... 10811
Javier Alvarez-Vizoso, Diego Cuevas, Carlos Beltrán, Ignacio Santamaria, Vit Tucek, Gunnar Peters

Attention-Guided Deep Learning Framework for Movement Quality Assessment .......................... 10816
Aditya Kanade, Mansi Sharma, Manivannan Muniyandi

Learning from Yourself: A Self-Distillation Method for Fake Speech Detection .......................... 10821
Jun Xue, Cunhang Fan, Jiangyan Yi, Chenglong Wang, Zhengqi Wen, Dan Zhang, Zhao Ly

Rate-Distortion Optimization with Alternative References for UGC Video Compression .................. 10826
Xin Xiong, Eduardo Pavez, Antonio Ortega, Balu Adsumilli

MDR-MFI: Multi-Branch Decoupled Regression and Multi-Scale Feature Interaction for Partial-to-
Partial Cloud Registration ....................................................................................... 10831
Weidong Dai, Xuejun Yan, Jingjing Wang, Di Xie, Shiliang Pu

Static-Scene Constrained Optimization for Matrix/Tensor-Decomposition-free Foreground-
Background Separation ......................................................................................... 10836
Kazuki Naganuma, Shunsuke Ono

Exploring Universal Singing Speech Language Identification Using Self-Supervised Learning
Based Front-End Features .................................................................................... 10841
Xingming Wang, Hao Wu, Chen Ding, Chuanzeng Huang, Ming Li

A Hybrid Deep Neural Network for Nonlinear Causality Analysis in Complex Industrial Control
System .................................................................................................................. 10846
Tian Feng, Qiming Chen, Yao Shi, Xun Lang, Lei Xie, Hongye Su

Binauralization Robust to Camera Rotation Using 360° Videos ................................................. 10851
Masaki Yoshida, Ren Togo, Takahiro Ogawa, Miki Haseyama

KEPS-NET: Robust Parking Slot Detection Based Keypoint Estimation for High Localization
Accuracy .............................................................................................................. 10856
Jaewoo Lee, Kapje Sung, Daepal Park, Younghan Jeon

Analysing Diffusion-Based Generative Approaches Versus Discriminative Approaches for Speech
Restoration ............................................................................................................ 10861
Jean-Marie Lemercier, Julius Richter, Simon Welker, Timo Gerkmann

EMIX: A Data Augmentation Method for Speech Emotion Recognition ...................................... 10866
An Dang, Toan H. Vu, Le Dinh Nguyen, Jia-Ching Wang

SDRNet: Shape Decoupled Regression Network for 3D Face Reconstruction ............................ 10871
Shikun Zhang, Fengyi Song, Ge Song, Ming Yang
Low in Resolution, High in Precision: UAV Detection with Super-Resolution and Motion Information Extraction .......................................................... 10876  
Hanzhuo Wang, Xingjian Wang, Chengwei Zhou, Wenchao Meng, Zhiguo Shi

Self-Supervised Hierarchical Metrical Structure Modeling .......................................................... 10881  
Junyan Jiang, Gus Xia

Cosmopolite Sound Monitoring (CoSMo): A Study of Urban Sound Event Detection Systems Generalizing to Multiple Cities .......................................................... 10886  
Florian Angulo, Slim Essid, Geoffroy Peeters, Christophe Mietlicki

Improving Heart Rate and Heart Rate Variability Estimation from Video Through a HR-RR-Tuned Filter ............................................................................ 10891  
Michael Chan, Li Zhu, Korosh Vatanparvar, Hewon Jung, Jilong Kuang, Alex Gao

Dense Adversarial Transfer Learning Based on Class-Invariance ............................................................................ 10896  
Bach-Tung Pham, Ting-Yu Wang, Phuong Le Thi, Khai-Thinh Nguyen, Yuan-Shan Lee, Tzu-Chiang Tai, Jia-Ching Wang

Uncertainty Estimation in Deep Speech Enhancement Using Complex Gaussian Mixture Models .......... 10901  
Huajian Fang, Timo Gerkmann

Multi-View Learning for Speech Emotion Recognition with Categorical Emotion, Categorical Sentiment, and Dimensional Scores .......................................................... 10906  
Daniel Tompkins, Dimitra Emmanouilidou, Soham Deshmukh, Benjamin Elizalde

Performance Above All? Energy Consumption Vs. Performance, a Study on Sound Event Detection with Heterogeneous Data .......................................................... 10911  
Romain Serizel, Samuele Cornell, Nicolas Turpault

Unsupervised Action Segmentation of Untrimmed Egocentric Videos ............................................................................ 10916  
Sam Perochon, Laurent Oudre

Self-Remixing: Unsupervised Speech Separation VIA Separation and Remixing .......................................................... 10921  
Kohei Saijo, Tetsuji Ogawa

Stargan-Vc Based Cross-Domain Data Augmentation for Speaker Verification .......................................................... 10926  
Hang-Rui Hu, Yan Song, Jian-Tao Zhang, Li-Rong Dai, Ian McLoughlin, Zhu Zhuo, Yu Zhou, Yu-Hong Li, Hui Xue

Investigation into Phone-Based Subword Units for Multilingual End-to-End Speech Recognition .......... 10931  
Saierdaer Yusuyin, Hao Huang, Junhua Liu, Cong Liu

Bat: Bi-Alignment Based on Transformation in Multi-Target Domain Adaptation for Semantic Segmentation ............................................................................ 10936  
Xian Zhong, Wei Li, Liang Liao, Jing Xiao, Wenzuan Liu, Wenzin Huang, Zheng Wang

Building Change Detection Using Cross-Temporal Feature Interaction Network .......................................................... 10941  
Yuchao Feng, Jiawei Jiang, Honghui Xu, Jianwei Zheng

Flowpose: Conditional Normalizing Flows for 3D Human Pose and Shape Estimation from Monocular Videos ............................................................................ 10946  
Yaoyao Du, Zixiao Zhang, Zhihao Li, Peng Wei, Qingmin Liao, Wenming Yang

Random Projector: Efficient Deep Image Prior ............................................................................ 10951  
Taihui Li, Zhong Zhuang, Hengkang Wang, Ju Sun
Background-Weakening Consistency Regularization for Semi-Supervised Video Action Detection
Xian Zhong, Aoyu Yi, Wenxuan Liu, Wenxin Huang, Chengming Zou, Zheng Wang

Convolutional NTF for Ambisonic Source Separation Under Reverberant Conditions
Mateusz Guzik, Konrad Kowalczyk

Calibrating AI Models for Few-Shot Demodulation VIA Conformal Prediction
Kfir M. Cohen, Sangwoo Park, Osvaldo Simeone, Shlomo Shamai Shitz

Rethinking Random Walk in Graph Representation Learning
Dingyi Zeng, Wenyu Chen, Wanlong Liu, Li Zhou, Hong Qu

Beyond Rate Coding: Signal Coding and Reconstruction Using Lean Spike Trains
Anik Chattopadhyay, Arunava Banerjee

Detecting Malicious Migration on Edge to Prevent Running Data Leakage
Yuchen Wong, Qingni Shen, Cong Li, Cunzhan Liu, Tianxiang Ai

Deformable Cross Attention for Learning Optical Flow
Rokia Abdein, Xuezhi Xiang, Ning Lv, Abdulmotaleb El Saddik

Optimized Quality Feature Learning for Video Quality Assessment
Ngai-Wing Kwong, Yui-Lam Chan, Sik-Ho Tsang, Daniel Pak-Kong Lun

Next-Speaker Prediction Based on Non-Verbal Information in Multi-Party Video Conversation
Saki Mizuno, Nobukatsu Hojo, Satoshi Kobashikawa, Ryo Masumura

Second-Order Statistic Deviation to Model Anomalies in the Design of Unsupervised Detectors
Andriy Enttsel, Filippo Martinini, Alex Marchioni, Mauro Mangia, Riccardo Rovatti, Gianluca Setti

Clustering-Based Supervised Contrastive Learning for Identifying Risk Items on Heterogeneous Graph
Ao Li, Yugang Ji, Guanyi Chu, Xiao Wang, Dong Li, Chuan Shi

DiffPhase: Generative Diffusion-Based STFT Phase Retrieval
Tal Peer, Simon Welker, Timo Gerkmann

LiteG2P: A Fast, Light and High Accuracy Model for Grapheme-To-Phoneme Conversion
Chunfeng Wang, Peisong Huang, Yuxiang Zou, Haoyu Zhang, Shichao Liu, Xiang Yin, Zejun Ma

Expressive-VC: Highly Expressive Voice Conversion with Attention Fusion of Bottleneck and Perturbation Features
Ziqian Ning, Qicong Xie, Pengcheng Zhu, Zhichao Wang, Liuneng Xue, Jixun Yao, Lei Xie, Mengxiao Bi

Recouple Event Field Via Probabilistic Bias for Event Extraction
Xingyu Bai, Taqiang Wu, Han Guo, Zhe Zhao, Xuefeng Yang, Jiayi Li, Weijie Liu, Qi Ju, Weigang Guo, Yajiu Yang

Self-Supervised Learning for Speech Enhancement Through Synthesis
Bryce Irvin, Marko Stamenovic, Mikolaj Kegler, Li-Chia Yang

Rigid-Body Sound Synthesis with Differentiable Modal Resonators
Rodrigo Diaz, Ben Hayes, Charalampos Saitis, György Fazekas, Mark Sandler
Unsupervised Voice Type Discrimination Score Adaptation Using X-Vector Clusters .................................................11041
Mark Lindsey, Tyler Vuong, Richard M. Stern

Toroidal Probabilistic Spherical Discriminant Analysis ..................................................................................................11046
Anna Silnova, Niko Brümmer, Albert Swart, Lukáš Burget

Large-Scale Nonverbal Vocalization Detection Using Transformers ..................................................................................11051
Panagiotis Tzirakis, Alice Baird, Jeffrey Brooks, Christopher Gagne, Lauren Kim, Michael Opara, Christopher Gregory, Jacob Metrick, Garrett Boseck, Vineet Tiruvadi, Björn Schuller, Dacher Keltner, Alan Cowen

Multilingual Query-By-Example Keyword Spotting with Metric Learning and Phoneme-to-Embedding Mapping .................................................................11056
Paul M. Reuter, Christian Rollwage, Bernd T. Meyer

Exploiting 3D Human Recovery for Action Recognition with Spatio-Temporal Bifurcation Fusion ..................................11061
Na Jiang, Wei Quan, Qichuan Geng, Zhiping Shi, Peng Xu

Fast and Efficient Speech Enhancement with Variational Autoencoders ........................................................................11066
Mostafa Sadeghi, Romain Serizel

Joint Compression and Demosaicking for Satellite Images ..................................................................................................11071
Pascal Bacchus, Renaud Fraisse, Aline Roumy, Christine Guillemot

Think Before You Speak: Concept-Guided Explicit Persona Reasoning for Personalized Dialogue Generation .................................................................11076
Yunpeng Li, Yue Hu, Wei Peng, Yuqiang Xie

Distributed Bayesian Tracking on the Special Euclidean Group Using Lie Algebra Parametric Approximations .................................................................11081
Claudio J. Bordin, Caio G. De Figueredo, Marcelo G. S. Bruno

Fast and Parallel Decoding for Transducer ..........................................................................................................................11086
Wei Kang, Liyong Guo, Fangjun Kuang, Long Lin, Mingshuang Luo, Zengwei Yao, Xiaoyu Yang, Piotr Zelasko, Daniel Povey

Difficulty-Aware Data Augmentor for Scene Text Recognition ..............................................................................................11091
Guanghao Meng, Tao Dai, Bin Chen, Naiqi Li, Yong Jiang, Shu-Tao Xia

Learning Interpretable Filters in Wav-UNet for Speech Enhancement ................................................................................11096
Félix Mathieu, Thomas Courtat, Gaël Richard, Geoffroy Peeters

Background Disturbance Mitigation for Video Captioning Via Entity-Action Relocation .....................................................11101
Zipeng Li, Xian Zhong, Shuqin Chen, Wenxuan Liu, Wenzin Huang, Lin Li

Sinusoidal Frequency Estimation by Gradient Descent ..........................................................................................................11106
Ben Hayes, Charalampos Saitis, György Fazekas

Multi-View K-Means with Laplacian Embedding ....................................................................................................................11111
Zhezheng Hao, Zhoumin Lu, Feiping Nie, Rong Wang, Xuelong Li

Going in Style: Audio Backdoors Through Stylistic Transformations ..................................................................................11116
Stefanos Koffas, Luca Pajola, Stjepan Picek, Mauro Conti

Model-Based Vs. Data-driven Approaches for Predicting Rain-induced Attenuation in Commercial Microwave Links: A Comparative Empirical Study ...................................................................................................................11121
Dror Jacoby, Jonatan Ostrometzky, Hagit Messer
Joint Multi-Level Feature Network for Lightweight Person Re-Identification ........................................11126
  Yunzuo Zhang, Weili Kang, Yameng Liu, Pengfei Zhu

Audio Signal Enhancement with Learning from Positive and Unlabeled Data ..................................11131
  Nobutaka Ito, Masashi Sugiyama

Active Perception System for Enhanced Visual Signal Recovery Using Deep Reinforcement Learning ........................................11136
  Gaurav Chaudhary, Laxmidhar Behera, Tushar Sandhan

Federated Learning for ASR Based on Wav2vec 2.0 .................................................................11141
  Tuan Nguyen, Salima Mdhaffar, Natalia Tomashenko, Jean-François Bonastre, Yannick Estève

Learning on Graphs Under Label Noise ..........................................................................................11146
  Jingyang Yuan, Xiao Luo, Yifang Qin, Yusheng Zhao, Wei Ju, Ming Zhang

Robust Hyperspectral Anomaly Detection with Simultaneous Mixed Noise Removal Via Constrained Convex Optimization ........................................11151
  Koyo Sato, Shunsuke Ono

Improving Knowledge Distillation for Non-Intrusive Load Monitoring Through Explainability Guided Learning .................................................................11156
  Djordje Batic, Giulia Tanoni, Lina Stankovic, Vladimir Stankovic, Emanuele Principi

Learning Unbiased Rewards with Mutual Information in Adversarial Imitation Learning ..................11161
  Lihua Zhang, Quan Liu, Zhigang Huang, Lan Wu

Confidence-Based Event-Centric Online Video Question Answering on a Newly Constructed ATBS Dataset ..................................................................................................................................11166
  Weikai Kong, Shuhong Ye, Chenglin Yao, Jianfeng Ren

Auto-AVSR: Audio-Visual Speech Recognition with Automatic Labels ..............................................11171
  Pingchuan Ma, Alexandros Haliassos, Adriana Fernandez-Lopez, Honglie Chen, Stavros Petridis, Maja Pantic

ClassA Entropy for the Analysis of Structural Complexity of Physiological Signals ............................11176
  Hongjian Xiao, Ling Li, Danilo P. Mandic

Classifying Pathological Images Based on Multi-Instance Learning and End-to-End Attention Pooling ..................................................................................................................................11181
  Yuqi Chen, Juan Liu, Zhiqun Zuo, Peng Jiang, Yu Jin, Guangsheng Wu

Imaginary Voice: Face-Styled Diffusion Model for Text-to-Speech .....................................................11186
  Jiyoung Lee, Joon Son Chung, Soo-Whan Chung

Learning to Balance the Global Coherence and Informativeness in Knowledge-Grounded Dialogue Generation ..................................................................................................................11191
  Chenxu Niu, Yue Hu, Wei Peng, Yuqiang Xie

Note and Playing Technique Transcription of Electric Guitar Solos in Real-World Music Performance ..........................................................................................................................11196
  Tung-Sheng Huang, Ping-Chung Yu, Li Su

Differential Analysis for Networks Obeying Conservation Laws ..........................................................11201
  Anirudh Rayas, Rajasekhar Anguluri, Jiajun Cheng, Gautam Dasarathy
CPD-GAN: Cascaded Pyramid Deformation GAN for Pose Transfer .................................................................11206
Yuan Huang, Yuting Tang, Xiu Zheng, Jie Tang

Training Robust Spiking Neural Networks with Viewpoint Transform and Spatiotemporal Stretching .................................................................11211
Haibo Shen, Juyu Xiao, Yihao Luo, Xiang Cao, Liangqi Zhang, Tianjiang Wang

BIRD-PCC: Bi-Directional Range Image-Based Deep Lidar Point Cloud Compression ..................................11216
Chia-Sheng Liu, Jia-Fong Yeh, Hao Hsu, Hung-Ting Su, Ming-Sui Lee, Winston H. Hsu

Sparse Error Correction for Power Network Parameters.....................................................................................11221
Dilan Senaratne, Jinsub Kim

Higher-Order Spatio-Temporal Neural Networks for Covid-19 Forecasting .......................................................11226
Yuzhou Chen, Sotiris Batsakis, H. Vincent Poor

Improving Accented Speech Recognition with Multi-Domain Training..........................................................11231
Lucas Maison, Yannick Esteve

Fully Unsupervised Topic Clustering of Unlabelled Spoken Audio Using Self-Supervised Representation Learning and Topic Model......................................................11236
Takashi Maekaku, Yuya Fujita, Xuankai Chang, Shinji Watanabe

Centralized Cascade Multi-Channel Noise Reduction and Acoustic Feedback Cancellation in a Wireless Acoustic Sensor and Actuator Network ..................................................11241
Santiago Ruiz, Toon Van Waterschoot, Marc Moonen

Hindi as a Second Language: Improving Visually Grounded Speech with Semantically Similar Samples ..................................................................................11246
Hyeonggon Ryu, Arda Senocak, In So Kweon, Joon Son Chung

Disentangling Speech from Surroundings with Neural Embeddings ..................................................................11251
Ahmed Omran, Neil Zeghidour, Zalán Borsos, Félix De Chaumont Quitry, Malcolm Slaney, Marco Tagliasacchi

Progressive Perception Learning for Distribution Modulation in Siamese Tracking .......................................11256
Kun Hu, Xianchen Zhou, Mingyu Cao, Mengshu Wang, Guangjie Gao, Wenjing Yang, Huibin Tan

Pooling Strategies for Simplicial Convolutional Networks ............................................................................11261
Domenico Mattia Cinque, Claudio Battiloro, Paolo Di Lorenzo

Efficient Uncertainty Estimation with Gaussian Process for Reliable Dialog Response Retrieval.................11266
Tong Ye, Zhitao Li, Jianzong Wang, Ning Cheng, Jing Xiao

Immersive Enhancement and Removal of Loudspeaker Sound Using Wireless Assistive Listening Systems and Binaural Hearing Devices .........................................................11271
Ryan M. Corey, Andrew C. Singer

Cross-Subject Mental Fatigue Detection Based on Separable Spatio-Temporal Feature Aggregation ..........11276
Yalan Ye, Yutuo He, Wanqing Huang, Qiaosen Dong, Chong Wang, Guoqing Wang

Preserving Background Sound in Noise-Robust Voice Conversion Via Multi-Task Learning ..................11281
Jixun Yao, Yi Lei, Qing Wang, Pengcheng Guo, Ziqian Ning, Lei Xie, Hai Li, Junhui Liu, Danming Xie
Any-To-Any Voice Conversion with F0 and Timbre Disentanglement and Novel Timbre Conditioning
Sudheer Kovela, Rafael Valle, Ambrish Dantrey, Bryan Catanzaro
Enhancing Unsupervised Speech Recognition with Diffusion GANS
Xianchao Wu
Learning Sparse Alignments Via Optimal Transport for Cross-Domain Fake News Detection
Wei Tang, Zuyao Ma, Haijeng Sun, Jingyu Wang
Sampling Order-Limited Signals on the Sphere
Muhammad Salaar Arif Khan, Salman Nadeem, Zubair Khalid
Large Dimensional Analysis of LS-SVM Transfer Learning: Application to Polars Classification
Cyprien Doz, Chengfang Ren, Jean-Philippe Ovarlez, Romain Couillet
Robust Autoencoders for Collective Corruption Removal
Taihui Li, Hengkang Wang, Le Peng, Xian’E Tang, Ju Sun
Representation of Vocal Tract Length Transformation Based on Group Theory
Atsushi Miyashita, Tomoki Toda
Make Your Enemy Your Friend: Improving Image Rotation Angle Estimation with Harmonics
Kun Yu, Morteza Darvish Morshed Hosseini, Anjie Peng, Hui Zeng, Miroslav Goljan
Jeffreys Divergence-Based Regularization of Neural Network Output Distribution Applied to
Speaker Recognition
Pierre-Michel Bousquet, Mickael Rouvier
Single-Channel Speech Enhancement with Deep Complex U-Networks and Probabilistic Latent
Space Models
Eike J. Nustede, Jörn Anemüller
Spatial Inference Using Censored Multiple Testing with Fdr Control
Martin Götz, Abdelhak M. Zoubir, Visa Koivunen
YOLOX-B: A Better Yolox Model for Real-Time Driver Behavior Detection
Xu Guo, Ming Ma, Jiaqiang Zhang, Shaojie Li
Towards a More Stable and General Subgraph Information Bottleneck
Hongzhi Liu, Kaizhong Zheng, Shujian Yu, Badong Chen
Adaptive Scale and Spatial Aggregation for Real-Time Object Detection
Wei Chen, Yulin He, Zhengfa Liang, Yulan Guo
LE-DTA: Local Extrema Convolution for Drug Target Affinity Prediction
Tanoj Langore, Te-Cheng Hsu, Yi-Hsien Hsieh, Che Lin
An Empirical Study and Improvement for Speech Emotion Recognition
Zhen Wu, Yizhe Lu, Xinyu Dai
Phonetic RNN-Transducer for Mispronunciation Diagnosis
Daniel Yue Zhang, Soumya Saha, Sarah Campbell
High-Acoustic Fidelity Text to Speech Synthesis with Fine-Grained Control of Speech Attributes
Rafael Valle, João Felipe Santos, Kevin J. Shih, Rohan Badlani, Bryan Catanzaro
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Localization for Extremely Large-Scale Antenna Arrays with Spatial Non-Stationarity</td>
<td>11376</td>
</tr>
<tr>
<td>Xiaohuan Wu, Ji Sun, Xiaoyuan Jia, Shuxin Wang</td>
<td></td>
</tr>
<tr>
<td>Rain2Avoid: Self-Supervised Single Image Deraining</td>
<td>11381</td>
</tr>
<tr>
<td>Yan-Tsung Peng, Wei-Hua Li</td>
<td></td>
</tr>
<tr>
<td>Joint Human Orientation-Activity Recognition Using WIFI Signals for Human-Machine Interaction</td>
<td>11386</td>
</tr>
<tr>
<td>Hojjat Salehinejad, Navid Hasanzadeh, Radomir Djogo, Shahrokh Valae</td>
<td></td>
</tr>
<tr>
<td>AE-Flow: Autoencoder Normalizing Flow</td>
<td>11391</td>
</tr>
<tr>
<td>Jakub Mosinski, Piotr Bilinski, Thomas Merritt, Abdelhamid Ezzerg, Daniel Korzekwa</td>
<td></td>
</tr>
<tr>
<td>An Empirical Study on Speech Restoration Guided by Self-Supervised Speech Representation</td>
<td>11396</td>
</tr>
<tr>
<td>Jaeuk Byun, Youna Ji, Soo-Whan Chung, Soyeon Choe, Min-Seok Choi</td>
<td></td>
</tr>
<tr>
<td>Transferring Quantified Emotion Knowledge for the Detection of Depression in Alzheimer’s Disease Using Foreestnets</td>
<td>11401</td>
</tr>
<tr>
<td>Unsupervised Noise Adaptation Using Data Simulation</td>
<td>11406</td>
</tr>
<tr>
<td>Chen Chen, Yuchen Hu, Heqing Zou, Linhui Sun, Eng Siong Chng</td>
<td></td>
</tr>
<tr>
<td>A Privacy-Preserving Trajectory Mining Model</td>
<td>11411</td>
</tr>
<tr>
<td>Ziyang Wang, Sissi Xiaoxiao Wu, Junjie Zhu, Yingying Zhu</td>
<td></td>
</tr>
<tr>
<td>Element Selection with Wide Class of Optimization Criteria Using Non-Convex Sparse Optimization</td>
<td>11416</td>
</tr>
<tr>
<td>Taiga Kawamura, Natsuki Ueno, Nobutaka Ono</td>
<td></td>
</tr>
<tr>
<td>Efficient Implementation of Robust CUSUM Algorithm to Characterize Nanogaps Measurements with Heavy-Tailed Noise</td>
<td>11421</td>
</tr>
<tr>
<td>Javier Kipen, Joakim Jaldén, Shyamprasad N. Raja, Saumey Jain</td>
<td></td>
</tr>
<tr>
<td>PMMSD: Development of the Matrix Sentence Intelligibility Dataset for Mandarin with Lombard Effect</td>
<td>11426</td>
</tr>
<tr>
<td>Hanchen Pei, Yuhong Yang, Xufeng Chen, Qingnu Liu, Hongyang Chen, Weiping Tu, Song Lin</td>
<td></td>
</tr>
<tr>
<td>Robust Content-Variant Reference Image Quality Assessment Via Similar Patch Matching</td>
<td>11431</td>
</tr>
<tr>
<td>Wenbo Shi, Wenming Yang, Qingmin Liao</td>
<td></td>
</tr>
<tr>
<td>Neighborhood Information-Based Label Refinement for Person Re-Identification with Label Noise</td>
<td>11436</td>
</tr>
<tr>
<td>Xian Zhong, Shuapeng Su, Wenzuan Liu, Xuemei Jia, Wexin Huang, Mengdie Wang</td>
<td></td>
</tr>
<tr>
<td>A Game of Snakes and Gans</td>
<td>11441</td>
</tr>
<tr>
<td>Siddarath Asokan, Fatwir Sheikh Mohammed, Chandra Sekhar Seelamantula</td>
<td></td>
</tr>
<tr>
<td>Joint Millimeter-Wave AoD and AoA Estimation Using One OFDM Symbol and Frequency-Dependent Beams</td>
<td>11446</td>
</tr>
<tr>
<td>Veljko Boljanovic, Danijela Cabric</td>
<td></td>
</tr>
<tr>
<td>Quantized Precoding and RIS-Assisted Modulation for Integrated Sensing and Communications Systems</td>
<td>11451</td>
</tr>
<tr>
<td>R. S. Prasobh Sankar, Sundeep Prabhakar Chepuri</td>
<td></td>
</tr>
</tbody>
</table>
Deep Architecture for DOA Trajectory Localization .......................................................... 11456
Shreyas Jaiswal, Ruchi Pandey, Santosh Nannuru

Ancient Chinese Word Segmentation and Part-Of-Speech Tagging Using Distant Supervision .......... 11461
Shuo Feng, Piji Li

SIAST: A Slot Imbalance-Aware Self-Training Scheme for Semi-Supervised Slot Filling ............... 11466
Jiachi Liu, Sishi Xiong, Yuehuan He, Tong Zhou, Liwen Wang, Xuefeng Li, Bo Xiao

A Novel Transformer-Based Pipeline for Lung Cytopathological Whole Slide Image Classification .... 11471
Gaojie Li, Qing Liu, Haotian Liu, Yixiong Liang

End-To-End Non-Autoregressive Image Captioning ................................................................... 11476
Hong Yu, Yuanqiu Liu, Baokun Qi, Zhaolong Hu, Han Liu

Spatio-Temporal Hybrid Fusion of CAE and SWin Transformers for Lung Cancer Malignancy Prediction ......................................................................................................................... 11481
Sadaf Khademi, Shahin Heidarian, Parnian Afshar, Farnoosh Naderkhani, Anastasia Oikonomou, Konstantinos N. Plataniotis, Arash Mohammadi

Moving Towards Non-Binary Gender Identification Via Analysis of System Errors in Binary Gender Classification .............................................................................................................................. 11486
Sebastian Ellis, Stefan Goetze, Heidi Christensen

Improving Few-Shot Learning for Talking Face System with TTS Data Augmentation .................... 11491
Qi Chen, Ziyang Ma, Tao Liu, Xu Tan, Qu Lu, Kai Yu, Xie Chen

On the Fairness of Multitask Representation Learning .................................................................... 11496
Yingcong Li, Samet Oymak

Animal Re-Identification Algorithm for Posture Diversity ............................................................. 11501
Zhimin He, Jiangbo Qian, Diqun Yan, Chong Wang, Yu Xin

Diffusion Probabilistic Modeling for Fine-Grained Urban Traffic Flow Inference with Relaxed Structural Constraint .................................................................................................................. 11506
Xovee Xu, Yutao Wei, Pengyu Wang, Xucheng Luo, Fan Zhou, Goce Trajcevski

PCQA-Graphpoint: Efficient Deep-Based Graph Metric for Point Cloud Quality Assessment ........ 11511
Marouane Tliba, Aladine Chetouani, Giuseppe Valenzise, Frédéric Dufaux

Low-Rank Plus Sparse Trajectory Decomposition for Direct Exoplanet Imaging ............................ 11516
Simon Vary, Hazan Daglayan, Laurent Jacques, P.-A. Absil

Classification of Synthetic Facial Attributes by Means of Hybrid Classification/Localization Patch-Based Analysis ............................................................................................................................ 11521
Jun Wang, Benedetta Tondi, Mauro Barni

Leveraging Pretrained Representations with Task-Related Keywords for Alzheimer’s Disease Detection ................................................................................................................................. 11526
Jinchao Li, Kaitao Song, Junan Li, Bo Zheng, Dongsheng Li, Xixin Wu, Xunying Liu, Helen Meng

Towards Learning Emotion Information from Short Segments of Speech ...................................... 11531
Tilak Purohit, Sarthak Yadav, Bogdan Vlasenko, S. Pavankumar Dubagunta, Mathew Magimai.-Doss
Unsupervised Word Segmentation Based on Word Influence .......................................................... 11536
Ruohao Yan, Huaping Zhang, Wushour Silamu, Askar Hamdulla

Classification-Based Dynamic Network for Efficient Super-Resolution ...................................... 11541
Qi Wang, Weiwei Fang, Meng Wang, Yusong Cheng

Vision Transformer with Progressive Tokenization for CT Metal Artifact Reduction ....................... 11546
Songwei Zheng, Dong Zhang, Chunyan Yu, Danhong Zhu, Longlong Zhu, Hao Liu, Zhongzheng Huang

The DKU Post-Challenge Audio-Visual Wake Word Spotting System for the 2021 MISP Challenge:
Deep Analysis ................................................................................................................................. 11551
Haoxu Wang, Ming Cheng, Qiang Fu, Ming Li

Absolute Decision Corrupts Absolutely: Conservative Online Speaker Diarisation ............................. 11556
Youngki Kwon, Hee-Soo Heo, Bong-Jin Lee, You Jin Kim, Jee-Weon Jung

Incorporating Uncertainty from Speaker Embedding Estimation to Speaker Verification .................... 11561
Qiongqiong Wang, Kong Aik Lee, Tianchi Li

DocRED-FE: A Document-Level Fine-Grained Entity and Relation Extraction Dataset ....................... 11566
Hongbo Wang, Weimin Xiong, Yifan Song, Dawei Zhu, Yu Xia, Sujian Li

Heart Rate Extraction from Abdominal Audio Signals .................................................................... 11571
Jake Stuchbury-Wass, Erika Bondareva, Kayla-Jade Butkow, Sanja Šcepanovic, Zoran Radivojevic, Cecilia Mascolo

PUFFIN: Pitch-Synchronous Neural Waveform Generation for Fullband Speech on Modest Devices ................................................................................................................................. 11576
Oliver Watts, Lovisa Wihlborg, Cassia Valentini-Botinhao

Deep Plug-And-Play for Tensor Robust Principal Component Analysis .......................................... 11581
Hao Tan, Jianjun Wang, Weichao Kong

ST-MVDNet++: Improve Vehicle Detection with Lidar-Radar Geometrical Augmentation Via Self-
Training ........................................................................................................................................ 11586
Yu-Jhe Li, Matthew O’Toole, Kris Kitani

Two-Stream Joint-Training for Speaker Independent Acoustic-to-Articulatory Inversion .................... 11591
Jianrong Wang, Jinyu Liu, Xuewei Li, Mei Yu, Jie Gao, Qiang Fang, Li Liu

Lattice-Free Sequence Discriminative Training for Phoneme-Based Neural Transducers .................... 11596
Zijian Yang, Wei Zhou, Ralf Schlüter, Hermann Ney

Optimization of Sensor Configurations for Fault Identification in Smart Buildings ............................ 11601
Naveed Ahmad, Malcolm Egan, Jean-Marie Gorce, Jilles S. Dibangoye, Frédéric Le Mouël

Prototype-Based Layered Federated Cross-Modal Hashing ................................................................ 11606
Jiale Liu, Yu-Wei Zhan, Xin Luo, Zhen-Duo Chen, Yongxin Wang, Xin-Shan Xu

Hankel Structured Low Rank and Sparse Representation Via L0-Norm Optimization for
Compressed Ultrasound Plane Wave Signal Reconstruction .................................................................. 11611
Miaomiao Zhang, Ji Chen, Xiaoyan Fu, Ge Xin, Jingzhi Zhang, Na Jiang, Jan D’Hooge

Variational Bayesian Channel Estimation in Wideband Multi-Scale Multi-Lag Channels ................. 11616
Niladri Halder, Arunkumar K. P., Chandra R. Murthy
Self-Healing Through Error Detection, Attribution, and Retraining .............................11621
  Ansel Maclaughlin, Anna Rumshisky, Rinat Khaziev, Anil Ramakrishna, Yuval Merhav, Rahul Gupta

Multi-Aspect Interest Neighbor-Augmented Network for Next-Basket Recommendation........11626
  Zhiying Deng, Jianjun Li, Zhiqiang Guo, Guohui Li

Wireless Deep Speech Semantic Transmission ..........................................................11631
  Zixuan Xiao, Shengshi Yao, Jincheng Dai, Sixian Wang, Kai Niu, Ping Zhang

Metric-Oriented Speech Enhancement Using Diffusion Probabilistic Model ........................11636
  Chen Chen, Yuchen Hu, Weiwei Weng, Eng Siong Chng

Decomposition, Interaction, Reconstruction Meets Global Context Learning in Visual Tracking........11641
  Huibin Tan, Kun Hu, Mingyu Cao, Mengzhu Wang, Liyang Xu, Wenjing Yang

Question Answering System with Sparse and Noisy Feedback ....................................11646
  Djallel Bouneffouf, Oznur Alkan, Raphael Feraud, Bathian Lin

Single-Sample Direction-of-Arrival Estimation for Fast and Robust 3D Localization with Real Measurements from a Massive MIMO System ..................................................11651
  Stepan Mazokha, Sanaz Naderi, Georgios I. Orfanidis, George Sklivanitis, Dimitris A. Pados,
  Jason O. Hallstrom

Factorized AED: Factorized Attention-Based Encoder-Decoder for Text-Only Domain Adaptive ASR .................................................................11656
  Xun Gong, Wei Wang, Hang Shao, Xie Chen, Yanmin Qian

Positive-Pair Redundancy Reduction Regularisation for Speech-Based Asthma Diagnosis Prediction ..........................................................11661
  Georgios Rizos, Rafael A. Calvo, Björn W. Schuller

Incorporating Lip Features into Audio-Visual Multi-Speaker DOA Estimation by Gated Fusion ........11666
  Ya Jiang, Hang Chen, Jun Du, Qing Wang, Chin-Hui Lee

A Large-Scale Pretrained Deep Model for Phishing URL Detection ................................11671
  Yanbin Wang, Weifan Zhu, Haitao Xu, Zhan Qin, Kui Ren, Wenrui Ma

Code-Switching Speech Synthesis Based on Self-Supervised Learning and Domain Adaptive Speaker Encoder ..................................................11676
  Yi-Xing Lin, Cheng-Hsun Pai, Phuoc Thi Le, Bima Prihasto, Chien-Ling Huang, Jia Ching
  Wang

Geometric Matrix Completion with Collaborative Routing Between Capsules .......................11681
  Xuan Li, Li Zhang

Distinguishable Speaker Anonymization Based on Formant and Fundamental Frequency Scaling ....11686
  Jixun Yao, Qing Wang, Yi Lei, Pengcheng Guo, Lei Xie, Namin Wang, Jie Liu

One-Shot Action Detection Via Attention Zooming in ..................................................11691
  He-Yen Hsieh, Ding-Jie Chen, Cheng-Wei Chang, Tyng-Luh Liu

Toward Universal Text-To-Music Retrieval .....................................................................11696
  Seungheon Doh, Minz Won, Keunwoo Choi, Juhan Nam

A Hierarchical Regression Chain Framework for Affective Vocal Burst Recognition .............11701
  Jinchao Li, Xixin Wu, Kaitao Song, Dongsheng Li, Xunying Liu, Helen Meng
Lightweight Annotation and Class Weight Training for Automatic Estimation of Alarm Audibility in Noise ................................................................. 11706
François Effa, Romain Sertzel, Jean-Pierre Arz, Nicolas Grimault

Inv-Senet: Invariant Self Expression Network for Clustering Under Biased Data ................................................................. 11711
Ashutosh Singh, Ashish Singh, Aria Masoomi, Tales Imbiriba, Erik Learned-Miller, Deniz Erdogmus

A Processing Framework to Access Large Quantities of Whispered Speech Found in ASMR .............................................. 11716
Pablo Pérez Zarazaga, Gustav Eje Henter, Zofia Malisz

Textless Speech-To-Music Retrieval Using Emotion Similarity .................................................................................. 11721
Seungheon Doh, Minz Won, Keunwoo Choi, Juhan Nam

Sequential Invariant Information Bottleneck .......................................................................................... 11726
Yichen Zhang, Shujian Yu, Badong Chen

Adaptive Semantic Fusion Framework for Unsupervised Monocular Depth Estimation ........................................ 11731
Ruoqi Li, Huimin Yu, Kaiyang Du, Zhuolong Xiao, Bo Yan, Zhengxi Yuan

FFEDCL: Fair Federated Learning with Contrastive Learning ............................................................................. 11736
Xiaoqiong Shi, Liping Yi, Xiaoguang Liu, Gang Wang

WITT: A Wireless Image Transmission Transformer for Semantic Communications ......................................... 11741
Ke Yang, Sixian Wang, Jincheng Dai, Kailin Tan, Kai Niu, Ping Zhang

Delay-Penalized Transducer for Low-Latency Streaming ASR ........................................................................... 11746
Wei Kang, Zengwei Yao, Fangjun Kuang, Liyong Guo, Xiaoyu Yang, Long Lin, Piotr Zelasko, Daniel Povey

SIGVIC: Spatial Importance Guided Variable-Rate Image Compression .......................................................... 11751
Jiaming Liang, Meiqin Liu, Chao Yao, Chunyu Lin, Yao Zhao

A Simple Scheme for Coupled Factorization for Hyperspectral Super-Resolution: Exploiting Sparsity in an Easy Way ...................................................................................... 11756
Yuening Li, Wing-Kin Ma, Ruiyuan Wu, Huikang Liu

The Multimodal Information Based Speech Processing (Misp) 2022 Challenge: Audio-Visual Diarization and Recognition ........................................................................................................ 11761
Zhe Wang, Shilong Wu, Hang Chen, Mao-Kui He, Jun Du, Chin-Hui Lee, Jingdong Chen, Shinji Watanabe, Sabato Siniscalchi, Odette Scharenborg, Diyuan Liu, Baocai Yin, Jia Pan, Jianqing Gao, Cong Liu

A Model-Based Hearing Compensation Method Using a Self-Supervised Framework .............................................. 11766
Yadong Niu, Nan Li, Xihong Wu, Jing Chen

Multitrack Music Transcription with a Time-Frequency Perceiver ............................................................................ 11771
Wei-Tsung Lu, Ju-Chiang Wang, Yun-Ning Hung

Singing Voice Synthesis Based on a Musical Note Position-Aware Attention Mechanism ........................................ 11776
Yukiya Hono, Kei Hashimoto, Yoshitomo Nankaku, Keiichi Tokuda

Relevance Propagation Through Deep Conditional Random Fields .............................................................................. 11781
Xiangyu Yang, Boris Joukovsky, Nikos Deligiannis

Boosting Transferability of Adversarial Example Via an Enhanced Euler’s Method ................................................ 11786
Anjie Peng, Zhi Lin, Hui Zeng, Wenxin Yu, Xiangui Kang
ScaleMix: Intra- And Inter-Layer Multiscale Feature Combination for Change Detection.........................................................11791
  Rui Huang, Qingyi Zhao, Ruofei Wang, Caihua Liu, Sihua Gao, Yuxiang Zhang, Wei Fan

Synthesizing Speech from ECoG with a Combination of Transformer-Based Encoder and Neural Vocoder.................................................................11796
  Kai Shigemi, Shuji Komeiji, Takumi Mitsushashi, Yasushi Hmura, Hiroharu Suzuki, Hidenori Gutano, Koichi Shinoda, Kohei Yatabe, Toshihisa Tanaka

AV-TAD: Audio-Visual Temporal Action Detection with Transformer.................................................................11801
  Yangcheng Li, Zefang Yu, Suncheng Xiang, Ting Liu, Yuzhuo Fu

Semantic-Preserving Augmentation for Robust Image-Text Retrieval..................................................................................11806
  Sunwoo Kim, Kyuhong Shim, Luong Trung Nguyen, Byonghyo Shim

Enhancing the Vocal Range of Single-Speaker Singing Voice Synthesis with Melody-Unsupervised Pre-Training.................................................................11811
  Shaohuan Zhou, Xu Li, Zhiyong Wu, Ying Shan, Helen Meng

Embedding a Differentiable Mel-Cepstral Synthesis Filter to a Neural Speech Synthesis System ..............................11816
  Takenori Yoshimura, Shinji Takaki, Kazuhiro Nakamura, Keiichiro Oura, Yukiya Hono, Kei Hashimoto, Yoshihiko Nankaku, Keiichi Tokuda

Boosting Semi-Supervised Federated Learning with Model Personalization and Client-Variance-Reduction ....................11821
  Shuai Wang, Yanqing Xu, Yani Yuan, Xiuhua Wang, Tony Q. S. Quek

SemGeo: Semantic Keywords for Cross-View Image Geo-Localization ..................................................................................11826
  Royston Rodrigues, Masahiro Tani

Modelling Black-Box Audio Effects with Time-Varying Feature Modulation.................................................................11831
  Marco Comunità, Christian J. Steinmetz, Huy Phan, Joshua D. Reiss

Symbol-Level Precoding is Related to Parameter Estimation from Quantized Data ..........................................................11836
  Mingjie Shao, Wing-Kin Ma, Yatao Liu

Runtime Prediction of Machine Learning Algorithms in Automl Systems .................................................................11841
  Parijat Dube, Theodoros Salonidis, Parikshit Ram, Ashish Verma

Pondering About Task Spatial Misalignment: Classification-Localization Equilibrated Object Detection ......................11846
  Yudong Zhang, Wei Lu, Xu Wang, Pengkun Wang, Yang Wang

Efficient Speech Quality Assessment Using Self-Supervised Framewise Embeddings ................................................11851
  Karl El Hajal, Zihan Wu, Neil Scheidwasser-Clow, Gasser Elbanna, Milos Cernak

History, Present and Future: Enhancing Dialogue Generation with Few-Shot History-Future Prompt ......................11856
  Yihe Wang, Yitong Li, Yasheng Wang, Fei Mi, Pingyi Zhou, Jin Liu, Xin Jiang, Qun Liu

Audio Quality Assessment of Vinyl Music Collections Using Self-Supervised Learning ...........................................11861
  Alessandro Ragano, Emmanouil Benetos, Andrew Hines

Massively Multilingual ASR on 70 Languages: Tokenization, Architecture, and Generalization Capabilities ................11866
  Andros Tjandra, Nayan Singhal, David Zhang, Ozlem Kalinli, Abdelrahman Mohamed, Duc Le, Michael L. Seltzer
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free-View Expressive Talking Head Video Editing</td>
<td>11871</td>
</tr>
<tr>
<td>Yuantian Huang, Satoshi Izuka, Kazuhiro Fukui</td>
<td></td>
</tr>
<tr>
<td>Wav2vec-Based Detection and Severity Level Classification of Dysarthria from Speech</td>
<td>11876</td>
</tr>
<tr>
<td>Farhad Javanmardi, Sasha Tirronen, Manila Kodali, Sudarsana Reddy Kadiri, Paavo Alku</td>
<td></td>
</tr>
<tr>
<td>Graph Based Semantic Ensemble of Riemannian Neural Structured Learning for BCI-EEG Signal Classification</td>
<td>11881</td>
</tr>
<tr>
<td>Vinay Gupta, Laxmidhar Behera, Tushar Sandhan</td>
<td></td>
</tr>
<tr>
<td>Measuring the Transferability of $\ell^\infty$ Attacks by the $\ell^2$ Norm</td>
<td>11886</td>
</tr>
<tr>
<td>Sizhe Chen, Qinghua Tao, Zhixing Ye, XiaoLIN Huang</td>
<td></td>
</tr>
<tr>
<td>An Efficient Relay Selection Scheme for Relay-Assisted HARQ</td>
<td>11891</td>
</tr>
<tr>
<td>Weihang Ding, Mohammad Shikh-Bahaei</td>
<td></td>
</tr>
<tr>
<td>Towards Privacy and Utility in Tourette TIC Detection Through Pretraining Based on Publicly Available Video Data of Healthy Subjects</td>
<td>11896</td>
</tr>
<tr>
<td>Enhanced Low-Resolution LiDAR-Camera Calibration Via Depth Interpolation and Supervised Contrastive Learning</td>
<td>11901</td>
</tr>
<tr>
<td>Zhikang Zhang, Zifan Yu, Suya You, Raghveer Rao, Sanjeev Agarwal, Fengbo Ren</td>
<td></td>
</tr>
<tr>
<td>Nonnegative Block-Term Decomposition with the $\beta$-Divergence: Joint Data Fusion and Blind Spectral Unmixing</td>
<td>11906</td>
</tr>
<tr>
<td>C. Prévost, V. Leplat</td>
<td></td>
</tr>
<tr>
<td>High-Level Feature Fusion Network for Session-Based Social Recommendation</td>
<td>11911</td>
</tr>
<tr>
<td>Liuyin Wang, Mingchao Li, Hai-Tao Zheng</td>
<td></td>
</tr>
<tr>
<td>Exploiting CCTV Cameras for Hand Hygiene Recognition in ICU</td>
<td>11916</td>
</tr>
<tr>
<td>Weijun Huang, Jia Huang, Guowei Wang, Hongzhou Lu, Min He, Wenjin Wang</td>
<td></td>
</tr>
<tr>
<td>A Database for Multi-Modal Short Video Quality Assessment</td>
<td>11921</td>
</tr>
<tr>
<td>Yukun Zhang, Chuan Wang, Sanyi Zhang, Xiaochun Cao</td>
<td></td>
</tr>
<tr>
<td>Sequence-Based Device-Free Gesture Recognition Framework for Multi-Channel Acoustic Signals</td>
<td>11926</td>
</tr>
<tr>
<td>Zhizheng Yang, Xun Wang, Dongyu Xia, Wei Wang, Haipeng Dai</td>
<td></td>
</tr>
<tr>
<td>Detecting Out-Of-Distribution Examples Via Class-Conditional Impressions Reappearing</td>
<td>11931</td>
</tr>
<tr>
<td>Jinggang Chen, Xiaoyang Qu, Junjie Li, Jianzong Wang, Jiguang Wan, Jing Xiao</td>
<td></td>
</tr>
<tr>
<td>Rethinking Rule-Based Approaches in Session-Based Recommendation</td>
<td>11936</td>
</tr>
<tr>
<td>Liuyin Wang, Mingchao Li, Hai-Tao Zheng</td>
<td></td>
</tr>
<tr>
<td>MLCGAN: Multi-Lead ECG Synthesis with Multi Label Conditional Generative Adversarial Network</td>
<td>11941</td>
</tr>
<tr>
<td>Jian Wu, Liping Wang, Hailin Pan, Binyu Wang</td>
<td></td>
</tr>
<tr>
<td>Multilingual Word Error Rate Estimation: E-Wer3</td>
<td>11946</td>
</tr>
<tr>
<td>Shammur Absar Chowdhury, Ahmed Ali</td>
<td></td>
</tr>
<tr>
<td>Scalable and Secure Federated XGBoost</td>
<td>11951</td>
</tr>
<tr>
<td>Quang Minh Nguyen, Nhan Khanh Le, Lam M. Nguyen</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>A DNN Based Normalized Time-Frequency Weighted Criterion for Robust Wideband DoA</td>
<td>11956</td>
</tr>
<tr>
<td>Kuan-Lin Chen, Ching-Hua Lee, Bhaskar D. Rao, Harinath Garudadri</td>
<td></td>
</tr>
<tr>
<td>DPP-Based Client Selection for Federated Learning with NON-IID DATA</td>
<td>11961</td>
</tr>
<tr>
<td>Yuxuan Zhang, Chao Xu, Howard H. Yang, Xijun Wang, Tony Q. S. Quek</td>
<td></td>
</tr>
<tr>
<td>Robust Acoustic and Semantic Contextual Biasing in Neural Transducers for Speech Recognition</td>
<td>11966</td>
</tr>
<tr>
<td>Xuandi Fu, Kanthashree Mysore Sathyendra, Ankur Gandhe, Jing Liu, Grant P. Strimel, Ross</td>
<td></td>
</tr>
<tr>
<td>McGowan, Athanasios Mouchtaris</td>
<td></td>
</tr>
<tr>
<td>Memory-Augmented Contrastive Learning for Talking Head Generation</td>
<td>11971</td>
</tr>
<tr>
<td>Jianrong Wang, Yaxin Zhao, Hongkai Fan, Tianyi Xu, Qi Li, Sen Li, Li Liu</td>
<td></td>
</tr>
<tr>
<td>Extreme Audio Time Stretching Using Neural Synthesis</td>
<td>11976</td>
</tr>
<tr>
<td>Leonardo Fierro, Alec Wright, Vesa Välimäki, Matti Hämäläinen</td>
<td></td>
</tr>
<tr>
<td>Group Personalized Federated Learning</td>
<td>11981</td>
</tr>
<tr>
<td>Zhe Liu, Yue Hui, Fuchun Peng</td>
<td></td>
</tr>
<tr>
<td>Guide and Select: A Transformer-Based Multimodal Fusion Method for Points of Interest</td>
<td>11986</td>
</tr>
<tr>
<td>Description Generation</td>
<td></td>
</tr>
<tr>
<td>Hanqing Liu, Wei Wang, Niu Hu, Hai-Tao Zheng, Rui Xie, Wei Wu, Yang Bai</td>
<td></td>
</tr>
<tr>
<td>Adaptive CSI Feedback with Hidden Semantic Information Transfer</td>
<td>11991</td>
</tr>
<tr>
<td>Jiaqi Cao, Lixiang Lian, Yijie Mao, Bruno Clerckx</td>
<td></td>
</tr>
<tr>
<td>Deep Probabilistic Model for Lossless Scalable Point Cloud Attribute Compression</td>
<td>11996</td>
</tr>
<tr>
<td>Dat Thanh Nguyen, Kamal Gopikrishnan Nambiar, André Kaup</td>
<td></td>
</tr>
<tr>
<td>Improved Wordpcfg for Passwords with Maximum Probability Segmentation</td>
<td>12001</td>
</tr>
<tr>
<td>Wenting Li, Jiahong Yang, Haibo Cheng, Ping Wang, Kaitai Liang</td>
<td></td>
</tr>
<tr>
<td>The Uniqueness Problem of Physical Law Learning</td>
<td>12006</td>
</tr>
<tr>
<td>Philipp Scholl, Aras Bacho, Holger Boche, Gitta Kutyniok</td>
<td></td>
</tr>
<tr>
<td>Visual Information Matters for ASR Error Correction</td>
<td>12011</td>
</tr>
<tr>
<td>Vanya Bannihatti Kumar, Shanbo Cheng, Ningxin Peng, Yuchen Zhang</td>
<td></td>
</tr>
<tr>
<td>SyncNet: Correlating Objective for Time Delay Estimation in Audio Signals</td>
<td>12016</td>
</tr>
<tr>
<td>Akshay Raina, Vipul Arora</td>
<td></td>
</tr>
<tr>
<td>Multiple Target Measurements: Bayesian Framework for Moving Object Detection in Mimo Radar</td>
<td>12021</td>
</tr>
<tr>
<td>Bastian Eisele, Ali Bereyhi, Ralf Müller</td>
<td></td>
</tr>
<tr>
<td>FindAdaptNet: Find and Insert Adapters by Learned Layer Importance</td>
<td>12026</td>
</tr>
<tr>
<td>Junwei Huang, Karthik Ganesan, Soumi Maiti, Young Min Kim, Xuankai Chang, Paul Liang,</td>
<td></td>
</tr>
<tr>
<td>Shinji Watanabe</td>
<td></td>
</tr>
<tr>
<td>Continuous Action Space-Based Spoken Language Acquisition Agent Using Residual Sentence</td>
<td>12031</td>
</tr>
<tr>
<td>Embedding and Transformer Decoder</td>
<td></td>
</tr>
<tr>
<td>Ryota Komatsu, Yusuke Kimura, Takuma Okamoto, Takahiro Shinozaki</td>
<td></td>
</tr>
<tr>
<td>On the Detection of Synthetic Images Generated by Diffusion Models</td>
<td>12036</td>
</tr>
<tr>
<td>Riccardo Corvi, Davide Cozzolino, Giada Zingarini, Giovanni Poggi, Koki Nagano, Luisa</td>
<td></td>
</tr>
<tr>
<td>Verdoliva</td>
<td></td>
</tr>
</tbody>
</table>
Automatic Classification of Vocal Intensity Category from Speech ................................................. 12041
Manila Kodali, Sudarsana Reddy Kadiri, Laura Laaksonen, Paavo Alku

A Sentiment and Syntactic-Aware Graph Convolutional Network for Aspect-Level Sentiment
Classification ............................................................................................................................ 12046
Yuxin Yang, Xia Sun, Qiang Lu, Richard Sutcliffe, Jun Feng

Effectiveness of Text, Acoustic, and Lattice-Based Representations in Spoken Language
Understanding Tasks ............................................................................................................... 12051
Esaú Villatoro-Tello, Srikanth Madikeri, Juan Zuluaga-Gomez, Bidisha Sharma, Seyyed
Saeed Sarfjoo, Iuliia Nigmatulina, Petr Motlicek, Alexei V. Ivanov, Aravind Ganapathiraju

Retrieval-Based Natural 3D Human Motion Generation ............................................................. N/A
Zehan Tan, Weidong Yang, Shuai Wu

Enhancing the Efficiency of WMMSE and FP for Beamforming by Minorization-Maximization ....... 12061
Zepeng Zhang, Ziping Zhao, Kaiming Shen

Efficient Learning of Balanced Signature Graphs ..................................................................... 12066
Gerald Matz, Claudio Verardo, Thomas Dittrich

Abusive Activity Detection with Multi-Modality Based on Convolutional Neural Network .......... 12071
Jisoo Kim, Hyebin Ahn, Byounghyun Yoo

Lyapunov-Driven Deep Reinforcement Learning for Edge Inference Empowered by
Reconfigurable Intelligent Surfaces ........................................................................................ 12076
Kyriakos Stylianopoulos, Mattia Merluzzi, Paolo Di Lorenzo, George C. Alexandropoulos

A Proximal Approach to IVA-G with Convergence Guarantees ............................................... 12081
Clément Cosserat, Ben Gabrielson, Emilie Chouzenoux, Jean-Christophe Pesquet, Tuilay
Adali

Low-Dose CT Reconstruction Via Optimization-Inspired GAN .................................................. 12086
Jiawei Jiang, Yuchao Feng, Honghui Xu, Jianwei Zheng

SPASHT: Semantic and Pragmatic Speech Features for Automatic Assessment of Autism .......... 12091
Ashwini B, Vrinda Narayan, Jainendra Shukla

Effectiveness of Inter- And Intra-Subarray Spatial Features for Acoustic Scene Classification .... 12096
Takao Kawamura, Yuma Kinoshita, Nobutaka Ono, Robin Scheibler

Estimating Shapley Values of Training Utterances for Automatic Speech Recognition Models .... 12101
Ali Raza Syed, Michael I. Mandel

Practice of the Conformer Enhanced Audio-Visual Hubert on Mandarin and English ................. 12106
Xiaoming Ren, Chao Li, Shenjian Wang, Biao Li

Piecewise Position Encoding in Convolutional Neural Network for Cough-Based Covid-19
Detection ................................................................................................................................. 12111
Jiakun Shen, Xueshuai Zhang, Pengyuan Zhang, Yonghong Yan, Shaoxing Zhang, Zhihua
Huang, Yanfen Tang, Yu Wang, Fujie Zhang, Aijun Sun

Transplayer: Timbre Style Transfer with Flexible Timbre Control ............................................. 12116
Yuxuan Wu, Yifan He, Xinlu Liu, Yi Wang, Roger B. Dannenberg
Robust M-Estimation Based Distributed Expectation Maximization Algorithm with Robust Aggregation......................................................................................................................... 12121
  Christian A. Schroth, Stefan Vlaski, Abdelhak M. Zoubir

Improving Transformer-Based Networks with Locality for Automatic Speaker Verification .................. 12126
  Mufan Sang, Yong Zhao, Gang Liu, John H. L. Hansen, Jian Wu

Causal Discovery and Causal Inference Based Counterfactual Fairness in Machine Learning .............. 12131
  Yajing Wang, Zongwei Luo

Learning Silhouettes with Group Sparse Autoencoders ........................................................................... 12136
  Emmanouil Theodosis, Demba Ba

Progressive Refinement Learning Based on Feature Cross Perception for Residential Areas
Semantic Segmentation ................................................................................................................................. 12141
  Xinran Lyu, Libao Zhang

HAG: Hierarchical Attention with Graph Network for Dialogue Act Classification in Conversation...... 12146
  Changzeng Fu, Zhenghan Chen, Jiaqi Shi, Bowen Wu, Chaoran Liu, Carlos Toshinori Ishi, Hiroshi Ishiguro

Neural Fourier Shift for Binaural Speech Rendering ................................................................................. 12151
  Jin Woo Lee, Kyoga Lee

CAN2V: Can-Bus Data-Based Seq2seq Model for Vehicle Velocity Prediction........................................ 12156
  Jae-Heung Cho, Joon-Hyuk Chang

Relative Dynamic Time Warping Comparison for Pronunciation Errors .................................................... 12161
  Caitlin Richter, Jón Guðnason

Unsupervised Out-Of-Distribution Detection Using Few in-Distribution Samples ................................. 12166
  Chandan Gautam, Aditya Kane, Savitha Ramasamy, Suresh Sundaram

G2CNN: Geometric Prior Based GCNN for Single-View 3D Reconstruction with Loop Subdivision.................. 12171
  Kun Cao, Na Qi, Wei Xu, Qing Zhu, Shibo Xu, Changxin Pan

Towards Building Text-To-Speech Systems for the Next Billion Users..................................................... 12176
  Gokul Karthik Kumar, Praveen S V, Pratyush Kumar, Mitesh M. Khapra, Karthik Nandakumar

Data Leakage in Cross-Modal Retrieval Training: A Case Study............................................................... 12181
  Benno Weck, Xavier Serra

Self Supervised Bert for Legal Text Classification.......................................................................................... 12186
  Arghya Pal, Sailaja Rajanala, Raphaël C.-W. Phan, Koksheik Wong

Egocentric Action Anticipation for Personal Health..................................................................................... 12191
  Ivan Rodin, Antonino Furnari, Dimitrios Mavroeidis, Giovanni Maria Farinella

Global Localisation in Continuous Magnetic Vector Fields Using Gaussian Processes........................... 12196
  William McDonald, Cedric Le Gentil, Teresa Vidal-Calleja

I3D: Transformer Architectures with Input-Dependent Dynamic Depth for Speech Recognition.............. 12201
  Yifan Peng, Jaesong Lee, Shinji Watanabe
SSGD: A Smartphone Screen Glass Dataset for Defect Detection ......................................................... 12206
   Haonan Han, Rui Yang, Shuyan Li, Runze Hu, Xiu Li

Joint Cryo-ET Alignment and Reconstruction with Neural Deformation Fields ........................................ 12211
   Valentin Debarnot, Sidharth Gupta, Konik Kothari, Ivan Dokmanic

Finer-Grained Decomposition for Parallel Quantum Mimo Processing ...................................................... 12216
   Minsung Kim, Kyle Jamieson

In-Sensor & Neuromorphic Computing Are All You Need for Energy Efficient Computer Vision .............. 12221
   Gourav Datta, Zeyu Liu, Md Abdullah-Al Kaiser, Souvik Kundu, Joe Mathai, Zihan Yin, Ajey P. Jacob, Akhilesh R. Jaiswal, Peter A. Beerel

Aiding Speech Harmonic Recovery in DNN-Based Single Channel Noise Reduction Using Cepstral Excitation Manipulation (CEM) Components ................................................................. 12226
   Yanjue Song, Nilesh Madhu

Noise-Aware Target Extension with Self-Distillation for Robust Speech Recognition ............................ 12231
   Ju-Seok Seong, Jeong-Hwan Choi, Jehyun Kyung, Ye-Rin Jeoung, Joon-Hyuk Chang

A Quantum Approach for Stochastic Constrained Binary Optimization .................................................. 12236
   Sarthak Gupta, Vassilis Kekatos

SVMV: Spatiotemporal Variance-Supervised Motion Volume for Video Frame Interpolation .................. 12241
   Yao Luo, Jinshan Pan, Jinhui Tang

LED: Label Correlation Enhanced Decoder for Multi-Label Text Classification ...................................... 12246
   Kefan Ma, Zheng Huang, Xinrui Deng, Jie Guo, Weidong Qiu

Unlimited Sampling of FRI Signals Independent of Sampling Rate ....................................................... 12251
   Ruiming Guo, Ayush Bhandari

New Interpretable Patterns and Discriminative Features from Brain Functional Network Connectivity Using Dictionary Learning ........................................................................................................ 12256
   F. Ghayem, H. Yang, F. Kantar, S.-J. Kim, V. D. Calhoun, T. Adali

MRML: Multimodal Rumor Detection by Deep Metric Learning .............................................................. 12261
   Liwen Peng, Songlei Jian, Dongsheng Li, Siqi Shen

DiffVoice: Text-To-Speech with Latent Diffusion ....................................................................................... 12266
   Zhijun Liu, Yiwei Guo, Kai Yu

Improved Appliance Transient Feature Extraction Via Template Matching ............................................. 12271
   Bo Liu, Fenglei Chang, Wenpeng Luan, Bochao Zhao

Asymptotic Distribution of Stochastic Mirror Descent Iterates in Average Ensemble Models ................. 12276
   Taylan Kargin, Fariborz Salehi, Babak Hassibi

DSPGAN: A Gan-Based Universal Vocoder for High-Fidelity TTS by Time-Frequency Domain Supervision from DSP ..................................................................................................................... 12281
   Kun Song, Yongmao Zhang, Yi Lei, Jian Cong, Hanzhao Li, Lei Xie, Gang He, Jinfeng Bai

Efficient Feature Fusion for Learning-Based Photometric Stereo .......................................................... 12286
   Yakun Ju, Kin-Man Lam, Jun Xiao, Cong Zhang, Cuixin Yang, Junyue Dong

Good Neighbors Are All You Need for Chinese Grapheme-To-Phoneme Conversion ................................ 12291
   Jungjun Kim, Changjin Han, Gyuhyeon Nam, Gyeongsu Chae
SPECTRANET-SO(3): Learning Satellite Orientation from Optical Spectra by Implicitly Modeling Mutually Exclusive Probability Distributions on the Rotation Manifold
Matthew Phelps, Thomas Swindle, J. Zachary Gazak, Andrew Vandenberg, Justin Fletcher

A Controllable Lifestyle Simulator for Use in Deep Reinforcement Learning Algorithms
Libio Gonçalves Braz, Allmin Susaiyah

Learning to Generate 3D Representations of Building Roofs Using Single-View Aerial Imagery
Maxim Khomiakov, Alejandro Valverde Mahou, Alba Reinders Sánchez, Jes Frellsen, Michael Riis Andersen

Building Keyword Search System from End-To-End Asr Systems
Ruizhe Huang, Matthew Wiesner, Leibny Paola Garcia-Perera, Dan Povey, Jan Trmal, Sanjeev Khudanpur

LMBAO: A Landmark Map for Bundle Adjustment Odometry in LiDAR SLAM
Letian Zhang, Jinping Wang, Lu Jie, Nanjie Chen, Xiaojun Tan, Zhifei Duan

Multi-Observation Hidden Semi-Markov Model for Photoplethysmogram Signal Semantic Segmentation
Navid Hasanzadeh, Shahrokh Valaee, Hojjat Salehinejad

Data2vec-Aqc: Search for the Right Teaching Assistant in the Teacher-Student Training Setup
Vasista Sai Lodagala, Sreyan Ghosh, S. Umesh

Exploiting Interactivity and Heterogeneity for Sleep Stage Classification Via Heterogeneous Graph Neural Network
Ziyu Jia, Youfang Lin, Yuhan Zhou, Xiyang Cai, Peng Zheng, Qiang Li, Jing Wang

Benchmark of Physiological Model Based and Deep Learning Based Remote Photoplethysmography in Automotive Applications
Zhiyu Wang, Xuezhi Yang, Hongzhou Lu, Caifeng Shan, Wenjin Wang

Learning Gradients of Convex Functions with Monotone Gradient Networks
Shreyas Chaudhari, Srinivasa Pranav, José M. F. Moura

Spatial Graph Signal Interpolation with an Application for Merging BCI Datasets with Various Dimensionalities
Yassine El Ouahidi, Lucas Drumetz, Giulia Lioi, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon

Phoneix: Acoustic Feature Processing Strategy for Enhanced Singing Pronunciation with Phoneme Distribution Predictor
Yuning Wu, Jiatong Shi, Tao Qian, Dongji Gao, Qin Jin

Codebook-Based User Tracking in IRS-Assisted mmWave Communication Networks
Moritz Garkisch, Vahid Jamali, Robert Schober

Efficient Large-Scale Multi-Unimodular Waveform Design with Good Correlation Properties Via Direct Phase Optimizations
Xiaohan Zhao, Yongzhe Li, Ran Tao

Diverse and Vivid Sound Generation from Text Descriptions
Guangwei Li, Xuenan Xu, Lingfeng Dai, Mengyue Wu, Kai Yu

Model-Based Spectral Reconstruction of Interferometric Acquisitions
Mohamad Jouni, Daniele Picone, Mauro Dalla Mura
Incorporating Visual Information Reconstruction into Progressive Learning for Optimizing Audio-Visual Speech Enhancement

Chen-Yue Zhang, Hang Chen, Jun Du, Bao-Cai Yin, Jia Pan, Chin-Hui Lee

Improving the Modality Representation with Multi-View Contrastive Learning for Multimodal Sentiment Analysis

Peipei Liu, Xin Zheng, Hong Li, Jie Liu, Yimo Ren, Hongsong Zhu, Limin Sun

Generative Model Based Highly Efficient Semantic Communication Approach for Image Transmission

Tianxiao Han, Jiancheng Tang, Qianqian Yang, Yiping Duan, Zhaoyang Zhang, Zhiguo Shi

Image Source Method Based on the Directional Impulse Responses

Jiarui Wang, Prasanga Samarasinghe, Thushara Abhayapala, Jihui Aimee Zhang

Music Rearrangement Using Hierarchical Segmentation

Christos Plachouras, Marius Miron

Graphit: Iterative Reweighted ℓ1 Algorithm for Sparse Graph Inference in State-Space Models

Émilie Chouzenoux, Victor Elvira

Modulo EEG Signal Recovery Using Transformer

Tianyu Geng, Feng Ji, Pratibha, Wee Peng Tay

Latent Iterative Refinement for Modular Source Separation

Dimitrios Bralios, Efthymios Tzinis, Gordon Wichern, Paris Smaragdis, Jonathan Le Roux

DisCoHead: Audio-And-Video-Driven Talking Head Generation by Disentangled Control of Head Pose and Facial Expressions

Geumbyeol Hwang, Sunwon Hong, Seunghyun Lee, Sungwoo Park, Gyeongsu Chae

MvCo-DoT: Multi-View Contrastive Domain Transfer Network for Medical Report Generation

Ruizhi Wang, Xiangtao Wang, Zhenghua Xu, Wenting Xu, Junyang Chen, Thomas Lukasiewicz

Motor Activity Recognition Using Eeg Data and Ensemble of Stacked BLSTM-LSTM Network and Transformer Model

Pallavi Kaushik, Ilina Tripathi, Partha Pratim Roy

G2PL: Lexicon Enhanced Chinese Polyphone Disambiguation Using Bert Adapter with a New Dataset

Haifeng Zhao, Hongzhi Wan, Lili Huang, Mingwei Cao

SSVMR: Saliency-Based Self-Training for Video-Music Retrieval

Xuxin Cheng, Zhihong Zhu, Hongxiang Li, Yaowei Li, Yuexian Zou

M3ST: Mix at Three Levels for Speech Translation

Xuxin Cheng, Qianqian Dong, Fengpeng Yue, Tom Ko, Mingxuan Wang, Yuexian Zou

On Weighted Cross-Entropy for Label-Imbalanced Separable Data: An Algorithmic-Stability Study

Paneesh Deora, Christos Thrampoulidis

Node-Wise Domain Adaptation Based on Transferable Attention for Recognizing Road Rage Via EEG

Xueqi Gao, Chao Xu, Yihang Song, Jing Hu, Jian Xiao, Zhaopeng Meng

Differentiable Adaptive Short-Time Fourier Transform with Respect to the Window Length

Maxime Leiber, Yosra Marnissi, Axel Barrau, Mohammed El Badaoui
Retiformer: Retinex-Based Enhancement in Transformer for Low-Light Image
Junxiang Ruan, Xiangtao Kong, Wenqi Huang, Wenming Yang

Learning Supervised Covariation Projection Through General Covariance
Xiangze Bao, Yun-Hao Yuan, Yun Li, Jipeng Qiang, Yi Zhu

A Contrastive Embedding-Based Domain Adaptation Method for Lung Sound Recognition in Children Community-Acquired Pneumonia
Dongmin Huang, Lingwei Wang, Hongzhou Lu, Wenjin Wang

On the Quantization of Recurrent Neural Networks for Smiles Generation
Adriano Durao, Joel P. Arrais, Bernardete Ribeiro, Gabriel Falcao

MFCCGAN: A Novel MFCC-Based Speech Synthesizer Using Adversarial Learning
Mohammad Reza Hasanabadi, Majid Behdad, Davood Gharavian

Passive Detection of Rank-One Gaussian Signals for Known Channel Subspaces and Arbitrary Noise
D. Ramírez, I. Santamaria, L. L. Scharf

Exploration into Translation-Equivariant Image Quantization
Woncheol Shin, Gyubok Lee, Jiyoung Lee, Eunyi Lyou, Jooseok Lee, Edward Choi

Towards Controllable Audio Texture Morphing
Chitralekha Gupta, Purnima Kamath, Yize Wei, Zhuoyao Li, Suranga Nanayakkara, Lonce Wyse

Self-Supervised Audio-Visual Speech Representations Learning by Multimodal Self-Distillation
Jing-Xuan Zhang, Genshun Wan, Zhen-Hua Ling, Jia Pan, Jianqing Gao, Cong Liu

A Novel Metric for Evaluating Audio Caption Similarity
Swapnil Bhosale, Rupayan Chakraborty, Sunil Kumar Kopparapu

Evaluating Speech–Phoneme Alignment and Its Impact on Neural Text-To-Speech Synthesis
Frank Zalkow, Prachi Govalkar, Meinard Müller, Emanuël A. P. Habets, Christian Dittmar

Continuous Learning for Blind Image Quality Assessment with Contrastive Transformer
Jifan Yang, Zhongyuan Wang, Baojin Huang, Lianbing Deng

Utilization of Bessel Beams in Wideband Sub Terahertz Communication Systems to Mitigate Beamsplit Effects in the Near-Field
Arjun Singh, Vitaly Petrov, Josep M. Jornet

NBA-OMP: Near-Field Beam-Split-Aware Orthogonal Matching Pursuit for Wideband THz Channel Estimation
Ahmet M. Elbir, Kumar Vijay Mishra, Symeon Chatzinotas

Composition of Motion from Video Animation Through Learning Local Transformations
Michalis Vrigkas, Virginia Tagka, Marina E. Plissiti, Christophoros Nikou

Delivering Speaking Style in Low-Resource Voice Conversion with Multi-Factor Constraints
Zhichao Wang, Xinheng Wang, Lei Xie, Yuanzhe Chen, Qiao Tian, Yuping Wang

Privacy-Enhanced Federated Learning Against Attribute Inference Attack for Speech Emotion Recognition
Huan Zhao, Haijiao Chen, Yufeng Xiao, Zixing Zhang
Stochastic Optimization of Vector Quantization Methods in Application to Speech and Image Processing

Mohammad Hassan Vali, Tom Bäckström

Multi-Speaker End-to-End Multi-Modal Speaker Diarization System for the MISP 2022 Challenge

Tao Liu, Zhengyang Chen, Yanmin Qian, Kai Yu

Improving Speech-To-Speech Translation Through Unlabeled Text

Xuan-Phi Nguyen, Sravya Popuri, Changhan Wang, Yun Tang, Ilia Kulikov, Hongyu Gong

HappyQuokka System for ICASSP 2023 Auditory EEG Challenge

Zhenyu Piao, Miseul Kim, Hyungchan Yoon, Hong-Goo Kang

High-Speed Drone Detection Based on Yolo-V8

Jun-Hwa Kim, Namho Kim, Chee Sun Won

Tensor Completion for Efficient and Accurate Hyperparameter Optimisation in Large-Scale Statistical Learning

Aaman Rebello, Kriton Konstantinidis, Yao Lei Xu, Danilo P. Mandic

The NPU-ASLP System for Audio-Visual Speech Recognition in MISP 2022 Challenge

Pengcheng Guo, He Wang, Bingshen Mu, Ao Zhang, Peikuan Chen

Personalized Speech Enhancement Combining Band-Split RNN and Speaker Attentive Module

Xiaoqiu Le, Li Chen, Chao He, Yiqing Guo, Cheng Chen, Xianjun Xia, Jing Lu

Multi-Task Sub-Band Network for Deep Residual Echo Suppression

Jiayao Sun, Dawei Luo, Zhaoxia Li, Jindong Li, Yukai Ju, Yang Li

Dialogue Context Modelling for Action Item Detection: Solution for ICASSP 2023 Mug Challenge Track 5

Jie Huang, Xiachong Feng, Yangfan Ye, Liang Zhao, Xiaocheng Feng, Bing Qin, Ting Liu

A Multi-Stage Low-Latency Enhancement System for Hearing Aids

Chengwei Ouyang, Kexin Fei, Haoshuai Zhou, Congxi Lu, Linkai Li

Relate Auditory Speech to EEG by Shallow–Deep Attention-Based Network

Fan Cui, Liyong Guo, Lang He, Jiayao Liu, Ercheng Pei, Yujun Wang, Dongmei Jiang

A Progressive Neural Network for Acoustic Echo Cancellation

Zhuangqi Chen, Xianjun Xia, Siyu Sun, Ziqian Wang, Cheng Chen, Guoliang Xie, Pingjian Zhang, Yijian Xiao

Inplace Cepstral Speech Enhancement System for the ICASSP 2023 Clarity Challenge

Jinjiang Liu, Xueliang Zhang

Tayloracnet: A Taylor Style Neural Network for Full-Band Echo Cancellation

Weiming Xu, Zhihao Guo

Multi-Head Attention and GRU for Improved Match-Mismatch Classification of Speech Stimulus and EEG Response

Marvin Borsdorf, Saurav Pahuja, Gabriel Ivucic, Siqi Cai, Haizhou Li, Tanja Schultz

Multi-Speaker Multi-Lingual VQ-TTS System for LIMMITS 2023 Challenge

Chenpeng Du, Yiwei Guo, Feiyu Shen, Kai Yu
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relating EEG Recordings to Speech Using Envelope Tracking and the Speech-FFR</td>
<td>12589</td>
</tr>
<tr>
<td>Mike Thornton, Danilo Mandic, Tobias Reichenbach</td>
<td></td>
</tr>
<tr>
<td>S-Feature Pyramid Network and Attention Model for Drone Detection</td>
<td>12591</td>
</tr>
<tr>
<td>Pengcheng Dong, Chuntao Wang, Zhenyong Lu, Kai Zhang, Wenbo Wan, Jianede Sun</td>
<td></td>
</tr>
<tr>
<td>E-Branchformer-Based E2E SLU Toward Stop on-Device Challenge</td>
<td>12593</td>
</tr>
<tr>
<td>Yosuke Kashiwagi, Siddhant Arora, Hayato Futami, Jessica Huynh, Shih-Lun Wu, Yifan Peng, Brian Yan, Emiru Tsunoo, Shinji Watanabe</td>
<td></td>
</tr>
<tr>
<td>W2KPE: Keyphrase Extraction with Word-Word Relation</td>
<td>12595</td>
</tr>
<tr>
<td>Wen Cheng, Shichen Dong, Wei Wang</td>
<td></td>
</tr>
<tr>
<td>CONSEN: Complementary and Simultaneous Ensemble for Alzheimer’s Disease Detection and MMSE Score Prediction</td>
<td>12597</td>
</tr>
<tr>
<td>Longbin Jin, Yealim Oh, Hyunseo Kim, Hyuntaek Jung, Hyo Jin Jon, Jung Eun Shin, Eun Yi Kim</td>
<td></td>
</tr>
<tr>
<td>The WHU-Alibaba Audio-Visual Speaker Diarization System for the MISP 2022 Challenge</td>
<td>12599</td>
</tr>
<tr>
<td>Ming Cheng, Haixu Wang, Ziteng Wang, Qiang Fu, Ming Li</td>
<td></td>
</tr>
<tr>
<td>A Two-Stage System for Spoken Language Understanding</td>
<td>12601</td>
</tr>
<tr>
<td>Gaosheng Zhang, Shilei Miao, Linghui Tang, Peijia Qian</td>
<td></td>
</tr>
<tr>
<td>A Low-Latency Deep Hierarchical Fusion Network for Fullband Acoustic Echo Cancellation</td>
<td>12603</td>
</tr>
<tr>
<td>Haoran Zhao, Nan Li, Runqiang Han, Xiguang Zheng, Chen Zhang, Liang Guo, Bing Yu</td>
<td></td>
</tr>
<tr>
<td>Transductive Matrix Completion with Calibration for Multi-Task Learning</td>
<td>12605</td>
</tr>
<tr>
<td>Hengfang Wang, Yasi Zhang, Xiaojun Mao, Zhonglei Wang</td>
<td></td>
</tr>
<tr>
<td>Deep Learning-Based Path Loss Prediction for Outdoor Wireless Communication Systems</td>
<td>12610</td>
</tr>
<tr>
<td>Kehai Qiu, Stefanos Bakirtzis, Hui Song, Ian Wassell, Jie Zhang</td>
<td></td>
</tr>
<tr>
<td>Ensemble and Personalized Transformer Models for Subject Identification and Relapse Detection in E-Prevention Challenge</td>
<td>12612</td>
</tr>
<tr>
<td>Salvatore Calcagno, Raffaele Mineo, Daniela Giordano, Concetto Spampinato</td>
<td></td>
</tr>
<tr>
<td>Speech Signal Improvement Using Causal Generative Diffusion Models</td>
<td>12614</td>
</tr>
<tr>
<td>Julius Richter, Simon Welker, Jean-Marie Lemercier, Bunlong Lay, Tal Peer, Timo Gerkmann</td>
<td></td>
</tr>
<tr>
<td>A Transformer-Based E2E SLU Model for Improved Semantic Parsing</td>
<td>12616</td>
</tr>
<tr>
<td>Othman Istaiteh, Yasmeen Kussad, Yahya Daqour, Maria Habib, Mohammad Habash, Dhananjaya Gowda</td>
<td></td>
</tr>
<tr>
<td>Half-Temporal and Half-Frequency Attention U2Net for Speech Signal Improvement</td>
<td>12618</td>
</tr>
<tr>
<td>Zehua Zhang, Shiyun Xu, Xuyi Zhuang, Yukun Qian, Lianyu Zhou, Mingjiang Wang</td>
<td></td>
</tr>
<tr>
<td>Drone-Vs-Bird: Drone Detection Using YOLOv7 with CSRT Tracker</td>
<td>12620</td>
</tr>
<tr>
<td>Sahaj K. Mistry, Shreyas Chatterjee, Ajeet K. Verma, Vinit Jakhettiya, Badri N. Subudhi, Sunil Jaiswal</td>
<td></td>
</tr>
<tr>
<td>Decoding Auditory EEG Responses Using an Adapted Wavenet</td>
<td>12622</td>
</tr>
<tr>
<td>Bob Van Dyck, Liuyin Yang, Marc M. Van Hulle</td>
<td></td>
</tr>
<tr>
<td>Agile Radio Map Prediction Using Deep Learning</td>
<td>12624</td>
</tr>
<tr>
<td>Enes Krijestorac, Hazem Sallouha, Shamik Sarkar, Daniela Cabric</td>
<td></td>
</tr>
</tbody>
</table>
A Study on the Integration of Pipeline and E2E SLU Systems for Spoken Semantic Parsing Towards Stop Quality Challenge .......................................................... 12626
Siddhant Arora, Hayato Futami, Shih-Lun Wu, Jessica Huynh, Yifan Peng, Yosuke Kashiwagi, Emiru Tsunoo, Brian Yan, Shinji Watanabe

Multi-Channel Speaker Extraction with Adversarial Training: The Wavlab Submission to the Clarity ICASSP 2023 Grand Challenge ........................................................................ 12628
Samuele Cornell, Zhong-Qiu Wang, Yoshiki Masuyama, Shinji Watanabe, Manuel Pariente, Nobutaka Ono, Stefano Squartini

PMNet: Large-Scale Channel Prediction System for ICASSP 2023 First Pathloss Radio Map Prediction Challenge ........................................................................ 12630
Ju-Hyung Lee, Joochan Lee, Seon-Ho Lee, Andreas F. Molisch

The Pipeline System of ASR and NLU with MLM-Based Data Augmentation Toward Stop Low-Resource Challenge ........................................................................ 12632
Hayato Futami, Jessica Huynh, Siddhant Arora, Shih-Lun Wu, Yosuke Kashiwagi, Yifan Peng, Brian Yan, Emiru Tsunoo, Shinji Watanabe

Exploring Language-Agnostic Speech Representations Using Domain Knowledge for Detecting Alzheimer’s Dementia ........................................................................ 12634
Zehra Shah, Shi-Ang Qi, Fei Wang, Mahtab Farrokh, Mashrura Tasnim, Eleni Stroulia, Russell Greiner, Manos Plitsis, Athanasios Katsamanis

The Ustc System for Adress-M Challenge ........................................................................ 12636
Kangdi Mei, Xinyun Ding, Yinlong Liu, Zhiqiang Guo, Feiyang Xu, Xin Li, Tuya Naren, Jiahong Yuan, Zhenhua Ling

Vani: Very-Lightweight Accent-Controllable TTS for Native and Non-Native Speakers with Identity Preservation ................................................................. 12638
Rohan Badlani, Akshit Arora, Subhankar Ghosh, Rafael Valle, Kevin J. Shih, João Felipe Santos, Boris Ginsburg, Bryan Catanzaro

Cross-Lingual Alzheimer’s Disease Detection Based on Paralinguistic and Pre-Trained Features ................................................................. 12640
Xuchu Chen, Yu Pu, Jinpeng Li, Wei-Qiang Zhang

Two-Step Band-Split Neural Network Approach for Full-Band Residual Echo Suppression ................................................................. 12642
Zihan Zhang, Shimin Zhang, Mingshuai Liu, Yanhong Leng, Zhe Han, Li Chen, Lei Xie

De’hubert: Disentangling Noise in a Self-Supervised Model for Robust Speech Recognition ................................................................. 12644
Dianwen Ng, Ruixi Zhang, Jia Qi Yip, Zhao Yang, Jinjie Ni, Chong Zhang, Yukun Ma, Chongjia Ni, Eng Siong Chng, Bin Ma

A Person Identification System for the ICASSP 2023 e-Prevention Challenge ................................................................. 12649
Jinting Wu, Mei Tu

LeanSpeech: The Microsoft Lightweight Speech Synthesis System for Limmits Challenge 2023 ................................................................. 12651
Chen Zhang, Shubham Bansal, Aakash Lakhera, Jinzhu Li, Gang Wang, Sandeepkumar Satpal, Sheng Zhao, Lei He

The NPU-Elevoc Personalized Speech Enhancement System for Icassp2023 DNS Challenge ................................................................. 12653
Xiaopeng Yan, Yindi Yang, Zhihao Guo, Liangliang Peng, Lei Xie

The XMU System for Audio-Visual Diarization and Recognition in MISP Challenge 2022 ................................................................. 12655
Tao Li, Haodong Zhou, Jie Wang, Qingyang Hong, Lin Li
A Low-Latency Hybrid Multi-Channel Speech Enhancement System for Hearing Aids

Tong Lei, Zhongshu Hou, Yuxiang Hu, Wanyu Yang, Tianchi Sun, Xiaobin Rong, Dahan Wang, Kai Chen, Jing Lu


Yukai Ju, Jun Chen, Shimin Zhang, Shulin He, Wei Rao, Weixin Zhu, Yannan Wang, Tao Yu, Shidong Shang

Real-Time Speech Interruption Analysis: from Cloud to Client Deployment

Quchen Fu, Szu-Wei Fu, Yaran Fan, Yu Wu, Zhuo Chen, Jayant Gupchup, Ross Cutler

TSpeech-AI System Description to the 5th Deep Noise Suppression (DNS) Challenge

Jianwei Yu, Hangting Chen, Yi Luo, Rongzhi Gu, Weihua Li, Chao Weng

SSI-Net: A Multi-Stage Speech Signal Improvement System for ICASSP 2023 SSI Challenge

Weixin Zhu, Zilin Wang, Jixiu Lin, Chang Zeng, Tao Yu

Post-Trained Language Model Adaptive to Extractive Summarization of Long Spoken Documents

Hyunjong Ok, Seong-Bae Park

Convolutional Recurrent MetriCGAN with Spectral Dimension Compression for Full-Band Speech Enhancement

Zhongshu Hou, Qinwen Hu, Tianchi Sun, Yuxiang Hu, Changbao Zhu, Kai Chen

The Ajmide Topic Segmentation System for the ICASSP 2023 General Meeting Understanding and Generation Challenge

Beibei Hu, Qiang Li, Xianjun Xia

Signal Processing Grand Challenge 2023 – E-Prevention: Sleep Behavior as an Indicator of Relapses in Psychotic Patients

Kleanthis Avramidis, Kranti Adsul, Digbalay Bose, Shrikanth Narayanan

Dual-Path Dilated Convolutional Recurrent Network with Group Attention for Multi-Channel Speech Enhancement

Jiaming Cheng, Cong Pang, Ruiyu Liang, Jingjie Fan, Li Zhao

Lightweight Prosody-TTS for Multi-Lingual Multi-Speaker Scenario

Giridhar Pamisetty, S Chaitanya Varun, K Sri Rama Murty

Two-Stage Neural Network for ICASSP 2023 Speech Signal Improvement Challenge

Mingshuai Liu, Shubo Lv, Zihan Zhang, Runduo Han, Xiang Hao, Xianjun Xia, Li Chen, Yijian Xiao, Lei Xie

Relapse Detection in Patients with Psychotic Disorders Using Unsupervised Learning on Smartwatch Signals

Salam Hamieh, Vincent Heiries, Hussein Al Osman, Christelle Godin

The NIO System for Audio-Visual Diarization and Recognition in MISP Challenge 2022

Gaopeng Xu, Xianliang Wang, Sang Wang, Junfeng Yuan, Wei Guo, Wei Li, Jie Gao

Stream Attention Based U-Net for L3DAS23 Challenge

Honglong Wang, Yanjie Fu, Junjie Li, Meng Ge, Longbiao Wang, Xinyuan Qian

3D Audio Signal Processing Systems for Speech Enhancement and Sound Localization and Detection

Jisheng Bai, Siwei Huang, Han Yin, Yafei Jia, Mou Wang, Jianfeng Chen
The NERCSLIP-USTC System for the L3DAS23 Challenge Task2: 3D Sound Event Localization and Detection (SELD)................................................................. 12692
Haoyin Yan, Haitao Xu, Qing Wang, Jie Zhang

Autonomous Soundscape Augmentation with Multimodal Fusion of Visual and Participant-Linked Inputs................................................................. 12694
Kenneth Ooi, Karn N. Watcharasupat, Bhan Lam, Zhen-Ting Ong, Woon-Seng Gan

Cross-Lingual Transfer Learning for Alzheimer’s Detection from Spontaneous Speech................................................................. 12699
Bastiaan Tamm, Rik Vandenberghe, Hugo Van Hamme

Person Identification with Wearable Sensing Using Missing Feature Encoding and Multi-Stage Modality Fusion ................................................................. 12701
Payal Mohapatra, Akash Pandey, Sinan Keten, Wei Chen, Qi Zhu

Encoder-Decoder Graph Convolutional Network for Automatic Timed-Up-and-Go and Sit-to-Stand Segmentation................................................................. 12703
Bo Wen, Chen Du, Truong Q. Nguyen

PhaseAug: A Differentiable Augmentation for Speech Synthesis to Simulate One-To-Many Mapping ................................................................. 12708
Junhyeok Lee, Seungu Han, Hyunjae Cho, Wonbin Jung

Lightweight Machine Learning for Seizure Detection on Wearable Devices................................................................. 12713
Baichuan Huang, Azra Abtahi, Amir Aminifar

Pretrained Transformers for Seizure Detection ......................................................................................................................... 12715
Saarang Panchavati, Samuel Vander Dussen, Hemal Semwal, Ahmed Ali, Justin Chen, Haoran Li, Corey Arnold, William Speier

Towards Interpretable Seizure Detection Using Wearables .................................................................................................. 12717
Irfan Al-Hussaini, Cassie S. Mitchell

Optimization of the Deep Neural Networks for Seizure Detection.......................................................................................... 12719
Alexey Shovkun, Andrey Kiryasov, Ilya Zakharov, Mariam Khayretdinova

Learnt Mutual Feature Compression for Machine Vision .................................................................................................. 12721
Tie Liu, Mai Xu, Shengxi Li, Chaoran Chen, Li Yang, Zhuoyi Lv

Boosting Person Re-Identification with Viewpoint Contrastive Learning and Adversarial Training ................................................................. 12726
Xingyue Shi, Hong Liu, Wei Shi, Zihui Zhou, Yidi Li

Graph-Graph Context Dependency Attention for Graph Edit Distance.................................................................................. 12731
Ruiqi Jia, Xianbing Feng, Xiaqing Lyu, Zhi Tang

Improving Sentence Similarity Estimation for Unsupervised Extractive Summarization................................................................. 12736
Shichao Sun, Ruifeng Yuan, Wenjie Li, Sujian Li

C2BN: Cross-Modality and Cross-Scale Balance Network for Multi-Modal 3D Object Detection ................................................................. 12741
Bonan Ding, Jin Xie, Jing Nie

An Attention-Based Approach to Hierarchical Multi-Label Music Instrument Classification .................................................................................................. 12746
Zhi Zhong, Masato Hirano, Kazuki Shimada, Kazuya Tateishi, Shusuke Takahashi, Yuki Mitsufuji

K2NN: Self-Supervised Learning with Hierarchical Nearest Neighbors for Remote Sensing .................................................................................................. 12751
Jianlong Yuan, Yuanhong Xu, Zhibin Wang
Light Field Compression Via Compact Neural Scene Representation .......................................................... 12756
  Jinglei Shi, Christine Guillemot

Logovit: Local-Global Vision Transformer for Object Re-Identification ....................................................... 12761
  Nguyen Phan, Ta Duc Huy, Soan T. M. Duong, Nguyen Tran Hoang, Sam Tran, Dao Huu Hung, Chanh D. Tr. Nguyen, Trung Bui, Steven Q. H. Truong

Semantics-Guided Object Removal for Facial Images: with Broad Applicability and Robust Style Preservation .......................................................................................................................... 12766
  Jookyung Song, Yeonjin Chang, Seonguk Park, Nojun Kwak

Adaptive Data Augmentation for Contrastive Learning ..................................................................................... 12771
  Yuhan Zhang, He Zhu, Shan Yu

Bipartite Graph Convolutional Networks with Adversarial Domain Transfer ............................................. 12776
  Dong Wu, Bin Liang, Xiangjun Liu, Xuan Zang, Mingmin Chi

Context-Aware Face Clustering with Graph Convolutional Networks .......................................................... 12781
  Dafeng Zhang, Jiangbo Guo, Zhezhu Jin

A Lightweight Convolutional Neural Network Using Feature Filtering Module ........................................... 12786
  Nan Jing, Yu Zhang

Meta++ Network for Few-Shot Aerospace Crack Segmentation ...................................................................... 12791
  Chengyuan Xu, Kang Liu, Xuelong Li

Time-Domain Speech Enhancement Assisted by Multi-Resolution Frequency Encoder and Decoder .......... 12796
  Hao Shi, Masato Mimura, Longbiao Wang, Jianwu Dang, Tatsuya Kawahara

Contextual Similarity is More Valuable than Character Similarity: An Empirical Study for Chinese Spell Checking ............................................................................................................. 12801
  Ding Zhang, Yinghui Li, Qingyu Zhou, Shirong Ma, Yangning Li, Yunbo Cao, Hai-Tao Zheng

FedVMR: A New Federated Learning Method for Video Moment Retrieval .................................................. 12806
  Yan Wang, Xin Luo, Zhen-Duo Chen, Peng-Fei Zhang, Meng Liu, Xin-Shun Xu

Prompt-Distiller: Few-Shot Knowledge Distillation for Prompt-Based Language Learners with Dual Contrastive Learning ........................................................................................................ 12811
  Boyu Hou, Chengyu Wang, Xiaqing Chen, Minhui Qiu, Liang Feng, Jun Huang

Hyernetwork-Based Adaptive Image Restoration ............................................................................................... 12816
  Shai Aharon, Gil Ben-Artzi

Vision2Touch: Imaging Estimation of Surface Tactile Physical Properties ................................................... 12821
  Jie Chen, Shizhe Zhou

Dialogue System with Missing Observation .................................................................................................. 12826
  Djallel Bouneffouf, Mayank Agarwal, Irina Rish

Adversarially Robust Fairness-Aware Regression ............................................................................................ 12831
  Yulu Jin, Lifeng Lai

I See What You Hear: A Vision-Inspired Method to Localize Words ............................................................. 12836
  Mohammad Samragh, Arnav Kundu, Ting-Yao Hu, Aman Chadha, Ashish Srivastava, Minsik Cho, Oncel Tuzel, Devang Naik
Stabilising and Accelerating Light Gated Recurrent Units for Automatic Speech Recognition

Adel Moumen, Titouan Parcollet

Constrained Independent Component Analysis Based on Entropy Bound Minimization for Subgroup Identification from Multi-Subject fMRI Data

H. Yang, F. Ghayem, B. Gabrielson, M. A. B. S. Akhonda, V. D. Calhoun, T. Adali

Step Restriction for Improving Adversarial Attacks

Keita Goto, Shinta Otake, Rei Kawakami, Nakamasa Inoue

Semi-Supervised Local Structured Feature Learning with Dynamic Maximum Entropy Graph

Rui Xu, Xun Liang

Neural Architecture of Speech

Subba Reddy Oota, Khushbu Pahwa, Mounika Marreddy, Manish Gupta, Bapi S. Raju

Cocktail Hubert: Generalized Self-Supervised Pre-Training for Mixture and Single-Source Speech

Maryam Fazel-Zarandi, Wei-Ning Hsu

Image Generation is May All You Need for VQA

Kyungho Kim, Junseo Lee, Jihwa Lee

Top-K Visual Tokens Transformer: Selecting Tokens for Visible-Infrared Person Re-Identification

Bin Yang, Jun Chen, Mang Ye

Weakly- And Semi-Supervised Object Localization

Zhen-Tang Huang, Yan-He Chen, Mei-Chen Yeh

Kalmanbot: Kalmannet-Aided Bollinger Bands for Pairs Trading

Haoran Deng, Guy Revach, Hai Morgenstern, Nir Shlezinger

Complementary Learning System Based Intrinsic Reward in Reinforcement Learning

Zijian Gao, Kele Xu, Hongda Jia, Tianjiao Wan, Bo Ding, Dawei Feng, Xinjun Mao, Huaimin Wang

Filter Pruning Via Filters Similarity in Consecutive Layers

Xiaorui Wang, Jun Wang, Xin Tang, Peng Gao, Rui Fang, Guotong Xie

Coarse-To-Fine Knowledge Selection for Document Grounded Dialogs

Ye Qin Zhang, Haomin Fu, Cheng Fu, Haiyang Yu, Yongbin Li, Cam-Tu Nguyen

Audio-Driven High Definition and Lip-Synchronized Talking Face Generation Based on Face Reenactment

Xianyu Wang, Yihan Zhang, Weihua He, Yaoyuan Wang, Minglei Li, Yuchen Wang, Jingyi Zhang, Shunbo Zhou, Ziyang Zhang

Advancing the Dimensionality Reduction of Speaker Embeddings for Speaker Diarisation: Disentangling Noise and Informing Speech Activity

You Jin Kim, Hee-Soo Heo, Jee-Weon Jung, Youngki Kwon, Bong-Jin Lee, Joon Son Chung

Improving Adversarial Robustness with Hypersphere Embedding and Angular-Based Regularizations

Olukorede Fakorede, Ashutosh Nirala, Modeste Atsague, Jin Tian

Matching-Based Term Semantics Pre-Training for Spoken Patient Query Understanding

Zefa Hu, Xiaoyi Chen, Haoran Wu, Mingle Han, Ziyi Ni, Jing Shi, Shuang Xu, Bo Xu
BTS-E: Audio Deepfake Detection Using Breathing-Talking-Silence Encoder................................................. 12926
  Thien-Phuc Doan, Long Nguyen-Vu, Souhwan Jung, Kihun Hong

Keyword-Specific Acoustic Model Pruning for Open-Vocabulary Keyword Spotting ..................................... 12931
  Yujie Yang, Kun Zhang, Zhiyong Wu, Helen Meng

Long-Memory Message-Passing for Spatially Coupled Systems................................................................. 12936
  Keigo Takeuchi

Hierarchical Multi-Task Learning for Fabric Component Analysis Based on NIR Spectral Signals .......... 12941
  Joseph Kim, Dong Wu, Mingmin Chi, Gaoqi Xu

Boosting Prompt-Based Few-Shot Learners Through Out-of-Domain Knowledge Distillation .............. 12946
  Xiaojing Chen, Chengyu Wang, Junwei Dong, Minghui Qiu, Liang Feng, Jun Huang

Transient Dictionary Learning for Compressed Time-Of-Flight Imaging............................................... 12951
  Miguel Heredia Conde

Biologically-Inspired Continual Learning of Human Motion Sequences................................................. 12956
  Joachim Ott, Shih-Chii Liu

Towards a Unified Training for Levenshtein Transformer ........................................................................... 12961
  Kangjie Zheng, Longyue Wang, Zhihao Wang, Binqi Chen, Ming Zhang, Zhaopeng Tu

Nested Attention Network with Graph Filtering for Visual Question and Answering............................. 12966
  Jing Lu, Chunjie Wu, Leiqian Wang, Shaozhu Yuan, Jie Wu

Unsupervised Deep Digital Staining for Microscopic Cell Images Via Knowledge Distillation .......... 12971
  Ziwen Xu, Lanqing Guo, Shuyan Zhang, Alex C. Kot, Bihan Wen

A Comparison of Semi-Supervised Learning Techniques for Streaming ASR at Scale ...................... 12976
  Cal Peyser, Michael Picheny, Kyunghyun Cho, Rohit Prabhavalkar, W. Ronny Huang, Tara N. Sainath

Extended Expectation Maximization for Under-Fitted Models...................................................................... 12981
  Aref Miri Rekavandi, Abd-Krim Seghouane, Farid Boussaid, Mohammed Bennamoun

WAVELET2VEC: A Filter Bank Masked Autoencoder for EEG-Based Seizure Subtype Classification ................................................................. 12986
  Ruimin Peng, Changming Zhao, Yifan Xu, Jun Jiang, Guangtao Kuang, Jianbo Shao, Dongrui Wu

Semantics-Aware Gamma Correction for Unsupervised Low-Light Image Enhancement ...................... 12991
  Yu-Hsuan Chen, Fu-Cheng Pan, Yu-Chien Liao, Jiao-Hong Kao, Yu-Chiang Frank Wang

ST360IQ: No-Reference Omnidirectional Image Quality Assessment with Spherical Vision Transformers......................................................................................................................... 12996
  Nafiseh Jabbari Tofighi, Mohamed Hedi Elfkir, Nevrez Imamoglu, Cagri Oczinac, Erkut Erdem, Aykut Erdem

Quantifying Catastrophic Forgetting in Continual Federated Learning ..................................................... 13001
  Christophe Dupuy, Jimit Majmudar, Jixuan Wang, Tanya G. Roosta, Rahul Gupta, Clement Chung, Jie Ding, Salman Avestimehr

Breaking the Trade-Off in Personalized Speech Enhancement with Cross-Task Knowledge Distillation .......................................................................................................................... 13006
  Hassan Taherian, Sefik Emre Eskimez, Takuya Yoshioka
Multi-Blank Transducers for Speech Recognition
Hainan Xu, Fei Jia, Somshubra Majumdar, Shinji Watanabe, Boris Ginsburg

E2E Segmentation in a Two-Pass Cascaded Encoder ASR Model
W. Ronny Huang, Shuo-Yii Chang, Tara N. Sainath, Yanzhang He, David Rybach, Robert David, Rohit Prabhavalkar, Cyril Allauzen, Cal Peyser, Trevor D. Strohman

CO-NET: Classification-Oriented Point Cloud Sampling Via Informative Feature Learning and Non-Overlapped Local Adjustment
Yanan Lin, Keyu Chen, Shihao Zhou, Yunan Huang, Yunqi Lei

A Study on the Invariance in Security Whatever the Dimension of Images for the Steganalysis by Deep-Learning
Kévin Planolles, Marc Chaumont, Frédéric Comby

Data-Aware Zero-Shot Neural Architecture Search for Image Recognition
Yi Fan, Zhong-Han Niu, Yu-Bin Yang

Structure-Preserving and Redundancy-Free Features Refinement for Generalized Zero-Shot Learning
Jian Ni, Yong Liao

Variational Message Passing-Based Respiratory Motion Estimation and Detection Using Radar Signals
Jakob Möderl, Erik Leitinger, Franz Pernkopf, Klaus Witrisal

Depth Estimation for a Single Omnidirectional Image with Reversed-Gradient Warming-up Thresholds Discriminator
Yihong Wu, Yuwen Heng, Mahesan Niranjan, Hansung Kim

Perceive and Predict: Self-Supervised Speech Representation Based Loss Functions for Speech Enhancement
George Close, William Ravenscroft, Thomas Hain, Stefan Goetze

A Template Matching Approach for Reference Picture Padding in Video Coding
Nicolas Horst, Priyanka Das, Mathias Wien

Channel Estimation in Massive MIMO with Heavy-Tailed Noise: Gaussian-Mixture Versus Cauchy Models
Ziya Gülgün, Erik G. Larsson

Learning to Locate Visual Answer in Video Corpus Using Question
Bin Li, Yixuan Weng, Bin Sun, Shutao Li

Subspace Hybrid Beamforming for Head-Worn Microphone Arrays
Sina Hafezi, Alastair H. Moore, Pierre Guiraud, Patrick A. Naylor, Jacob Donley, Vladimir Tourbabin, Thomas Lunner

Overlay Cognitive Radio Using Symbol Level Precoding with Quantized CSI
Lu Liu, A. Lee Swindlehurst

Subject-Specific Adaptation for a Causally-Trained Auditory-Attention Decoding System
Christine Beauchene, Michael Brandstein, Stephanie Haro, Thomas F. Quattieri, Christopher J. Smalt

Synthesizer Preset Interpolation Using Transformer Auto-Encoders
Gwendal Le Vaillant, Thierry Dutoit
Audiodec: An Open-Source Streaming High-Fidelity Neural Audio Codec .................................................. 13090
Yi-Chiao Wu, Israel D. Gebru, Dejan Markovic, Alexander Richard

Learning Cross-Modal Audiovisual Representations with Ladder Networks for Emotion Recognition ................................................................. 13095
Lucas Goncalves, Carlos Busso

Wasserstein Gan Synthesis for Time Series with Complex Temporal Dynamics: Frugal Architectures and Arbitrary Sample-Size Generation .................................................................................. 13100

Sine: Similarity-Regularized Intra-Class Exploitation for Cross-Granularity Few-Shot Learning .................. 13105
Jinhai Yang, Hua Yang

Simulating Realistic Speech Overlaps Improves Multi-Talker ASR .......................................................... 13110
Muqiao Yang, Naoyuki Kanda, Xiaofei Wang, Jian Wu, Sunit Sivasankaran, Zhuo Chen, Jinyu Li, Takuya Yoshioka

Paaploss: A Phonetic-Aligned Acoustic Parameter Loss for Speech Enhancement ...................................... 13115
Muqiao Yang, Joseph Konan, David Bick, Yinyang Zeng, Shuo Han, Anurag Kumar, Shinji Watanabe, Bhiksha Raj

Deep Quantigraphic Image Enhancement Via Comparametric Equations .................................................. 13120
Xiaomeng Wu, Yongqing Sun, Akisato Kimura

Transadapt: A Transformative Framework for Online Test Time Adaptive Semantic Segmentation .......... 13125
Debasmit Das, Shubhankar Borse, Hyojin Park, Kambiz Azarian, Hong Cai, Risheek Garrepalli, Fatih Porikli

Inverse Quadratic Transform for Minimizing a Sum of Ratios ................................................................... 13130
Yannan Chen, Licheng Zhao, Yaowen Zhang, Kaiming Shen

Boosting Fine-Grained Sketch-Based Image Retrieval with Self-Supervised Learning ............................. 13135
Zhaolong Zhang, Yangdong Chen, Yuejie Zhang, Rui Feng, Tao Zhang

D2Former: A Fully Complex Dual-Path Dual-Decoder Conformer Network Using Joint Complex Masking and Complex Spectral Mapping for Monaural Speech Enhancement .................................................. 13140
Shengkui Zhao, Bin Ma

MossFormer: Pushing the Performance Limit of Monaural Speech Separation Using Gated Single-Head Transformer with Convolution-Augmented Joint Self-Attentions .............................................. 13145
Shengkui Zhao, Bin Ma

Target Speaker Voice Activity Detection with Transformers and Its Integration with End-To-End Neural Diarization .................................................................................................................. 13150
Dongmei Wang, Xiong Xiao, Naoyuki Kanda, Takuya Yoshioka, Jian Wu

Space-Time Variable Density Samplings for Sparse Bandlimited Graph Signals Driven by Diffusion Operators .............................................................................. 13155
Qing Yao, Longxiu Huang, Sui Tang

Long-Term Synchronization of Wireless Acoustic Sensor Networks with Nonpersistent Acoustic Activity Using Coherence State ................................................................................................. 13160
Aleksej Chinaev, Niklas Knaepper, Gerald Enzner

Mitigating Unintended Memorization in Language Models Via Alternating Teaching .................................. 13165
Zhe Liu, Xuedong Zhang, Fuchun Peng
Video Captioning Via Relation-Aware Graph Learning ................................................................. 13170
Yi Zheng, Heming Jing, Qiujuie Xie, Yuejie Zhang, Rui Feng, Tao Zhang, Shang Gao

Acoustically-Driven Phoneme Removal that Preserves Vocal Affect Cues........................................ 13175
Camille Noufi, Jonathan Berger, Karen J. Parker, Daniel L. Bowling

Multi-Source Templates Learning for Real-Time Aerial Tracking .................................................. 13180
Yiming Sun, Yang Li, Changbo Wang

Diffroll: Diffusion-Based Generative Music Transcription with Unsupervised Pretraining Capability ............................................................................ 13185
Kin Wai Cheuk, Ryosuke Sawata, Toshimitsu Uesaka, Naoki Murata, Naoya Takahashi,
Shusuke Takahashi, Doreen Herremans, Yuki Mitsufuji

Sparsity-Smoothness-Aware Power Spectral Density Estimation with Application to Phased Array Weather Radar ........................................................................................................ 13190
Hiroki Kuroda, Daichi Kitahara, Eiichi Yoshikawa, Hiroshi Kikuchi, Tomoo Ushio

Optimal Mixed-ADC Arrangement for DOA Estimation Via CRB Using ULA ........................................ 13195
Xinnan Zhang, Yuanbo Cheng, Xiaolei Shang, Jun Liu

Higher-Order Sparse Convolutions in Graph Neural Networks ............................................................ 13200
Jhony H. Giraldo, Sajid Javed, Arif Mahmood, Fragkiskos D. Malliaros, Thierry Bouwmans

Improving Weakly Supervised Sound Event Detection with Causal Intervention ............................... 13205
Yifei Xin, Dongchao Yang, Fan Cui, Yujun Wang, Yuexian Zou

Robust Iterative Solution for Linear Array-Based 3-D Localization by Message Passing ...................... 13210
Yimao Sun, K. C. Ho, Yanbing Yang, Lei Zhang, Liangyun Chen

Backdoor Defense Via Suppressing Model Shortcuts ........................................................................... 13215
Sheng Yang, Yiming Li, Yong Jiang, Shu-Tao Xia

Sparse Representations with Cone Atoms ............................................................................................. 13220
Denis C. Ilie-Ablachim, Andra Baltoiu, Bogdan Dumitrescu

A Spatio-Temporal Decomposition Network for Compressed Video Quality Enhancement .................. 13225
Kai Wang, Fangdong Chen, Zongming Ye, Li Wang, Xiaoyang Wu, Shiliang Pu

A Physically Explainable Framework for Human-Related Anomaly Detection ........................................ 13230
Yalong Jiang, Huining Li, Changkang Li

High-Resolution Neural Network Processing of LFM Radar Pulses .................................................... 13235
Jabran Akhtar

Author Index