

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

**Rhodes Island, Greece
4-10 June 2023**

Pages 1-365



**IEEE Catalog Number: CFP23ICA-POD
ISBN: 978-1-7281-6328-4**

**Copyright © 2023 by the Institute of Electrical and Electronics Engineers, Inc.
All Rights Reserved**

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

****** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

| | |
|-------------------------|-------------------|
| IEEE Catalog Number: | CFP23ICA-POD |
| ISBN (Print-On-Demand): | 978-1-7281-6328-4 |
| ISBN (Online): | 978-1-7281-6327-7 |
| ISSN: | 1520-6149 |

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

TABLE OF CONTENTS

| | |
|---|----|
| Counterfactual Two-Stage Debiasing for Video Corpus Moment Retrieval..... | 1 |
| <i>Sunjae Yoon, Ji Woo Hong, Soohwan Eom, Hee Suk Yoon, Eunseop Yoon, Daehyeok Kim, Junyeong Kim, Chanwoo Kim, Chang D. Yoo</i> | |
| Test-Time Training-Free Domain Adaptation..... | 6 |
| <i>Yongxiang Feng, Weihua He, Kaichao You, Bing Liu, Ziyang Zhang, Yaoyuan Wang, Minglei Li, Yihang Lou, Jiawei Li, Guoqi Li, Jianxing Liao</i> | |
| AdapITN: A Fast, Reliable, and Dynamic Adaptive Inverse Text Normalization | 11 |
| <i>Thai-Binh Nguyen, Le Duc Minh Nhat, Quang Minh Nguyen, Quoc Truong Do, Chi Mai Luong, Alexander Waibel</i> | |
| Compressed Distributed Regression Over Adaptive Networks | 16 |
| <i>Marco Carpentiero, Vincenzo Matta, Ali H. Sayed</i> | |
| The Role of Memory in Social Learning When Sharing Partial Opinions | 21 |
| <i>Michele Cirillo, Virginia Bordignon, Vincenzo Matta, A. H. Sayed</i> | |
| IoU-Aware Multi-Expert Cascade Network Via Dynamic Ensemble for Long-Tailed Object Detection | 26 |
| <i>Wan-Cyuan Fan, Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu</i> | |
| TransLink: Transformer-Based Embedding for Tracklets' Global Link | 31 |
| <i>Yanting Zhang, Shuanghong Wang, Yuxuan Fan, Gaoang Wang, Cairong Yan</i> | |
| Burst Perception-Distortion Tradeoff: Analysis and Evaluation..... | 36 |
| <i>Danna Xue, Luis Herranz, Javier Vazquez Corral, Yanning Zhang</i> | |
| Approximation Error Back-Propagation for Q-Function in Scalable Reinforcement Learning with Tree Dependence Structure..... | 41 |
| <i>Yuzi Yan, Yu Dong, Kai Ma, Yuan Shen</i> | |
| EMC ² -Net: Joint Equalization and Modulation Classification Based on Constellation Network | 46 |
| <i>Hyun Ryu, Junil Choi</i> | |
| Improving Prosody for Cross-Speaker Style Transfer by Semi-Supervised Style Extractor and Hierarchical Modeling in Speech Synthesis | 51 |
| <i>Chunyu Qiang, Peng Yang, Hao Che, Ying Zhang, Xiaorui Wang, Zhongyuan Wang</i> | |
| CyFi-TTS: Cyclic Normalizing Flow with Fine-Grained Representation for End-to-End Text-to-Speech | 56 |
| <i>In-Sun Hwang, Young-Sub Han, Byoung-Ki Jeon</i> | |
| Learning Graph Laplacian from Intrinsic Patterns Via Gaussian Process | 61 |
| <i>Koshi Watanabe, Keisuke Maeda, Takahiro Ogawa, Miki Haseyama</i> | |
| Overcoming the Seesaw in Monocular 3D Object Detection Via Language Knowledge Transferring | 66 |
| <i>Weichen Xu, Tianhao Fu</i> | |
| Improving Dropout in Graph Convolutional Networks for Recommendation Via Contrastive Loss | 71 |
| <i>Hiroki Okamura, Keisuke Maeda, Ren Togo, Takahiro Ogawa, Miki Haseyama</i> | |

| | |
|---|-----|
| Estimation of Visual Contents from Human Brain Signals Via VQA Based on Brain-Specific Attention | 76 |
| <i>Ryo Shichida, Ren Togo, Keisuke Maeda, Takahiro Ogawa, Miki Haseyama</i> | |
| Combining the Silhouette and Skeleton Data for Gait Recognition | 81 |
| <i>Likai Wang, Ruize Han, Wei Feng</i> | |
| Unobtrusive Respiratory Monitoring System for Intensive Care | 86 |
| <i>Xudong Tan, Menghan Hu, Guangtao Zhai, Yan Zhu, Wenfang Li, Xiaoping Zhang</i> | |
| End-To-End Amp Modeling: from Data to Controllable Guitar Amplifier Models | 91 |
| <i>Lauri Juvola, Eero-Pekka Damsk agg, Aleksi Peussa, Jaakko M akinen, Thomas Sherson, Stylianos I. Mimitakis, Kimmo Rauhanen, Athanasios Gotsopoulos</i> | |
| Provably Convergent Plug & Play Linearized ADMM, Applied to Deblurring Spatially Varying Kernels | 96 |
| <i>Charles Laroche, Andr es Almansa, Eva Coupet e, Matias Tassano</i> | |
| Robust Video Anomaly Detection Framework Via Prior Knowledge and Multi-Path Frame Prediction | 101 |
| <i>Menghao Zhang, Jingyu Wang, Jing Wang, Qi Qi, Zirui Zhuang, Haifeng Sun, Ning Xiao</i> | |
| Unrolled Fourier Disparity Layer Optimization for Scene Reconstruction from Few-Shots Focal Stacks | 106 |
| <i>Brandon Le Bon, Mika el Le Pendu, Christine Guillemot</i> | |
| Adaptive Axonal Delays in Feedforward Spiking Neural Networks for Accurate Spoken Word Recognition | 111 |
| <i>Pengfei Sun, Ehsan Eqlimi, Yansong Chua, Paul Devos, Dick Botteldooren</i> | |
| EEG2IMAGE: Image Reconstruction from EEG Brain Signals | 116 |
| <i>Prajwal Singh, Pankaj Pandey, Krishna Miyapuram, Shanmuganathan Raman</i> | |
| Multispectral Image Fusion Based on Super Pixel Segmentation | 121 |
| <i>Nati Ofir</i> | |
| NF-PCAC: Normalizing Flow Based Point Cloud Attribute Compression | 126 |
| <i>Rodrigo B. Pinheiro, Jean-Eudes Marvie, Giuseppe Valenzise, Fr ed eric Dufaux</i> | |
| An Adaptive DFE Using Light-Pattern-Protection Algorithm in 12 NM CMOS Technology | 131 |
| <i>Shiyuan Xing, Changlong Lin, Yuchen Li, Huandong Wang</i> | |
| Hierarchical Transformer for Multi-Label Trailer Genre Classification | 136 |
| <i>Zihui Cai, Hongwei Ding, Xuemeng Wu, Mohan Xu, Xiaohui Cui</i> | |
| A Contrastive Framework to Enhance Unsupervised Sentence Representation Learning | 141 |
| <i>Haoyang Ma, Zeyu Li, Hongyu Guo</i> | |
| Self-Attention Based Action Segmentation Using Intra-And Inter-Segment Representations | 146 |
| <i>Constantin Patsch, Eckehard Steinbach</i> | |
| A Deep Disentangled Approach for Interpretable Hyperspectral Unmixing | 151 |
| <i>Ricardo Augusto Borsoi, Tales Imbiriba, Deniz Erdogmus</i> | |
| Residual Squeeze-And-Excitation U-Shaped Network for Minutia Extraction in Contactless Fingerprint Images | 156 |
| <i>Anderson Nogueira Cotrim, Helio Pedrini</i> | |

| | |
|---|-----|
| Learnable Flow Model Conditioned on Graph Representation Memory for Anomaly Detection | 161 |
| <i>Ziyu Zhu, Wenlei Liu, Zhidong Deng</i> | |
| Building Blocks for a Complex-Valued Transformer Architecture | 166 |
| <i>Florian Eilers, Xiaoyi Jiang</i> | |
| PCSalmix: Gradient Saliency-Based Mix Augmentation for Point Cloud Classification | 171 |
| <i>Tao Hong, Zeren Zhang, Jinwen Ma</i> | |
| Entropy Based Feature Regularization to Improve Transferability of Deep Learning Models | 176 |
| <i>Raphael Baena, Lucas Drumetz, Vincent Gripon</i> | |
| Exploring Complementary Features in Multi-Modal Speech Emotion Recognition..... | 181 |
| <i>Suzhen Wang, Yifeng Ma, Yu Ding</i> | |
| Database-Aware ASR Error Correction for Speech-to-SQL Parsing..... | 186 |
| <i>Yutong Shao, Arun Kumar, Ndapa Nakashole</i> | |
| Super-Resolution Harmonic Retrieval of Non-Circular Signals..... | 191 |
| <i>Yu Zhang, Yue Wang, Zhi Tian, Geert Leus, Gong Zhang</i> | |
| Semi-Swinderain: Semi-Supervised Image Deraining Network Using SWIN Transformer | 196 |
| <i>Chun Ren, Danfeng Yan, Yuanqiang Cai, Yangchun Li</i> | |
| Learning Generalizable Light Field Networks from Few Images | 201 |
| <i>Qian Li, Franck Multon, Adnane Boukhayma</i> | |
| Distributionally Robust Multiclass Classification and Applications in Deep Image Classifiers | 206 |
| <i>Ruidi Chen, Boran Hao, Ioannis Ch. Paschalidis</i> | |
| Adaptive Mask Co-Optimization for Modal Dependence in Multimodal Learning..... | 211 |
| <i>Ying Zhou, Xuefeng Liang, Shiquan Zheng, Huijun Xuan, Takatsune Kumada</i> | |
| Improving Occluded Human Pose Estimation Via Linked Joints..... | 216 |
| <i>Suhang Ye, Zebo Hong, Jiawen Zheng, Shengchuan Zhang</i> | |
| DST: Deformable Speech Transformer for Emotion Recognition..... | 221 |
| <i>Weidong Chen, Xiaofen Xing, Xiangmin Xu, Jianxin Pang, Lan Du</i> | |
| Permutation Invariant Training for Paraphrase Identification | 226 |
| <i>Jun Bai, Chuantao Yin, Hanhua Hong, Jianfei Zhang, Chen Li, Yanmeng Wang, Wenge Rong</i> | |
| Code-Enhanced Fine-Grained Semantic Matching for Tag Recommendation in Software Information Sites | 231 |
| <i>Lin Li, Peipei Wang, Xinhao Zheng, Qing Xie</i> | |
| Audio Coding with Unified Noise Shaping and Phase Contrast Control | 236 |
| <i>Byeongho Jo, Seungkwon Beack, Taejin Lee</i> | |
| Binary Image Fast Perfect Recovery from Sparse 2D-DFT Coefficients..... | 241 |
| <i>Soo-Chang Pei, Kuo-Wei Chang</i> | |
| Reducing the GAP Between Streaming and Non-Streaming Transducer-Based ASR by Adaptive Two-Stage Knowledge Distillation..... | 246 |
| <i>Haitao Tang, Yu Fu, Lei Sun, Jiabin Xue, Dan Liu, Yongchao Li, Zhiqiang Ma, Minghui Wu, Jia Pan, Genshun Wan, Ming'En Zhao</i> | |

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

| | |
|--|-----|
| Improved Indoor Localization with NLOS Signal Propagations..... | 341 |
| <i>Wei Huang, Yixin Zhao, Xuechao Wu, Le Yin</i> | |
| HQP-MVS:High-Quality Plane Priors Assisted Multi-View Stereo for Low-Textured Areas | 346 |
| <i>Zefan Tian, Rongjie Wang, Zhenyu Wang, Ronggang Wang</i> | |
| MSN-Net: Multi-Scale Normality Network for Video Anomaly Detection | 351 |
| <i>Yang Liu, Di Li, Wei Zhu, Dingkan Yang, Jing Liu, Liang Song</i> | |
| High-Dynamic Range ADC for Finite-Rate-of-Innovation Signals | 356 |
| <i>Satish Mulleti, Yonina C. Eldar</i> | |
| Lip-To-Speech Synthesis in the Wild with Multi-Task Learning | 361 |
| <i>Minsu Kim, Joanna Hong, Yong Man Ro</i> | |
| NC-WAMKD: Neighborhood Correction Weight-Adaptive Multi-Teacher Knowledge Distillation for Graph-Based Semi-Supervised Node Classification..... | 366 |
| <i>Jiahao Liu, Pengcheng Guo, Yonghong Song</i> | |
| Toward Asymptotic Optimality: Sequential Unsupervised Regression of Density Ratio for Early Classification..... | 371 |
| <i>Akinori F. Ebihara, Taiki Miyagawa, Kazuyuki Sakurai, Hitoshi Imaoka</i> | |
| Improving Image Captioning with Control Signal of Sentence Quality..... | 376 |
| <i>Zhangzi Zhu, Shuai Wang, Hong Qu</i> | |
| Multiple Acoustic Features Speech Emotion Recognition Using Cross-Attention Transformer..... | 381 |
| <i>Yurun He, Nobuaki Minematsu, Daisuke Saito</i> | |
| Graph Neural Networks for Object Type Classification Based on Automotive Radar Point Clouds and Spectra | 386 |
| <i>Loveneet Saini, Axel Acosta, Gor Hakobyan</i> | |
| Frame-Level Multi-Label Playing Technique Detection Using Multi-Scale Network and Self- Attention Mechanism | 391 |
| <i>Dichucheng Li, Mingjin Che, Wenwu Meng, Yulun Wu, Yi Yu, Fan Xia, Wei Li</i> | |
| FedPrompt: Communication-Efficient and Privacy-Preserving Prompt Tuning in Federated Learning | 396 |
| <i>Haodong Zhao, Wei Du, Fangqi Li, Peixuan Li, Gongshen Liu</i> | |
| Uncer2Natural: Uncertainty-Aware Unsupervised Image Denoising..... | 401 |
| <i>Chenyu Huang, Weimin Tan, Jiaxing Shi, Zhen Xing, Bo Yan</i> | |
| Simple Pooling Front-Ends for Efficient Audio Classification | 406 |
| <i>Xubo Liu, Haohe Liu, Qiuqiang Kong, Xinhao Mei, Mark D. Plumbley, Wenwu Wang</i> | |
| A Compensated Shrinkage Affine Projection Algorithm for Debaised Sparse Adaptive Filtering..... | 411 |
| <i>Yi Zhang, Isao Yamada</i> | |
| CTCBERT: Advancing Hidden-Unit Bert with CTC Objectives..... | 416 |
| <i>Ruchao Fan, Yiming Wang, Yashesh Gaur, Jinyu Li</i> | |
| Analysis and Re-Synthesis of Natural Cricket Sounds Assessing the Perceptual Relevance of Idiosyncratic Parameters | 421 |
| <i>Marco Oliveira, Vitor Almeida, João Silva, Anibal Ferreira</i> | |

| | |
|---|-----|
| Hierarchical Spatial-Temporal Transformer with Motion Trajectory for Individual Action and Group Activity Recognition | 426 |
| <i>Xiaolin Zhu, Dongli Wang, Yan Zhou</i> | |
| S3I-PointHop: SO(3)-Invariant PointHop for 3D Point Cloud Classification..... | 431 |
| <i>Pranav Kadam, Hardik Prajapati, Min Zhang, Jintang Xue, Shan Liu, C.-C. Jay Kuo</i> | |
| Salient Co-Speech Gesture Synthesizing with Discrete Motion Representation..... | 436 |
| <i>Zijie Ye, Jia Jia, Haozhe Wu, Shuo Huang, Shikun Sun, Junliang Xing</i> | |
| Hierarchical Diffusion Models for Singing Voice Neural Vocoder | 441 |
| <i>Naoya Takahashi, Mayank Kumar, Singh, Yuki Mitsufuji</i> | |
| Period VITS: Variational Inference with Explicit Pitch Modeling for End-To-End Emotional Speech Synthesis | 446 |
| <i>Yuma Shirahata, Ryuichi Yamamoto, Eunwoo Song, Ryo Terashima, Jae-Min Kim, Kentaro Tachibana</i> | |
| Sanet: Spatial Attention Network with Global Average Contrast Learning for Infrared Small Target Detection | 451 |
| <i>Jiewen Zhu, Shengjia Chen, Lexiao Li, Luping Ji</i> | |
| An ASR-Free Fluency Scoring Approach with Self-Supervised Learning..... | 456 |
| <i>Wei Liu, Kaiqi Fu, Xiaohai Tian, Shuju Shi, Wei Li, Zejun Ma, Tan Lee</i> | |
| Memory-Augmented U-Transformer for Multivariate Time Series Anomaly Detection | 461 |
| <i>Shuxin Qin, Yongcan Luo, Gaofeng Tao</i> | |
| Optimizing Distributed Multi-Sensor Multi-Target Tracking Algorithm Based on Labeled Multi-Bernoulli Filter | 466 |
| <i>Honggang Liu, Jinlong Yang, Yue Xu, Le Yang</i> | |
| Rethink Pair-Wise Self-Supervised Cross-Modal Retrieval from a Contrastive Learning Perspective..... | 471 |
| <i>Tiantian Gong, Junsheng Wang, Liyan Zhang</i> | |
| Grad-StyleSpeech: Any-Speaker Adaptive Text-to-Speech Synthesis with Diffusion Models | 476 |
| <i>Minki Kang, Dongchan Min, Sung Ju Hwang</i> | |
| Classification Via Subspace Learning Machine (SLM): Methodology and Performance Evaluation | 481 |
| <i>Hongyu Fu, Yijing Yang, Vinod K. Mishra, C.-C. Jay Kuo</i> | |
| T5-SR: A Unified Seq-to-Seq Decoding Strategy for Semantic Parsing | 486 |
| <i>Yuntao Li, Zhenpeng Su, Yutian Li, Hanchu Zhang, Sirui Wang, Wei Wu, Yan Zhang</i> | |
| Two-Stream Decoder Feature Normality Estimating Network for Industrial Anomaly Detection..... | 491 |
| <i>Chaewon Park, Minhyeok Lee, Suhwan Cho, Donghyeong Kim, Sangyoun Lee</i> | |
| TriAAN-VC: Triple Adaptive Attention Normalization for Any-to-Any Voice Conversion | 496 |
| <i>Hyun Joon Park, Seok Woo Yang, Jin Sob Kim, Wooseok Shin, Sung Won Han</i> | |
| AD-YOLO: You Look Only Once in Training Multiple Sound Event Localization and Detection | 501 |
| <i>Jin Sob Kim, Hyun Joon Park, Wooseok Shin, Sung Won Han</i> | |
| Dual-Cycle: Self-Supervised Dual-View Fluorescence Microscopy Image Reconstruction Using CycleGAN | 506 |
| <i>Tomas Kerepecky, Jiaming Liu, Xue Wen Ng, David W. Piston, Ulugbek S. Kamilov</i> | |

| | |
|---|-----|
| Global-Context Aware Generative Protein Design | 511 |
| <i>Cheng Tan, Zhangyang Gao, Jun Xia, Bozhen Hu, Stan Z. Li</i> | |
| Semi-Supervised Speech Enhancement Based on Speech Purity | 516 |
| <i>Zihao Cui, Shilei Zhang, Yanan Chen, Yingying Gao, Chao Deng, Junlan Feng</i> | |
| Modify: Model-Driven Face Stylization Without Style Images | 521 |
| <i>Yuhe Ding, Jian Liang, Jie Cao, Aihua Zheng, Ran He</i> | |
| Precognition in Contextual Spoken Language Understanding Via Knowledge Distillation..... | 526 |
| <i>Nan Su, Bingzhu Du, Yuchi Zhang, Chao Liu, Yongliang Wang, Hong Chen, Xin Lu</i> | |
| GaPP: Multi-Target Tracking with Gaussian Processes | 531 |
| <i>Fred Goodyer, Bashar I. Ahmad, Simon Godsill</i> | |
| Adversarial Network Pruning by Filter Robustness Estimation | 536 |
| <i>Xinlu Zhuang, Yunjie Ge, Baolin Zheng, Qian Wang</i> | |
| Leveraging Phone-Level Linguistic-Acoustic Similarity for Utterance-Level Pronunciation Scoring | 541 |
| <i>Wei Liu, Kaiqi Fu, Xiaohai Tian, Shuju Shi, Wei Li, Zejun Ma, Tan Lee</i> | |
| CTTSR: A Hybrid CNN-Transformer Network for Scene Text Image Super-Resolution..... | 546 |
| <i>Kaiwei Dai, Nan Kang, Li Kuang</i> | |
| CF-VTON: Multi-Pose Virtual Try-on with Cross-Domain Fusion | 551 |
| <i>Chenghu Du, Shengwu Xiong</i> | |
| In Search of Strong Embedding Extractors for Speaker Diarisation | 556 |
| <i>Jee-Weon Jung, Hee-Soo Heo, Bong-Jin Lee, Jaesung Huh, Andrew Brown, Youngki Kwon, Shinji Watanabe, Joon Son Chung</i> | |
| Towards Making a Trojan-Horse Attack on Text-to-Image Retrieval..... | 561 |
| <i>Fan Hu, Aozhu Chen, Xirong Li</i> | |
| Elastic Graph Transformer Networks for EEG-Based Emotion Recognition..... | 566 |
| <i>Wei-Bang Jiang, Xu Yan, Wei-Long Zheng, Bao-Liang Lu</i> | |
| Raw Ultrasound-Based Phonetic Segments Classification Via Mask Modeling | 571 |
| <i>Kang You, Bo Liu, Kele Xu, Yunsheng Xiong, Qisheng Xu, Ming Feng, Tamás Gábor Csapó, Boqing Zhu</i> | |
| Content-Insensitive Dynamic Lip Feature Extraction for Visual Speaker Authentication Against Deepfake Attacks..... | 576 |
| <i>Zihao Guo, Shilin Wang</i> | |
| Sample-Aware Knowledge Distillation for Long-Tailed Learning..... | 581 |
| <i>Shanshan Zheng, Yachao Zhang, Hongyi Huang, Yanyun Qu</i> | |
| A3S: Adversarial Learning of Semantic Representations for Scene-Text Spotting..... | 586 |
| <i>Masato Fujitake</i> | |
| A Novel Cross-Component Context Model for End-to-End Wavelet Image Coding | 591 |
| <i>Anna Meyer, André Kaup</i> | |
| Multi-Head Uncertainty Inference for Adversarial Attack Detection..... | 596 |
| <i>Yuqi Yang, Songyun Yang, Jiyang Xie, Zhongwei Si, Kai Guo, Ke Zhang, Kongming Liang</i> | |

| | |
|---|-----|
| ESCL: Equivariant Self-Contrastive Learning for Sentence Representations | 601 |
| <i>Jie Liu, Yixuan Liu, Xue Han, Chao Deng, Junlan Feng</i> | |
| Vision Transformer-Based Feature Extraction for Generalized Zero-Shot Learning | 606 |
| <i>Jiseob Kim, Kyuhong Shim, Junhan Kim, Byonghyo Shim</i> | |
| A Robust Kalman Filter Based Approach for Indoor Robot Positioning with Multi-Path Contaminated UWB Data..... | 611 |
| <i>Justin Cano, Yi Ding, Gaël Pages, Eric Chaumette, Jerome Le Ny</i> | |
| Semi-Supervised Sound Event Detection with Pre-Trained Model..... | 616 |
| <i>Liang Xu, Lizhong Wang, Sijun Bi, Hanyue Liu, Jing Wang</i> | |
| Design Choices for Learning Embeddings from Auxiliary Tasks for Domain Generalization in Anomalous Sound Detection | 621 |
| <i>Kevin Wilkinghoff</i> | |
| Tensor-Based Complex-valued Graph Neural Network for Dynamic Coupling Multimodal Brain Networks | 626 |
| <i>Yanwu Yang, Guoqing Cai, Chenfei Ye, Yang Xiang, Ting Ma</i> | |
| Sparse Graph Learning with Spectrum Prior for Deep Graph Convolutional Networks..... | 631 |
| <i>Jin Zeng, Yang Liu, Gene Cheung, Wei Hu</i> | |
| Divcon: Learning Concept Sequences for Semantically Diverse Image Captioning..... | 636 |
| <i>Yue Zheng, Ya-Li Li, Shengjin Wang</i> | |
| EBEN: Extreme Bandwidth Extension Network Applied to Speech Signals Captured with Noise- Resilient Body-Conduction Microphones | 641 |
| <i>Julien Hauret, Thomas Joubaud, Véronique Zimpfer, Éric Bavu</i> | |
| Long Range Imaging Using Multispectral Fusion of RGB and NIR Images | 646 |
| <i>Hao Zhang, Lin Mei, Cheolkon Jung</i> | |
| FedEEG: Federated EEG Decoding Via inter-Subject Structure Matching..... | 651 |
| <i>Wenlong Hang, Jiaxing Li, Shuang Liang, Yuan Wu, Baiying Lei, Jing Qin, Yu Zhang, Kup-Sze Choi</i> | |
| Visual Onoma-To-Wave: Environmental Sound Synthesis from Visual Onomatopoeias and Sound- Source Images | 656 |
| <i>Hien Ohnaka, Shinnosuke Takamichi, Keisuke Imoto, Yuki Okamoto, Kazuki Fujii, Hiroshi Saruwatari</i> | |
| Asynchronous Social Learning..... | 661 |
| <i>Mert Cemri, Virginia Bordignon, Mert Kayaalp, Valentina Shumovskaia, Ali H. Sayed</i> | |
| Untargeted Backdoor Attack Against Object Detection | 666 |
| <i>Chengxiao Luo, Yiming Li, Yong Jiang, Shu-Tao Xia</i> | |
| Improving Contextual Biasing with Text Injection | 671 |
| <i>Tara N. Sainath, Rohit Prabhavalkar, Diamantino Caseiro, Pat Rondon, Cyril Allauzen</i> | |
| Noise PSD Insensitive RTF Estimation in a Reverberant and Noisy Environment..... | 676 |
| <i>Changheng Li, Richard C. Hendriks</i> | |
| Error Analysis of Convolutional Beamspace Algorithms..... | 681 |
| <i>Po-Chih Chen, P. P. Vaidyanathan</i> | |

| | |
|--|-----|
| Unitary Esprit for Coprime Arrays | 686 |
| <i>Po-Chih Chen, P. P. Vaidyanathan</i> | |
| UX-Net: Filter-and-Process-Based Improved U-Net for Real-time Time-domain Audio Separation | 691 |
| <i>Kashyap Patel, Anton Kovalyov, Issa Panahi</i> | |
| SARdBScene: Dataset and Resnet Baseline for Audio Scene Source Counting and Analysis | 696 |
| <i>Michael Nigro, Sridhar Krishnan</i> | |
| CD-FSOD: A Benchmark for Cross-Domain Few-Shot Object Detection | 701 |
| <i>Wuti Xiong</i> | |
| Fast and Accurate Factorized Neural Transducer for Text Adaption of End-To-End Speech Recognition Models | 706 |
| <i>Rui Zhao, Jian Xue, Partha Parthasarathy, Veljko Miljanic, Jinyu Li</i> | |
| Augmentation Robust Self-Supervised Learning for Human Activity Recognition..... | 711 |
| <i>Cong Xu, Yuhang Li, Dae Lee, Dae Hoon Park, Hongda Mao, Huyen Do, Jonathan Chung, Dinesh Nair</i> | |
| SPADE: Self-Supervised Pretraining for Acoustic Disentanglement | 716 |
| <i>John Harvill, Jarred Barber, Arun Nair, Ramin Pishehvar</i> | |
| Unified Keyword Spotting and Audio Tagging on Mobile Devices with Transformers..... | 721 |
| <i>Heinrich Dinkel, Yongqing Wang, Zhiyong Yan, Junbo Zhang, Yujun Wang</i> | |
| Anomalous Sound Detection Using Audio Representation with Machine ID Based Contrastive Learning Pretraining | 726 |
| <i>Jian Guan, Feiyang Xiao, Youde Liu, Qiaoxi Zhu, Wenwu Wang</i> | |
| Cross-Modality Depth Estimation Via Unsupervised Stereo RGB-to-infrared Translation | 731 |
| <i>Shi Tang, Xinchun Ye, Fei Xue, Rui Xu</i> | |
| Solving Jigsaw Puzzle of Large Eroded Gaps Using Puzzlet Discriminant Network | 736 |
| <i>Xingke Song, Xiaoying Yang, Jianfeng Ren, Ruibin Bai, Xudong Jiang</i> | |
| A New Semi-Supervised Classification Method Using a Supervised Autoencoder for Biomedical Applications..... | 741 |
| <i>Cyprien Gille, Frederic Guyard, Michel Barlaud</i> | |
| A Frequency-Domain Recursive Least-Squares Adaptive Filtering Algorithm Based on a Kronecker Product Decomposition | 746 |
| <i>Hongsen He, Jingdong Chen, Jacob Benesty, Yi Yu</i> | |
| Asymptotic Bias and Variance of Kernel Ridge Regression..... | 751 |
| <i>Victor Solo</i> | |
| Rethinking Implicit Neural Representations for Vision Learners..... | 756 |
| <i>Yiran Song, Qianyu Zhou, Lizhuang Ma</i> | |
| Streaming Voice Conversion Via Intermediate Bottleneck Features and Non-Streaming Teacher Guidance..... | 761 |
| <i>Yuanzhe Chen, Ming Tu, Tang Li, Xin Li, Qiuqiang Kong, Jiaxin Li, Zhichao Wang, Qiao Tian, Yuping Wang, Yuxuan Wang</i> | |
| Learning Sparse auto-Encoders for Green AI Image Coding | 766 |
| <i>Cyprien Gille, Frédéric Guyard, Marc Antonini, Michel Barlaud</i> | |

| | |
|---|-----|
| Bias Identification with RankPix Saliency | 771 |
| <i>Salamata Konate, Léo Lebrat, Rodrigo Santa Cruz, Clinton Fookes, Andrew Bradley, Olivier Salvado</i> | |
| Efficient Multi-Scale Attention Module with Cross-Spatial Learning | 776 |
| <i>Daliang Ouyang, Su He, Guozhong Zhang, Mingzhu Luo, Huaiyong Guo, Jian Zhan, Zhijie Huang</i> | |
| Pseudo-Query Generation for Semi-Supervised Visual Grounding with Knowledge Distillation | 781 |
| <i>Jianglin Jin, Jiabo Ye, Xin Lin, Liang He</i> | |
| One-Shot Neural Band Selection for Spectral Recovery | 786 |
| <i>Hai-Miao Hu, Zhenbo Xu, Wenshuai Xu, You Song, Yitao Zhang, Liu Liu, Zhilin Han, Ajin Meng</i> | |
| Neural Band-To-Piano Score Arrangement with Stepless Difficulty Control | 791 |
| <i>Moyu Terao, Eita Nakamura, Kazuyoshi Yoshii</i> | |
| Interference Leakage Minimization in RIS-Assisted MIMO Interference Channels | 796 |
| <i>Ignacio Santamaria, Mohammad Soleymani, Eduard Jorswieck, Jesús Gutiérrez</i> | |
| HTNet: Human Topology Aware Network for 3D Human Pose Estimation | 801 |
| <i>Jialun Cai, Hong Liu, Runwei Ding, Wenhao Li, Jianbing Wu, Miaoju Ban</i> | |
| Designing Transformer Networks for Sparse Recovery of Sequential Data Using Deep Unfolding | 806 |
| <i>Brent De Weerd, Yonina C. Eldar, Nikos Deligiannis</i> | |
| Fast 3D Human Pose Estimation Using RF Signals | 811 |
| <i>Cong Yu, Dongheng Zhang, Zhi Wu, Chunyang Xie, Zhi Lu, Yang Hu, Yan Chen</i> | |
| Knowledge-Aware Graph Convolutional Network with Utterance-Specific Window Search for Emotion Recognition in Conversations | 816 |
| <i>Xiaotong Zhang, Peng He, Han Liu, Zhengxi Yin, Xinyue Liu, Xianchao Zhang</i> | |
| BadRes: Reveal the Backdoors Through Residual Connection..... | 821 |
| <i>Mingrui He, Tianyu Chen, Haoyi Zhou, Shanghang Zhang, Jianxin Li</i> | |
| Framewise Multiple Sound Source Localization and Counting Using Binaural Spatial Audio Signals | 826 |
| <i>Lei Wang, Zhibin Jiao, Qiyong Zhao, Jie Zhu, Yang Fu</i> | |
| Kernel Estimation and Deconvolution for Blind Image Super-Resolution | 831 |
| <i>Jiali Gong, Hongfan Gao, Jiahao Chao, Zhou Zhou, Zhengfeng Yang, Zhenbing Zeng</i> | |
| Dual Meta Calibration Mix for Improving Generalization in Meta-Learning | 836 |
| <i>Ze-Yu Mi, Yu-Bin Yang</i> | |
| Tensorized LSSVMS for Multitask Regression..... | 841 |
| <i>Jiani Liu, Qinghua Tao, Ce Zhu, Yipeng Liu, Johan A. K. Suykens</i> | |
| An Improved Optimal Transport Kernel Embedding Method with Gating Mechanism for Singing Voice Separation and Speaker Identification | 846 |
| <i>Weitao Yuan, Yuren Bian, Shengbei Wang, Masashi Unoki, Wenwu Wang</i> | |
| A Unitary Transform Based Generalized Approximate Message Passing | 851 |
| <i>Jiang Zhu, Xiangming Meng, Xupeng Lei, Qinghua Guo</i> | |

| | |
|---|-----|
| Infrared and Visible Image Fusion by Using Multi-Scale Transformation and Fractional-Order Gradient Information | 856 |
| <i>Shiwei Wu, Kang Zhang, Xia Yuan, Chunxia Zhao</i> | |
| Optimal Transport with a Diversified Memory Bank for Cross-Domain Speaker Verification..... | 861 |
| <i>Ruiteng Zhang, Jianguo Wei, Xugang Lu, Wenhuan Lu, Di Jin, Lin Zhang, Junhai Xu</i> | |
| BHE-DARTS: Bilevel Optimization Based on Hypergradient Estimation for Differentiable Architecture Search | 866 |
| <i>Zicheng Cai, Lei Chen, Hai-Lin Liu</i> | |
| Not All Classes Are Equal: Adaptively Focus-Aware Confidence for Semi-Supervised Object Detection | 871 |
| <i>Hui Zhu, Yongchun Lu, Hongyu Zhao, Guoqing Zhao, Xiaofang Zhao</i> | |
| Combining Dual-Tree Wavelet Analysis and Proximal Optimization for Anisotropic Scale-Free Texture Segmentation..... | 876 |
| <i>Leo Davy, Nelly Pustelnik, Patrice Abry</i> | |
| Interaction-Assisted Multi-Modal Representation Learning for Recommendation..... | 881 |
| <i>Hao Wu, Jiajie Wang, Zhonglin Zu</i> | |
| Transaudio: Towards the Transferable Adversarial Audio Attack Via Learning Contextualized Perturbations..... | 886 |
| <i>Gege Qi, Yuefeng Chen, Yao Zhu, Binyuan Hui, Xiaodan Li, Xiaofeng Mao, Rong Zhang, Hui Xue</i> | |
| Semi-Supervised Remote Sensing Image Change Detection Using Mean Teacher Model for Constructing Pseudo-Labels | 891 |
| <i>Zan Mao, Xinyu Tong, Ze Luo</i> | |
| SAR Image Despeckling with Residual-In-Residual Dense Generative Adversarial Network | 896 |
| <i>Yunpeng Bai, Yayuan Xiao, Xuan Hou, Ying Li, Changjing Shang, Qiang Shen</i> | |
| Int-GNN: A User Intention Aware Graph Neural Network for Session-Based Recommendation..... | 901 |
| <i>Guangning Xu, Jinyang Yang, Jinjin Guo, Zhichao Huang, Bowen Zhang</i> | |
| Co-Design for MIMO Radar and MIMO Communication Aided by Reconfigurable Intelligent Surface..... | 906 |
| <i>Da Li, Bo Tang, Lei Xue</i> | |
| CLMAE: A Lighter and Faster Masked Autoencoders | 911 |
| <i>Yiran Song, Lizhuang Ma</i> | |
| GCC-Speaker: Target Speaker Localization with Optimal Speaker-Dependent Weighting in Multi-Speaker Scenarios..... | 916 |
| <i>Guanjun Li, Wei Xue, Wenju Liu, Jiangyan Yi, Jianhua Tao</i> | |
| MID-Attribute Speaker Generation Using Optimal-Transport-Based Interpolation of Gaussian Mixture Models | 921 |
| <i>Aya Watanabe, Shinnosuke Takamichi, Yuki Saito, Detai Xin, Hiroshi Saruwatari</i> | |
| M ² -CTTS: End-to-End Multi-Scale Multi-Modal Conversational Text-to-Speech Synthesis..... | 926 |
| <i>Jinlong Xue, Yayue Deng, Fengping Wang, Ya Li, Yingming Gao, Jianhua Tao, Jianqing Sun, Jiaen Liang</i> | |

| | |
|--|------|
| Nord: Non-Matching Reference Based Relative Depth Estimation from Binaural Speech..... | 931 |
| <i>Pranay Manocha, Israel D. Gebru, Anurag Kumar, Dejan Markovic, Alexander Richard</i> | |
| ROI-Based Deep Image Compression with Swin Transformers | 936 |
| <i>Binglin Li, Jie Liang, Haisheng Fu, Jingning Han</i> | |
| Interpretability in the Context of Sequential Cost-Sensitive Feature Acquisition | 941 |
| <i>Yasitha Warahena Liyanage, Daphney–Stavroula Zois</i> | |
| Neural Transducer Training: Reduced Memory Consumption with Sample-Wise Computation | 946 |
| <i>Stefan Braun, Erik McDermott, Roger Hsiao</i> | |
| Multi-Resolution Location-Based Training for Multi-Channel Continuous Speech Separation | 951 |
| <i>Hassan Taherian, Deliang Wang</i> | |
| CyPMLI: WISL-Minimized Unimodular Sequence Design Via Power Method-Like Iterations | 956 |
| <i>Arian Eamaz, Farhang Yeganegi, Mojtaba Soltanalian</i> | |
| Weight-Sharing Supernet for Searching Specialized Acoustic Event Classification Networks Across Device Constraints | 961 |
| <i>Guan-Ting Lin, Qingming Tang, Chieh-Chi Kao, Viktor Rozgic, Chao Wang</i> | |
| Fast Convolution Algorithm for Real-Valued Finite Length Sequences..... | 966 |
| <i>Weiwei Wang, Victor Debrunner, Linda S. Debrunner</i> | |
| Do Prosody Transfer Models Transfer Prosody? | 971 |
| <i>Atli Thor Sigurgeirsson, Simon King</i> | |
| Do Coarser Units Benefit Cluster Prediction-Based Speech Pre-Training? | 976 |
| <i>Ali Elkahky, Wei-Ning Hsu, Paden Tomasello, Tu-Anh Nguyen, Robin Algayres, Yossi Adi, Jade Copet, Emmanuel Dupoux, Abdelrahman Mohamed</i> | |
| Learned Video Coding with Motion Compensation Mixture Model | 981 |
| <i>Khanh Quoc Dinh, Kwang Pyo Choi</i> | |
| HEiMDaL: Highly Efficient Method for Detection and Localization of Wake-Words | 986 |
| <i>Arnav Kundu, Mohammad Samragh, Minsik Cho, Priyanka Padmanabhan, Devang Naik</i> | |
| Spherical Vector Quantization for Spatial Direction Coding..... | 991 |
| <i>Stéphane Ragot, Adriana Vasilache</i> | |
| Sequential Datum–Wise Joint Feature Selection and Classification in the Presence of External Classifier..... | 996 |
| <i>Sachini Piyoni Ekanayake, Daphney–Stavroula Zois, Charalampos Chelmis</i> | |
| Multi-Scale Compositional Constraints for Representation Learning on Videos | 1001 |
| <i>Georgios Paraskevopoulos, Chandrashekhar Lavania, Lovish Chum, Shiva Sundaram</i> | |
| Active Learning of non-Semantic Speech Tasks with Pretrained Models | 1006 |
| <i>Harlin Lee, Aaqib Saeed, Andrea L. Bertozzi</i> | |
| A Magnetic Framelet-Based Convolutional Neural Network for Directed Graphs..... | 1011 |
| <i>Lequan Lin, Junbin Gao</i> | |
| Multi-Output RNN-T Joint Networks for Multi-Task Learning of ASR and Auxiliary Tasks..... | 1016 |
| <i>Weiran Wang, Ding Zhao, Shaojin Ding, Hao Zhang, Shuo-Yiin Chang, David Rybach, Tara N. Sainath, Yanzhang He, Ian McGraw, Shankar Kumar</i> | |

| | |
|--|------|
| Visual-Aware Text-to-Speech..... | 1021 |
| <i>Mohan Zhou, Yalong Bai, Wei Zhang, Ting Yao, Tiejun Zhao, Tao Mei</i> | |
| Output-Dependent Gaussian Process State-Space Model..... | 1026 |
| <i>Zhidi Lin, Lei Cheng, Feng Yin, Lexi Xu, Shuguang Cui</i> | |
| A Mathematical Model for Neuronal Activity and Brain Information Processing Capacity | 1031 |
| <i>Yu Zheng, David Zhu, Jian Ren, Taosheng Liu, Karl Friston, Tongtong Li</i> | |
| Attention Based Relation Network for Facial Action Units Recognition | 1036 |
| <i>Yao Wei, Haoxiang Wang, Mingze Sun, Jiawang Liu</i> | |
| Learning Hybrid Representations of Semantics and Distortion for Blind Image Quality Assessment..... | 1041 |
| <i>Xiaoqi Wang, Jian Xiong, Bo Li, Jinli Suo, Hao Gao</i> | |
| Affinity Learning with Blind-Spot Self-Supervision for Image Denoising..... | 1046 |
| <i>Yuhongze Zhou, Liguang Zhou, Issam Hadj Laradji, Tin Lun Lam, Yangsheng Xu</i> | |
| A Comprehensive Comparison of Projections in Omnidirectional Super-Resolution | 1051 |
| <i>Huicheng Pi, Senmao Tian, Ming Lu, Jiaming Liu, Yandong Guo, Shunli Zhang</i> | |
| APGP: Accuracy-Preserving Generative Perturbation for Defending Against Model Cloning Attacks..... | 1056 |
| <i>Anda Cheng, Jian Cheng</i> | |
| HDNet: Hierarchical Dynamic Network for Gait Recognition Using Millimeter-Wave Radar | 1061 |
| <i>Yanyan Huang, Yong Wang, Kun Shi, Chaojie Gu, Yu Fu, Cheng Zhuo, Zhiguo Shi</i> | |
| Liveness Score-Based Regression Neural Networks for Face Anti-Spoofing | 1066 |
| <i>Youngjun Kwak, Minyoung Jung, Hunjae Yoo, Jinho Shin, Changick Kim</i> | |
| Efficient Online Convolutional Dictionary Learning Using Approximate Sparse Components | 1071 |
| <i>Farshad G. Veshki, Sergiy A. Vorobyov</i> | |
| Laryngeal Leukoplakia Classification Via Dense Multiscale Feature Extraction in White Light Endoscopy Images..... | 1076 |
| <i>Zhenzhen You, Yan Yan, Zhenghao Shi, Minghua Zhao, Jing Yan, Haiqin Liu, Xinhong Hei, Xiaoyong Ren</i> | |
| String-Based Molecule Generation Via Multi-Decoder VAE..... | 1081 |
| <i>Kisoo Kwon, Kuhwan Jeong, Junghyun Park, Hwidong Na, Jinwoo Shin</i> | |
| Contrastive Domain Adaptation Via Delimitation Discriminator | 1086 |
| <i>Xing Wei, Bin Wen, Lei Chen, Yujie Liu, Chong Zhao, Yang Lu</i> | |
| Towards Robust Audio-Based Vehicle Detection Via Importance-Aware Audio-Visual Learning..... | 1091 |
| <i>Jung Uk Kim, Seong Tae Kim</i> | |
| Modulation-Based Center Alignment and Motion Mining for Spatial Temporal Action Detection | 1096 |
| <i>Weiji Zhao, Kefeng Huang, Chongyang Zhang</i> | |
| Randmasking Augment: A Simple and Randomized Data Augmentation for Acoustic Scene Classification | 1101 |
| <i>Jubum Han, Mateusz Matuszewski, Olaf Sikorski, Hosang Sung, Hoonyoung Cho</i> | |

| | |
|---|------|
| MHSCNET: A Multimodal Hierarchical Shot-Aware Convolutional Network for Video Summarization..... | 1106 |
| <i>Wujiang Xu, Runzhong Wang, Xiaobo Guo, Shaoshuai Li, Qiongxu Ma, Yunan Zhao, Sheng Guo, Zhenfeng Zhu, Junchi Yan</i> | |
| Defending Against Universal Patch Attacks by Restricting Token Attention in Vision Transformers | 1111 |
| <i>Hongwei Yu, Jiansheng Chen, Huimin Ma, Cheng Yu, Xinlong Ding</i> | |
| Two-Stage Video De-Raining with Spatio-Temporal Fusion and Illumination-Invariant Detail Preservation | 1116 |
| <i>Yufeng Tan, Youjun Xiang, Lei Cai, Pengcheng Wang, Ying Zhang, Yuli Fu</i> | |
| AugTarget Data Augmentation for Infrared Small Target Detection | 1121 |
| <i>Shengjia Chen, Jiewen Zhu, Luping Ji, Hongjun Pan, Yuhao Xu</i> | |
| MMATR: A Lightweight Approach for Multimodal Sentiment Analysis Based on Tensor Methods..... | 1126 |
| <i>Panagiotis Koromilas, Mihalis A. Nicolaou, Theodoros Giannakopoulos, Yannis Panagakis</i> | |
| Real-Time Speech Enhancement with Dynamic Attention Span..... | 1131 |
| <i>Chengyu Zheng, Yuan Zhou, Xiulian Peng, Yuan Zhang, Yan Lu</i> | |
| DQFORMER: Dynamic Query Transformer for Lane Detection..... | 1136 |
| <i>Hao Yang, Shuyuan Lin, Runqing Jiang, Yang Lu, Hanzi Wang</i> | |
| CROSSSPEECH: Speaker-Independent Acoustic Representation for Cross-Lingual Speech Synthesis..... | 1141 |
| <i>Ji-Hoon Kim, Hong-Sun Yang, Yoon-Cheol Ju, Il-Hwan Kim, Byeong-Yeol Kim</i> | |
| Focusing on Targets for Improving Weakly Supervised Visual Grounding | 1146 |
| <i>Viet-Quoc Pham, Nao Mishima</i> | |
| Lightweight Portrait Segmentation Via Edge-Optimized Attention | 1151 |
| <i>Xinyue Zhang, Guodong Wang, Lijuan Yang, Chenglizhao Chen</i> | |
| Leveraging Sparsity with Spiking Recurrent Neural Networks for Energy-Efficient Keyword Spotting | 1156 |
| <i>Manon Dampffoffer, Thomas Mesquida, Emmanuel Hardy, Alexandre Valentian, Lorena Anghel</i> | |
| TSPTQ-ViT: Two-Scaled Post-Training Quantization for Vision Transformer..... | 1161 |
| <i>Yu-Shan Tai, Ming-Guang Lin, An-Yeu Andy Wu</i> | |
| Explanations for Automatic Speech Recognition | 1166 |
| <i>Xiaoliang Wu, Peter Bell, Ajitha Rajan</i> | |
| Contrast-PLC: Contrastive Learning for Packet Loss Concealment | 1171 |
| <i>Huaying Xue, Xiulian Peng, Yan Lu</i> | |
| Efficient Large-Scale Audio Tagging Via Transformer-to-CNN Knowledge Distillation | 1176 |
| <i>Florian Schmid, Khaled Koutini, Gerhard Widmer</i> | |
| Masking Speech Contents by Random Splicing: is Emotional Expression Preserved? | 1181 |
| <i>Felix Burkhardt, Anna Derington, Matthias Kahlau, Klaus Scherer, Florian Eyben, Björn Schuller</i> | |
| On Out-Of-Distribution Detection for Audio with Deep Nearest Neighbors | 1186 |
| <i>Zaharah Bukhsh, Aaqib Saeed</i> | |

| | |
|--|------|
| Improving Scheduled Sampling for Neural Transducer-Based ASR..... | 1191 |
| <i>Takafumi Moriya, Takanori Ashihara, Hiroshi Sato, Kohei Matsuura, Tomohiro Tanaka, Ryo Masumura</i> | |
| Estimating Inharmonic Signals with Optimal Transport Priors..... | 1196 |
| <i>Filip Elvander</i> | |
| Gluformer: Transformer-Based Personalized Glucose Forecasting with Uncertainty Quantification..... | 1201 |
| <i>Renat Sergazinov, Mohammadreza Armandpour, Irina Gaynanova</i> | |
| Global HRTF Interpolation Via Learned Affine Transformation of Hyper-Conditioned Features..... | 1206 |
| <i>Jin Woo Lee, Sungho Lee, Kyogu Lee</i> | |
| Efficient Compressed Video Action Recognition Via Late Fusion with a Single Network..... | 1211 |
| <i>Hayato Terao, Wataru Noguchi, Hiroyuki Iizuka, Masahito Yamamoto</i> | |
| GOP-Based Latent Refinement for Learned Video Coding..... | 1216 |
| <i>Mohsen Abdoli, Gordon Clare, Félix Henry</i> | |
| Effective Training of RNN Transducer Models on Diverse Sources of Speech and Text Data..... | 1221 |
| <i>Takashi Fukuda, Samuel Thomas</i> | |
| On Minimal Variations for Unsupervised Representation Learning..... | 1226 |
| <i>Vivien Cabannes, Alberto Bietti, Randall Balestrieri</i> | |
| Möbius Total Variation for Directed Acyclic Graphs..... | 1231 |
| <i>Vedran Mihal, Markus Püschel</i> | |
| Spoofed Training Data for Speech Spoofing Countermeasure Can Be Efficiently Created Using Neural Vocoders..... | 1236 |
| <i>Xin Wang, Junichi Yamagishi</i> | |
| Multi-Speaker Data Augmentation for Improved End-to-end Automatic Speech Recognition..... | 1241 |
| <i>Samuel Thomas, Hong-Kwang J. Kuo, George Saon, Brian Kingsbury</i> | |
| Alignment Entropy Regularization..... | 1246 |
| <i>Ehsan Variani, Ke Wu, David Rybach, Cyril Allauzen, Michael Riley</i> | |
| A Targeted Sampling Strategy for Compressive Cryo Focused Ion Beam Scanning Electron Microscopy..... | 1251 |
| <i>D. Nicholls, J. Wells, A. W. Robinson, A. Moshtaghpour, M. Kobylenska, R. A. Fleck, A. I. Kirkland, N. D. Browning</i> | |
| Diagonal State Space Augmented Transformers for Speech Recognition..... | 1256 |
| <i>George Saon, Ankit Gupta, Xiaodong Cui</i> | |
| Healthcall Corpus and Transformer Embeddings from Healthcare Customer-Agent Conversations..... | 1261 |
| <i>Nikola Lackovic, Claude Montacié, Cédric Lequilliec, Marie-José Caraty</i> | |
| Gaitcotr: Improved Spatial-Temporal Representation for Gait Recognition with a Hybrid Convolution-Transformer Framework..... | 1266 |
| <i>Jingqi Li, Yuzhen Zhang, Hongming Shan, Junping Zhang</i> | |
| Named Entity Detection and Injection for Direct Speech Translation..... | 1271 |
| <i>Marco Gaido, Yun Tang, Ilia Kulikov, Rongqing Huang, Hongyu Gong, Hirofumi Inaguma</i> | |

| | |
|---|------|
| C2KD: Cross-Lingual Cross-Modal Knowledge Distillation for Multilingual Text-Video Retrieval | 1276 |
| <i>Andrew Rouditchenko, Yung-Sung Chuang, Nina Shvetsova, Samuel Thomas, Rogerio Feris, Brian Kingsbury, Leonid Karlinsky, David Harwath, Hilde Kuehne, James Glass</i> | |
| Amicable Aid: Perturbing Images to Improve Classification Performance..... | 1281 |
| <i>Juyeop Kim, Jun-Ho Choi, Soobeom Jang, Jong-Seok Lee</i> | |
| Mutual Information Based Reweighting for Precipitation Nowcasting..... | 1286 |
| <i>Yuan Cao, Danchen Zhang, Xin Zheng, Hongming Shan, Junping Zhang</i> | |
| Structured State Space Decoder for Speech Recognition and Synthesis | 1291 |
| <i>Koichi Miyazaki, Masato Murata, Tomoki Koriyama</i> | |
| Implicit Vehicle Positioning with Cooperative Lidar Sensing..... | 1296 |
| <i>Luca Barbieri, Bernardo Camajori Tedeschini, Mattia Brambilla, Monica Nicoli</i> | |
| Channel-Driven Decentralized Bayesian Federated Learning for Trustworthy Decision Making in D2D Networks..... | 1301 |
| <i>Luca Barbieri, Osvaldo Simeone, Monica Nicoli</i> | |
| Multilevel FISTA for Image Restoration | 1306 |
| <i>Guillaume Lauga, Elisa Riccietti, Nelly Pustelnik, Paulo Gonçalves</i> | |
| TrimTail: Low-Latency Streaming ASR with Simple but Effective Spectrogram-Level Length Penalty | 1311 |
| <i>Xingchen Song, Di Wu, Zhiyong Wu, Binbin Zhang, Yuekai Zhang, Zhendong Peng, Wenpeng Li, Fuping Pan, Changbao Zhu</i> | |
| Cross-Device Federated Learning for Mobile Health Diagnostics: A First Study on COVID-19 Detection | 1316 |
| <i>Tong Xia, Jing Han, Abhirup Ghosh, Cecilia Mascolo</i> | |
| Masked Spectrogram Prediction for Self-Supervised Audio Pre-Training..... | 1321 |
| <i>Dading Chong, Helin Wang, Peilin Zhou, Qingcheng Zeng</i> | |
| Benchmarking Cross-Domain Face Recognition with Avatars, Caricatures and Sketches..... | 1326 |
| <i>A. Foroughi, C. Rathgeb, M. Ibsen, C. Busch</i> | |
| Unbiased Unsupervised Stimulus Reconstruction for EEG-Based Auditory Attention Decoding | 1331 |
| <i>Nicolas Heintz, Simon Geirnaert, Tom Francart, Alexander Bertrand</i> | |
| Projected Hierarchical ALS for Generalized Boolean Matrix Factorization | 1335 |
| <i>Rodrigo Cabral Farias, Sebastian Miron</i> | |
| Using Received Power in Microphone Arrays to Estimate Direction of Arrival..... | 1340 |
| <i>Gustav Zetterqvist, Fredrik Gustafsson, Gustaf Hendeby</i> | |
| Energy Regularized RNNs for Solving Non-Stationary Bandit Problems..... | 1345 |
| <i>Michael Rotman, Lior Wolf</i> | |
| Line Segment Matching Based on Intersection-Enhanced Point Correspondences | 1350 |
| <i>Zhiyu Liu, Baojiang Zhong</i> | |
| Text is All You Need: Personalizing ASR Models Using Controllable Speech Synthesis..... | 1355 |
| <i>Karren Yang, Ting-Yao Hu, Jen-Hao Rick Chang, Hema Swetha Koppula, Oncel Tuzel</i> | |

| | |
|--|------|
| On-The-Fly Text Retrieval for End-To-end ASR Adaptation | 1360 |
| <i>Bolaji Yusuf, Aditya Gourav, Ankur Gandhe, Ivan Bulyko</i> | |
| Deep Learning-Based Stereo Camera Multi-Video Synchronization | 1365 |
| <i>Nicolas Boizard, Kevin El Haddad, Thierry Ravet, François Cresson, Thierry Dutoit</i> | |
| Provable Computational and Statistical Guarantees for Efficient Learning of Continuous-Action Graphical Games | 1370 |
| <i>Adarsh Barik, Jean Honorio</i> | |
| Instance-Aware Hierarchical Structured Policy for Prompt Learning in Vision-Language Models | 1375 |
| <i>Xun Wu, Guolong Wang, Zhaoyuan Liu, Xuan Dang, Zheng Qin</i> | |
| Analyzing Acoustic Word Embeddings from Pre-Trained Self-Supervised Speech Models | 1380 |
| <i>Ramon Sanabria, Hao Tang, Sharon Goldwater</i> | |
| Probabilistic Back-Ends for Online Speaker Recognition and Clustering | 1385 |
| <i>Alexey Sholokhov, Nikita Kuzmin, Kong Aik Lee, Eng Siong Chng</i> | |
| A Unified Uncertainty-Aware Exploration: Combining Epistemic and Aleatory Uncertainty | 1390 |
| <i>Parvin Malekzadeh, Ming Hou, Konstantinos N. Plataniotis</i> | |
| Independent Vector Analysis with Multivariate Gaussian Model: A Scalable Method by Multilinear Regression | 1395 |
| <i>Ben Gabrielson, Mingyu Sun, M. A. B. S. Akhonda, Vince D. Calhoun, Tülay Adalı</i> | |
| Toward a Multimodal Approach for Disfluency Detection and Categorization | 1400 |
| <i>Amrit Romana, Kazuhito Koishida</i> | |
| Towards Diverse and Coherent Augmentation for Time-Series Forecasting | 1405 |
| <i>Xiyuan Zhang, Ranak Roy Chowdhury, Jingbo Shang, Rajesh Gupta, Dezhi Hong</i> | |
| On Using the UA-Speech and Torgo Databases to Validate Automatic Dysarthric Speech Classification Approaches | 1410 |
| <i>Guilherme Schu, Parvaneh Janbakhshi, Ina Kodrasi</i> | |
| Detection of Real-Time Deepfakes in Video Conferencing with Active Probing and Corneal Reflection | 1415 |
| <i>Hui Guo, Xin Wang, Siwei Lyu</i> | |
| Sparse Mixture Once-For-all Adversarial Training for Efficient In-situ Trade-off Between Accuracy and Robustness of DNNs | 1420 |
| <i>Souvik Kundu, Sairam Sundaresan, Sharath Nittur Sridhar, Shunlin Lu, Han Tang, Peter A. Beerel</i> | |
| Context-Aware End-to-end ASR Using Self-Attentive Embedding and Tensor Fusion | 1425 |
| <i>Shuo-Yiin Chang, Chao Zhang, Tara N. Sainath, Bo Li, Trevor Strohman</i> | |
| EH-Enabled Distributed Detection Over Temporally Correlated Markovian MIMO Channels | 1430 |
| <i>Ghazaleh Ardeshiri, Azadeh Vosoughi</i> | |
| Cross Modality Knowledge Distillation for Robust Pedestrian Detection in Low Light and Adverse Weather Conditions | 1435 |
| <i>Mazin Hnewa, Alireza Rahimpour, Justin Miller, Devesh Upadhyay, Hayder Radha</i> | |
| Direction-Of-Arrival Estimation Using Gaussian Process Interpolation | 1440 |
| <i>Ishan D. Khurjekar, Peter Gerstoft, Christoph F. Mecklenbräuer, Zoi-Heleni Michalopoulou</i> | |

| | |
|--|------|
| SQId: Measuring Speech Naturalness in Many Languages..... | 1445 |
| <i>Thibault Sellam, Ankur Bapna, Joshua Camp, Diana Mackinnon, Ankur P. Parikh, Jason Riesa</i> | |
| Linear Microphone Array Parallel to the Driving Direction for in-Car Speech Enhancement..... | 1450 |
| <i>Masanori Tsujikawa, Akihiko Sugiyama, Ken Hanazawa, Yoshinobu Kajikawa</i> | |
| Facial Texture Perceiver: Towards High-Fidelity Facial Texture Recovery with Input-Level Inductive Biased Perceiver IO..... | 1455 |
| <i>Seungeun Lee</i> | |
| Adaptive Noise Canceller Algorithm with SNR-Based Stepsize and Data-Dependent Averaging..... | 1460 |
| <i>Akihiko Sugiyama</i> | |
| Exploring Instance Relation for Decentralized Multi-Source Domain Adaptation | 1465 |
| <i>Yikang Wei, Yahong Han</i> | |
| Contrastive Learning with Dialogue Attributes for Neural Dialogue Generation | 1470 |
| <i>Jie Tan, Hengyi Cai, Hongshen Chen, Hong Cheng, Helen Meng, Zhuoye Ding</i> | |
| Open-Set Automatic Target Recognition..... | 1475 |
| <i>Bardia Safaei, Vibashan Vs, Celso M. De Melo, Shuowen Hu, Vishal M. Patel</i> | |
| Training Robust Spiking Neural Networks on Neuromorphic Data with Spatiotemporal Fragments..... | 1480 |
| <i>Haibo Shen, Yihao Luo, Xiang Cao, Liangqi Zhang, Juyi Xiao, Tianjiang Wang</i> | |
| On Bidirectional Preestimates and Their Application to Identification of Fast Time-Varying Systems..... | 1485 |
| <i>Maciej Niedzwiecki, Artur Gancza, Lu Shen, Yuriy Zakharov</i> | |
| Matrix Resolvent Eigenembeddings for Dynamic Graphs | 1490 |
| <i>Vasileios Kalantzis, Panagiotis A. Traganitis</i> | |
| Class-Aware Shared Gaussian Process Dynamic Model | 1495 |
| <i>Ryosuke Sawata, Takahiro Ogawa, Miki Haseyama</i> | |
| Hierarchical Multi-Agent Reinforcement Learning with Intrinsic Reward Rectification | 1500 |
| <i>Zhihao Liu, Zhiwei Xu, Guoliang Fan</i> | |
| On Multiple-Input/Binaural-Output Antiphase Speaker Signal Extraction..... | 1505 |
| <i>Xianrui Wang, Ningning Pan, Jacob Benesty, Jingdong Chen</i> | |
| Speech Dereverberation with a Reverberation Time Shortening Target..... | 1510 |
| <i>Rui Zhou, Wenye Zhu, Xiaofei Li</i> | |
| Spatially Informed Independent Vector Analysis for Source Extraction Based on the Convolutional Transfer Function Model..... | 1515 |
| <i>Xianrui Wang, Andreas Brendel, Gongping Huang, Yichen Yang, Walter Kellermann, Jingdong Chen</i> | |
| Dewarping Documents Using C2 Continuous Boundary Estimation..... | 1520 |
| <i>Prasenjit Mondal, Ayush Pant, Sachin Soni</i> | |
| Dynamic Selection of P-Norm in Linear Adaptive Filtering Via Online Kernel-based Reinforcement Learning..... | 1525 |
| <i>Minh Vu, Yuki Akiyama, Konstantinos Slavakis</i> | |

| | |
|---|------|
| LiQuiD-MIMO Radar: Distributed MIMO Radar with Low-Bit Quantization..... | 1530 |
| <i>Yikun Xiang, Feng Xi, Shengyao Chen</i> | |
| IAST: Instance Association Relying on Spatio-Temporal Features for Video Instance Segmentation..... | 1535 |
| <i>Junhao Chen, Sheng Liu, Ruixiang Chen, Bingnan Guo, Feng Zhang</i> | |
| UPGLADE: Unplugged Plug-And-Play Audio Declipper Based on Consensus Equilibrium of DNN and Sparse Optimization..... | 1540 |
| <i>Tomoro Tanaka, Kohei Yatabe, Yasuhiro Oikawa</i> | |
| Streaming Stroke Classification of Online Handwriting | 1545 |
| <i>Jing-Yu Liu, Yan-Ming Zhang, Fei Yin, Cheng-Lin Liu</i> | |
| This Changes to That: Combining Causal and Non-Causal Explanations to Generate Disease Progression in Capsule Endoscopy..... | 1550 |
| <i>Anuja Vats, Ahmed Mohammed, Marius Pedersen, Nirmalie Wiratunga</i> | |
| Masked Modeling Duo: Learning Representations by Encouraging Both Networks to Model the Input | 1555 |
| <i>Daisuke Niizumi, Daiki Takeuchi, Yasunori Ohishi, Noboru Harada, Kunio Kashino</i> | |
| CANet: Curved Guide Line Network with Adaptive Decoder for Lane Detection | 1560 |
| <i>Zhongyu Yang, Chen Shen, Wei Shao, Tengfei Xing, Runbo Hu, Pengfei Xu, Hua Chai, Ruini Xue</i> | |
| Quantum Variational Bayes on Manifolds..... | 1565 |
| <i>Anna Lopatnikova, Minh-Ngoc Tran</i> | |
| Robust FIR Filters for Wireless Low-Frequency Sound Zones | 1570 |
| <i>Mo Zhou, Martin Bo Møller, Christian Sejer Pedersen, Jan Østergaard</i> | |
| UCONV-Conformer: High Reduction of Input Sequence Length for End-to-End Speech Recognition | 1575 |
| <i>Andrei Andrusenko, Rauf Nasretdinov, Aleksei Romanenko</i> | |
| SFEMGN: Image Denoising with Shallow Feature Enhancement Network and Multi-Scale ConvGRU..... | 1580 |
| <i>Qidong Wang, Lili Guo, Shifei Ding, Jian Zhang, Xiao Xu</i> | |
| Time-Aware Multiway Adaptive Fusion Network for Temporal Knowledge Graph Question Answering | 1585 |
| <i>Yonghao Liu, Di Liang, Fang Fang, Sirui Wang, Wei Wu, Rui Jiang</i> | |
| FAPM: Fast Adaptive Patch Memory for Real-Time Industrial Anomaly Detection | 1590 |
| <i>Donghyeong Kim, Chaewon Park, Suhwan Cho, Sangyoun Lee</i> | |
| CRFAST: Clip-Based Reference-Guided Facial Image Semantic Transfer..... | 1595 |
| <i>Ailin Li, Lei Zhao, Zhiwen Zuo, Zhizhong Wang, Wei Xing, Dongming Lu</i> | |
| Introducing Topography in Convolutional Neural Networks | 1600 |
| <i>Maxime Poli, Emmanuel Dupoux, Rachid Riad</i> | |
| Structure-Aware Sparse Bayesian Learning-Based Channel Estimation for Intelligent Reflecting Surface-Aided MIMO | 1605 |
| <i>Yanbin He, Geethu Joseph</i> | |

| | |
|---|------|
| Aleatoric Uncertainty Estimation of Overnight Sleep Statistics Through Posterior Sampling Using Conditional Normalizing Flows | 1610 |
| <i>Hans Van Gorp, Merel M. Van Gilst, Pedro Fonseca, Sebastiaan Overeem, Ruud J. G. Van Sloun</i> | |
| Hardware Friendly Spline Sketched Lidar | 1615 |
| <i>Michael P. Sheehan, Julián Tachella, Mike E. Davies</i> | |
| Explicit Ziv-Zakai Bound for Multiple Sources Doa Estimation | 1620 |
| <i>Zongyu Zhang, Yujie Gu, Zhiguo Shi</i> | |
| Adaptive Large Margin Fine-Tuning for Robust Speaker Verification | 1625 |
| <i>Leying Zhang, Zhengyang Chen, Yanmin Qian</i> | |
| Radio-Astronomy Imaging and Interference Excision Using Tensor Decomposition and Canonical Correlation Analysis | 1630 |
| <i>Mikael Sørensen, Nicholas D. Sidiropoulos</i> | |
| Hybridformer: Improving Squeezeformer with Hybrid Attention and NSR Mechanism | 1635 |
| <i>Yuguang Yang, Yu Pan, Jingjing Yin, Jiangyu Han, Lei Ma, Heng Lu</i> | |
| EGAN: A Neural Excitation Generation Model Based on Generative Adversarial Networks with Harmonics and Noise Input | 1640 |
| <i>Yen-Ting Lin, Chen-Yu Chiang</i> | |
| Training Set Cleansing of Backdoor Poisoning by Self-Supervised Representation Learning | 1645 |
| <i>Hang Wang, Sahar Karami, Ousmane Dia, Hippolyt Ritter, Ehsan Emamjomeh-Zadeh, Jiahui Chen, Zhen Xiang, David J. Miller, George Kesidis</i> | |
| HIFI++: A Unified Framework for Bandwidth Extension and Speech Enhancement | 1650 |
| <i>Pavel Andreev, Aibek Alanov, Oleg Ivanov, Dmitry Vetrov</i> | |
| SUVR: A Search-Based Approach to Unsupervised Visual Representation Learning | 1655 |
| <i>Yi-Zhan Xu, Chih-Yao Chen, Cheng-Te Li</i> | |
| Continuous Interaction with a Smart Speaker Via Low-Dimensional Embeddings of Dynamic Hand Pose | 1660 |
| <i>Songpei Xu, Chaitanya Kaul, Xuri Ge, Roderick Murray-Smith</i> | |
| Making Synchrosqueezing Locally Adaptive in the Time-Frequency Plane | 1665 |
| <i>Marcelo A. Colominas, Sylvain Meignen</i> | |
| Refined Pseudo Labeling for Source-Free Domain Adaptive Object Detection..... | 1670 |
| <i>Siqi Zhang, Lu Zhang, Zhiyong Liu</i> | |
| Learning Quantum Entanglement Distillation with Noisy Classical Communications | 1675 |
| <i>Hari Hara Suthan Chittoor, Osvaldo Simeone</i> | |
| Hybrid Ris-Assisted Interference Mitigation for Spectrum Sharing | 1680 |
| <i>Fangzhou Wang, A. Lee Swindlehurst</i> | |
| Disentangled Training with Adversarial Examples for Robust Small-Footprint Keyword Spotting | 1685 |
| <i>Zhenyu Wang, Li Wan, Biqiao Zhang, Yiteng Huang, Shang-Wen Li, Ming Sun, Xin Lei, Zhaojun Yang</i> | |
| Play it Back: Iterative Attention for Audio Recognition..... | 1690 |
| <i>Alexandros Stergiou, Dima Damen</i> | |

| | |
|---|------|
| Anchored Speech Recognition with Neural Transducers | 1695 |
| <i>Desh Raj, Junteng Jia, Jay Mahadeokar, Chunyang Wu, Niko Moritz, Xiaohui Zhang, Ozlem Kalinli</i> | |
| Can Knowledge of End-To-End Text-To-Speech Models Improve Neural Midi-To-Audio Synthesis Systems? | 1700 |
| <i>Xuan Shi, Erica Cooper, Xin Wang, Junichi Yamagishi, Shrikanth Narayanan</i> | |
| Perspective Projection-Based 3D CT Reconstruction from Biplanar X-Rays | 1705 |
| <i>Daeun Kyung, Kyungmin Jo, Jaegul Choo, Joonseok Lee, Edward Choi</i> | |
| Shadow Removal of Text Document Images Using Background Estimation and Adaptive Text Enhancement | 1710 |
| <i>Wenjie Liu, Bingshu Wang, Jiangbin Zheng, Wenmin Wang</i> | |
| A Quantum Kernel Learning Approach to Acoustic Modeling for Spoken Command Recognition | 1715 |
| <i>Chao-Han Huck Yang, Bo Li, Yu Zhang, Nanxin Chen, Tara N. Sainath, Sabato Marco Siniscalchi, Chin-Hui Lee</i> | |
| Speech Emotion Recognition Via Two-Stream Pooling Attention with Discriminative Channel Weighting | 1720 |
| <i>Ke Liu, Dekui Wang, Dongya Wu, Jun Feng</i> | |
| Identifying Entrainment in Task-Oriented Conversations | 1725 |
| <i>Run Chen, Seokhwan Kim, Alexandros Papangelis, Julia Hirschberg, Yang Liu, Dilek Hakkani-Tür</i> | |
| Spatial Active Noise Control Method Based on Sound Field Interpolation from Reference Microphone Signals..... | 1730 |
| <i>Kazuyuki Arikawa, Shoichi Koyama, Hiroshi Saruwatari</i> | |
| A Perturbation-Based Policy Distillation Framework with Generative Adversarial Nets | 1735 |
| <i>Lihua Zhang, Quan Liu, Xiongzen Zhang, Yapeng Xu</i> | |
| Domain Generalized Fundus Image Segmentation Via Dual-Level Mixing | 1740 |
| <i>Xin Luo, Wei Chen, Chen Li, Bin Zhou, Yusong Tan</i> | |
| A Holistic Cascade System, Benchmark, and Human Evaluation Protocol for Expressive Speech-To-Speech Translation | 1745 |
| <i>Wen-Chin Huang, Benjamin Peloquin, Justine Kao, Changhan Wang, Hongyu Gong, Elizabeth Salesky, Yossi Adi, Ann Lee, Peng-Jen Chen</i> | |
| Class-Incremental Learning on Multivariate Time Series Via Shape-Aligned Temporal Distillation | 1750 |
| <i>Zhongzheng Qiao, Minghui Hu, Xudong Jiang, Ponnuthurai Nagaratnam Suganthan, Ramasamy Savitha</i> | |
| Distributed Online Learning with Adversarial Participants in an Adversarial Environment | 1755 |
| <i>Xingrong Dong, Zhaoxian Wu, Qing Ling, Zhi Tian</i> | |
| A Geometric Surrogate for Simulation Calibration | 1760 |
| <i>Lincon S. Souza, Bojan Batalo, Keisuke Yamazaki</i> | |
| Role of Bias Terms in Dot-Product Attention..... | 1765 |
| <i>Mahdi Namazifar, Devamanyu Hazarika, Dilek Hakkani-Tür</i> | |
| Jazznet: A Dataset of Fundamental Piano Patterns for Music Audio Machine Learning Research..... | 1770 |
| <i>Tosiron Adegbija</i> | |

| | |
|---|------|
| Multi-Functional Reconfigurable Intelligent Surface..... | 1775 |
| <i>Wen Wang, Wanli Ni, Hui Tian, Yonina C. Eldar</i> | |
| Heuristic Masking for Text Representation Pretraining | 1780 |
| <i>Yimeng Zhuang</i> | |
| Supervised Hierarchical Clustering Using Graph Neural Networks for Speaker Diarization | 1785 |
| <i>Prachi Singh, Amrit Kaul, Sriram Ganapathy</i> | |
| Kernel Interpolation of Acoustic Transfer Functions with Adaptive Kernel for Directed and Residual Reverberations | 1790 |
| <i>Juliano G. C. Ribeiro, Shoichi Koyama, Hiroshi Saruwatari</i> | |
| SLBERT: A Novel Pre-Training Framework for Joint Speech and Language Modeling | 1795 |
| <i>Onkar Susladkar, Prajwal Gatti, Santosh Kumar Yadav</i> | |
| CANDY: Category-Kernelized Dynamic Convolution for Instance Segmentation..... | 1800 |
| <i>Yao Lu, Zhiyi Chen, Zehui Chen, Jie Hu, Liujuan Cao, Shengchuan Zhang</i> | |
| A Study on Bias and Fairness in Deep Speaker Recognition | 1805 |
| <i>Amirhossein Hajavi, Ali Etemad</i> | |
| Thermal Infrared Image inpainting Via Edge-Aware Guidance | 1810 |
| <i>Zeyu Wang, Haibin Shen, Changyou Men, Quan Sun, Kejie Huang</i> | |
| On Word Error Rate Definitions and Their Efficient Computation for Multi-Speaker Speech Recognition Systems | 1815 |
| <i>Thilo Von Neumann, Christoph Boeddeker, Keisuke Kinoshita, Marc Delcroix, Reinhold Haeb-Umbach</i> | |
| Deep Low Light Image Enhancement Via Multi-Scale Recursive Feature Enhancement and Curve Adjustment | 1820 |
| <i>Haiyan Jin, Dawei Wei, Haonan Su</i> | |
| Log-Can: Local-Global Class-Aware Network for Semantic Segmentation of Remote Sensing Images | 1825 |
| <i>Xiaowen Ma, Mengting Ma, Chenlu Hu, Zhiyuan Song, Ziyang Zhao, Tian Feng, Wei Zhang</i> | |
| TAMformer: Multi-Modal Transformer with Learned Attention Mask for Early Intent Prediction..... | 1830 |
| <i>Nada Osman, Guglielmo Camporese, Lamberto Ballan</i> | |
| Long-Tailed Recognition with Causal Invariant Transformation | 1835 |
| <i>Yahong Zhang, Sheng Shi, Chenchen Fan, Yixin Wang, Wenli Ouyang, Weifan, Jianpin Fan</i> | |
| “Prediction of Sleepiness Ratings from Voice by Man and Machine”: A Perceptual Experiment Replication Study | 1840 |
| <i>Vincent P. Martin, Aymeric Ferron, Jean-Luc Rouas, Pierre Philip</i> | |
| Online Residual-Based Key Frame Sampling with Self-Coach Mechanism and Adaptive Multi-Level Feature Fusion | 1845 |
| <i>Rui Zhang, Yang Hua, Tao Song, Zhengui Xue, Ruhui Ma, Haibing Guan</i> | |
| Unsupervised Word Segmentation Using Temporal Gradient Pseudo-Labels..... | 1850 |
| <i>Tzeviya Sylvia Fuchs, Yedid Hoshen</i> | |

| | |
|---|------|
| Exploiting PRNU and Linear Patterns in Forensic Camera Attribution Under Complex Lens Distortion Correction..... | 1855 |
| <i>Andrea Montibeller, Fernando Pérez-González</i> | |
| Gradient Remedy for Multi-Task Learning in End-to-End Noise-Robust Speech Recognition..... | 1860 |
| <i>Yuchen Hu, Chen Chen, Ruizhe Li, Qiushi Zhu, Eng Siong Chng</i> | |
| Region-Aware Transformer with Asymmetric Loss in Multi-Label Classification | 1865 |
| <i>Lei Zhang, Jie Liu, Yanqi Bao, Jie Wang</i> | |
| A Multi-Stage Triple-Path Method for Speech Separation in Noisy and Reverberant Environments | 1870 |
| <i>Zhaoxi Mu, Xinyu Yang, Xiangyuan Yang, Wenjing Zhu</i> | |
| Scaling Law Analysis for Covariance Based Activity Detection in Cooperative Multi-Cell Massive MIMO..... | 1875 |
| <i>Ziyue Wang, Ya-Feng Liu, Zhaorui Wang, Wei Yu</i> | |
| Blind Polynomial Regression..... | 1880 |
| <i>Alberto Natali, Geert Leus</i> | |
| Empathetic Response Generation Via Emotion Cause Transition Graph | 1885 |
| <i>Yushan Qian, Bo Wang, Ting-En Lin, Yinhe Zheng, Ying Zhu, Dongming Zhao, Yuexian Hou, Yuchuan Wu, Yongbin Li</i> | |
| Semantic-Aware Gated Fusion Network for Interactive Colorization..... | 1890 |
| <i>Jie Zhang, Yi Xiao, Yan Zhenga, Zhenni Wang, Chi-Sing Leung</i> | |
| CDHD: Contrastive Dreamer for Hint Distillation..... | 1895 |
| <i>Le Yu, Tongyan Hua, Wenming Yang, Peng Ye, Qingmin Liao</i> | |
| Masked Token Similarity Transfer for Compressing Transformer-Based ASR Models..... | 1900 |
| <i>Euntae Choi, Youshin Lim, Byeong-Yeol Kim, Hyung Yong Kim, Hanbin Lee, Yunkyu Lim, Seung Woo Yu, Sungjoo Yoo</i> | |
| Which Country is This Picture From? New Data and Methods for Dnn-Based Country Recognition..... | 1905 |
| <i>Omran Alamayreh, Giovanna Maria Dimitri, Jun Wang, Benedetta Tondi, Mauro Barni</i> | |
| Improving Speech Enhancement Via Event-Based Query | 1910 |
| <i>Yifei Xin, Xiulian Peng, Yan Lu</i> | |
| Capturing Cross-Scale Disparity for Stereo Image Super-Resolution..... | 1915 |
| <i>Kun He, Changyu Li, Dongyang Zhang, Jie Shao</i> | |
| Byzantine-Robust and Communication-Efficient Personalized Federated Learning..... | 1920 |
| <i>Xuechao He, Jiaojiao Zhang, Qing Ling</i> | |
| Better Together: Dialogue Separation and Voice Activity Detection for Audio Personalization in TV..... | 1925 |
| <i>Matteo Torcoli, Emanuel A. P. Habets</i> | |
| Decaying Contrast for Fine-Grained Video Representation Learning..... | 1930 |
| <i>Heng Zhang, Bing Su</i> | |
| A Causal Convolutional Approach for Packet Loss Concealment in Low Powered Devices..... | 1935 |
| <i>Steven Davy, Niamh Belton, Joshua Tobin, Owais Bin Zuber, Liu Dong, Yuan Xuwen</i> | |
| The MBSTOI Binaural Intelligibility Metric Using a Close-Talking Microphone Reference | 1940 |
| <i>Pierre Guiraud, Alastair H. Moore, Rebecca R. Vos, Patrick A. Naylor, Mike Brookes</i> | |

| | |
|--|------|
| D ² Q-DETR: Decoupling and Dynamic Queries for Oriented Object Detection with Transformers | 1945 |
| <i>Qiang Zhou, Chaohui Yu, Zhibin Wang, Fan Wang</i> | |
| Cross-Head Supervision for Crowd Counting with Noisy Annotations | 1950 |
| <i>Mingliang Dai, Zhizhong Huang, Jiaqi Gao, Hongming Shan, Junping Zhang</i> | |
| Digital Phenotype Representation by Statistical, Information Theory, Data-Driven Approach with Digital Health Data..... | 1955 |
| <i>Binh Nguyen, Michael Nigro, Alice Rueda, Venkat Bhat, Sri Krishnan</i> | |
| Robust Data2VEC: Noise-Robust Speech Representation Learning for ASR by Combining Regression and Improved Contrastive Learning | 1960 |
| <i>Qiu-Shi Zhu, Long Zhou, Jie Zhang, Shu-Jie Liu, Yu-Chen Hu, Li-Rong Dai</i> | |
| N2MVSNet: Non-Local Neighbors Aware Multi-View Stereo Network | 1965 |
| <i>Zhe Zhang, Huachen Gao, Yuxi Hu, Ronggang Wang</i> | |
| Batch Normalization Damages Federated Learning on NON-IID Data: Analysis and Remedy | 1970 |
| <i>Yanmeng Wang, Qingjiang Shi, Tsung-Hui Chang</i> | |
| Pitch Mark Detection from Noisy Speech Waveform Using Wave-U-Net..... | 1975 |
| <i>Hyun-Joon Nam, Hong-June Park</i> | |
| A Framework for Unified Real-Time Personalized and Non-Personalized Speech Enhancement | 1980 |
| <i>Zhepei Wang, Ritwik Giri, Devansh Shah, Jean-Marc Valin, Michael M. Goodwin, Paris Smaragdis</i> | |
| Look and Think: Intrinsic Unification of Self-Attention and Convolution for Spatial-Channel Specificity..... | 1985 |
| <i>Xiang Gao, Honghui Lin, Yu Li, Ruiyan Fang, Xin Zhang</i> | |
| Sandformer: CNN and Transformer Under Gated Fusion for Sand Dust Image Restoration..... | 1990 |
| <i>Jun Shi, Bingcai Wei, Gang Zhou, Liye Zhang</i> | |
| Filterbank Learning for Noise-Robust Small-Footprint Keyword Spotting | 1995 |
| <i>Iván López-Espejo, Ram C. M. C. Shekar, Zheng-Hua Tan, Jesper Jensen, John H. L. Hansen</i> | |
| Learning ASR Pathways: A Sparse Multilingual ASR Model | 2000 |
| <i>Mu Yang, Andros Tjandra, Chunxi Liu, David Zhang, Duc Le, Ozlem Kalinli</i> | |
| HIFI: A Hierarchical Performer Identification Model Based on Symbolic Representation of Music | 2005 |
| <i>Syed Rifat Mahmud Rafee, György Fazekas, Geraint Wiggins</i> | |
| Neural Speech Phase Prediction Based on Parallel Estimation Architecture and Anti-Wrapping Losses | 2010 |
| <i>Yang Ai, Zhen-Hua Ling</i> | |
| Adversarial Permutation Invariant Training for Universal Sound Separation | 2015 |
| <i>Emilian Postolache, Jordi Pons, Santiago Pascual, Joan Serra</i> | |
| Super-Resolution for Macro X-Ray Fluorescence Data Collected from Old Master Paintings..... | 2020 |
| <i>Su Yarn, Herman Verinaz-Jadan, Jun-Jie Huang, Nathan Daly, Catherine Higgitt, Pier Luigi Dragotti</i> | |
| Audio-Text Models Do Not Yet Leverage Natural Language | 2025 |
| <i>Ho-Hsiang Wu, Oriol Nieto, Juan Pablo Bello, Justin Salamon</i> | |

| | |
|--|------|
| ImagineNet: Target Speaker Extraction with Intermittent Visual Cue Through Embedding Inpainting | 2030 |
| <i>Zexu Pan, Wupeng Wang, Marvin Borsdorf, Haizhou Li</i> | |
| Audio-To-Intent Using Acoustic-Textual Subword Representations from End-To-End ASR..... | 2035 |
| <i>Pranay Dighe, Prateeth Nayak, Oggi Rudovic, Erik Marchi, Xiaochuan Niu, Ahmed Tewfik</i> | |
| DATA2VEC-SG: Improving Self-Supervised Learning Representations for Speech Generation Tasks..... | 2040 |
| <i>Heming Wang, Yao Qian, Hemin Yang, Naoyuki Kanda, Peidong Wang, Takuya Yoshioka, Xiaofei Wang, Yiming Wang, Shujie Liu, Zhuo Chen, Deliang Wang, Michael Zeng</i> | |
| On the Importance of Different Cough Phases for COVID-19 Detection | 2045 |
| <i>Yi Zhu, Mahil Hussain Shaik, Tiago H. Falk</i> | |
| Privacy-Preserving Occupancy Estimation | 2050 |
| <i>Jennifer Williams, Vahid Yazdanpanah, Sebastian Stein</i> | |
| Towards Simultaneous Segmentation of Liver Tumors and Intrahepatic Vessels Via Cross-Attention Mechanism | 2055 |
| <i>Haopeng Kuang, Dingkan Yang, Shunli Wang, Xiaoying Wang, Lihua Zhang</i> | |
| Robust Data-Driven Accelerated Mirror Descent..... | 2060 |
| <i>Hong Ye Tan, Subhadip Mukherjee, Junqi Tang, Andreas Hauptmann, Carola-Bibiane Schönlieb</i> | |
| Naturalistic Head Motion Generation from Speech | 2065 |
| <i>Trisha Mittal, Zakaria Aldeneh, Masha Fedzechkina, Anurag Ranjan, Barry-John Theobald</i> | |
| On the Role of LIP Articulation in Visual Speech Perception | 2070 |
| <i>Zakaria Aldeneh, Masha Fedzechkina, Skyler Seto, Katherine Metcalf, Miguel Sarabia, Nicholas Apostoloff, Barry-John Theobald</i> | |
| Accelerating Matrix Trace Estimation by Aitken's Δ_2 Process..... | 2075 |
| <i>Vassilis Kalantzis, Georgios Kollias, Shashanka Ubaru, Theodoros Salonidis</i> | |
| Long-Short Attention Network for the Spectral Super-Resolution of Multispectral Images..... | 2080 |
| <i>Kai Zhang, Tian Jin, Feng Zhang, Jiande Sun</i> | |
| MADI: Inter-Domain Matching and Intra-Domain Discrimination for Cross-Domain Speech Recognition | 2085 |
| <i>Jiaming Zhou, Shiwan Zhao, Ning Jiang, Guoqing Zhao, Yong Qin</i> | |
| Weight-Based Mask for Domain Adaptation..... | 2090 |
| <i>Eunseop Lee, Inhan Kim, Daijin Kim</i> | |
| Analysing Discrete Self Supervised Speech Representation for Spoken Language Modeling | 2095 |
| <i>Amitay Sicherman, Yossi Adi</i> | |
| Robust Multi-Object Tracking with Spatial Uncertainty..... | 2100 |
| <i>Pin-Jie Liao, Yu-Cheng Huang, Chen-Kuo Chiang, Shang-Hong Lai</i> | |
| SADE: A Self-Adaptive Expert for Multi-Dataset Question Answering..... | 2105 |
| <i>Yixing Peng, Quan Wang, Zhendong Mao, Yongdong Zhang</i> | |

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

| | |
|---|------|
| Hybrid Transformers for Music Source Separation..... | 2195 |
| <i>Simon Rouard, Francisco Massa, Alexandre Défossez</i> | |
| Audio-Driven Facial Landmark Generation in Violin Performance Using 3DCNN Network with Self Attention Model | 2200 |
| <i>Ting-Wei Lin, Chao-Lin Liu, Li Su</i> | |
| Gesper: A Unified Framework for General Speech Restoration..... | 2205 |
| <i>Jun Chen, Yupeng Shi, Wenzhe Liu, Wei Rao, Shulin He, Andong Li, Yannan Wang, Zhiyong Wu, Shidong Shang, Chengshi Zheng</i> | |
| Flexible Beam Design for Vital Sign Monitoring Using a Phased Array Equipped with Double-Phase Shifters | 2207 |
| <i>Zhaoyi Xu, Donglin Gao, Shuping Li, Chung-Tse Michael Wu, Athina Petropulu</i> | |
| Graph Neural Networks for Sound Source Localization on Distributed Microphone Networks | 2212 |
| <i>Eric Grinstein, Mike Brookes, Patrick A. Naylor</i> | |
| FedSD: A New Federated Learning Structure Used in Non-Iid Data | 2217 |
| <i>Minmin Yi, Houchun Ning, Peng Liu</i> | |
| Controllable Music Inpainting with Mixed-Level and Disentangled Representation..... | 2222 |
| <i>Shiqi Wei, Ziyu Wang, Weiguo Gao, Gus Xia</i> | |
| Reliable Beamforming at Terahertz Bands: Are Causal Representations the Way Forward?..... | 2227 |
| <i>Christo Kurisummoottil Thomas, Walid Saad</i> | |
| Light-Weight CNN-Attention Based Architecture for Hand Gesture Recognition Via Electromyography | 2232 |
| <i>Soheil Zabih, Elahe Rahimian, Amir Asif, Arash Mohammadi</i> | |
| Conformer-Based Target-Speaker Automatic Speech Recognition for Single-Channel Audio | 2237 |
| <i>Yang Zhang, Krishna C. Puvvada, Vitaly Lavrukhin, Boris Ginsburg</i> | |
| Effectiveness of Mining Audio and Text Pairs from Public Data for Improving ASR Systems for Low-Resource Languages | 2242 |
| <i>Kaushal Bhogale, Abhigyan Raman, Tahir Javed, Sumanth Doddapaneni, Anoop Kunchukuttan, Pratyush Kumar, Mitesh M. Khapra</i> | |
| Mixed Far-Field and Near-Field Source Localization Based on Low-Rank Matrix Reconstruction..... | 2247 |
| <i>Yunchang Liu, Hong Jiang, Qi Zhang</i> | |
| Cancelling Intermodulation Distortions for Otoacoustic Emission Measurements with Earbuds..... | 2252 |
| <i>Berken Utku Demirel, Khaldoon Al-Naimi, Fahim Kawsar, Alessandro Montanari</i> | |
| Adversarial Guitar Amplifier Modelling with Unpaired Data | 2257 |
| <i>Alec Wright, Vesa Välimäki, Lauri Juvela</i> | |
| Deep Adaptive Superpixels for Hadamard Single Pixel Imaging in Near-Infrared Spectrum..... | 2262 |
| <i>Brayan Monroy, Jorge Bacca, Henry Arguello</i> | |
| Speechlmscore: Evaluating Speech Generation Using Speech Language Model..... | 2267 |
| <i>Soumi Maiti, Yifan Peng, Takaaki Saeki, Shinji Watanabe</i> | |
| Adaptive Filtering Algorithms for Set-Valued Observations-Symmetric Measurement Approach to Unlabeled and Anonymized Data | 2272 |
| <i>Vikram Krishnamurthy</i> | |

| | |
|---|------|
| Training Large-Vocabulary Neural Language Models by Private Federated Learning for Resource-Constrained Devices | 2277 |
| <i>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</i> | |
| Bytecover3: Accurate Cover Song Identification on Short Queries | 2282 |
| <i>Xingjian Du, Zijie Wang, Xia Liang, Huidong Liang, Bilei Zhu, Zejun Ma</i> | |
| A Multi-Modal Approach for Context-Aware Network Traffic Classification | 2287 |
| <i>Bo Pang, Yongquan Fu, Siyuan Ren, Siqi Shen, Ye Wang, Qing Liao, Yan Jia</i> | |
| Wav2Seq: Pre-Training Speech-to-Text Encoder-Decoder Models Using Pseudo Languages | 2292 |
| <i>Felix Wu, Kwangyoum Kim, Shinji Watanabe, Kyu J. Han, Ryan McDonald, Kilian Q. Weinberger, Yoav Artzi</i> | |
| Improved Mask-Based Neural Beamforming for Multichannel Speech Enhancement by Snapshot Matching Masking..... | 2297 |
| <i>Ching-Hua Lee, Chouchang Yang, Yilin Shen, Hongxia Jin</i> | |
| Gct: Gated Contextual Transformer for Sequential Audio Tagging..... | 2302 |
| <i>Yuanbo Hou, Yun Wang, Wenwu Wang, Dick Botteldooren</i> | |
| Diffusion Motion: Generate Text-Guided 3D Human Motion by Diffusion Model | 2307 |
| <i>Zhiyuan Ren, Zhihong Pan, Xin Zhou, Le Kang</i> | |
| Real-Time Wireless ECG-Derived Respiration Rate Estimation Using an Autoencoder with a DCT Layer..... | 2312 |
| <i>Hongyi Pan, Xin Zhu, Zhilu Ye, Pai-Yen Chen, Ahmet Enis Cetin</i> | |
| Transceiver Design for MIMO-DFRC Systems | 2317 |
| <i>Cai Wen, Timothy N. Davidson</i> | |
| Explicit and Implicit Knowledge Distillation Via Unlabeled Data..... | 2322 |
| <i>Yuzheng Wang, Zuhao Ge, Zhaoyu Chen, Xian Liu, Chuangjia Ma, Yunquan Sun, Lizhe Qi</i> | |
| Robust Audio-Visual ASR with Unified Cross-Modal Attention | 2327 |
| <i>Jiahong Li, Chenda Li, Yifei Wu, Yanmin Qian</i> | |
| Similarity Relation Preserving Cross-Modal Learning for Multispectral Pedestrian Detection Against Adversarial Attacks | 2332 |
| <i>Jung Uk Kim, Yong Man Ro</i> | |
| DailyTalk: Spoken Dialogue Dataset for Conversational Text-To-Speech..... | 2337 |
| <i>Keon Lee, Kyumin Park, Daeyoung Kim</i> | |
| Resolving Doppler Ambiguity Via Spread Phase Alignment in FDA-MIMO Radar | 2342 |
| <i>Yanxing Wang, Shengqi Zhu, Guisheng Liao, Lan Lan, Zhuochen Chen, Feilong Liu</i> | |
| Adversarial Contrastive Distillation with Adaptive Denoising | 2347 |
| <i>Yuzheng Wang, Zhaoyu Chen, Dingkang Yang, Yang Liu, Siao Liu, Wenqiang Zhang, Lizhe Qi</i> | |
| Interpretable Multi-Scale Neural Network for Granger Causality Discovery | 2352 |
| <i>Chenchen Fan, Yixin Wang, Yahong Zhang, Wenli Ouyang</i> | |
| Reliable Cluster-Based Framework for Open Set Domain Adaptation | 2357 |
| <i>Xiu Zheng, Yuan Huang, Jie Tang</i> | |

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

| | |
|--|------|
| PAGE: A Position-Aware Graph-Based Model for Emotion Cause Entailment in Conversation..... | 2447 |
| <i>Xiaojie Gu, Renze Lou, Lin Sun, Shangxin Li</i> | |
| Switching Kronecker Product Linear Filtering for Multispeaker Adaptive Speech Dereverberation..... | 2452 |
| <i>Gongping Huang, Jacob Benesty, Israel Cohen, Emil Winebrand, Jingdong Chen, Walter Kellermann</i> | |
| Diffusionnet: An Efficient Framework to Classify Single-Molecule Images with Latent Entropy Minimization | 2457 |
| <i>Soumee Guha, Olivia De Cuba, Andreas Gahlmann, Scott T. Acton</i> | |
| Antenna Impedance Estimation in Correlated Rayleigh Fading Channels | 2462 |
| <i>Shaohan Wu, Brian L. Hughes</i> | |
| Hierarchical Filtering with Online Learned Priors for ECG Denoising | 2467 |
| <i>Timur Locher, Guy Revach, Nir Shlezinger, Ruud J. G. Van Sloun, Rik Vullings</i> | |
| Batch-Ensemble Stochastic Neural Networks for Out-of-Distribution Detection | 2472 |
| <i>Xiongjie Chen, Yunpeng Li, Yongxin Yang</i> | |
| Is Multi-Task Learning an Upper Bound for Continual Learning? | 2477 |
| <i>Zihao Wu, Huy Tran, Hamed Pirsiavash, Soheil Kolouri</i> | |
| M ² TSR: Multi-Range and Mix-Grained Transformer for Single Image Super-Resolution | 2482 |
| <i>Zhong-Han Niu, Qing-Long Zhang, Yi Fan, Yu-Bin Yang</i> | |
| Large Covariance Matrix Estimation with Oracle Statistical Rate | 2487 |
| <i>Quan Wei, Ziping Zhao</i> | |
| Improving Performance of Real-Time Full-Band Blind Packet-Loss Concealment with Predictive Network..... | 2492 |
| <i>Viet-Anh Nguyen, Anh H. T. Nguyen, Andy W. H. Khong</i> | |
| Exploring Self-Supervised Pre-Trained ASR Models for Dysarthric and Elderly Speech Recognition | 2497 |
| <i>Shujie Hu, Xurong Xie, Zengrui Jin, Mengzhe Geng, Yi Wang, Mingyu Cui, Jiajun Deng, Xunying Liu, Helen Meng</i> | |
| Ultimate Negative Sampling for Contrastive Learning | 2502 |
| <i>Huijie Guo, Lei Shi</i> | |
| Designing a 3D-Aware StyleNerf Encoder for Face Editing..... | 2507 |
| <i>Songlin Yang, Wei Wang, Bo Peng, Jing Dong</i> | |
| Bias Reduced Semidefinite Relaxation Method for Multistatic Localization in the Absence of Transmitter Position and Its Synchronization..... | 2512 |
| <i>Jian Pei, Gang Wang, K. C. Ho, Lei Huang</i> | |
| Super-Resolution Information Enhancement for Crowd Counting | 2517 |
| <i>Jiahao Xie, Wei Xu, Dingkang Liang, Zhanyu Ma, Kongming Liang, Weidong Liu, Rui Wang, Ling Jin</i> | |
| Shadocnet: Learning Spatial-Aware Tokens in Transformer for Document Shadow Removal | 2522 |
| <i>Xuhang Chen, Xiaodong Cun, Chi-Man Pun, Shuqiang Wang</i> | |
| Fine-Grained Blind Face inpainting with 3D Face Component Disentanglement | 2527 |
| <i>Yu Bai, Ruian He, Weimin Tan, Bo Yan, Yangle Lin</i> | |

| | |
|--|------|
| Sample-Adapt Fusion Network for RGB-D Hand Detection in the Wild..... | 2532 |
| <i>Xingyu Liu, Pengfei Ren, Yuchen Chen, Cong Liu, Jing Wang, Haifeng Sun, Qi Qi, Jingyu Wang</i> | |
| Inter-Pulse Estimation for Sperm Whale Click Detection..... | 2537 |
| <i>Guy Gubnitsky, Roee Diamant</i> | |
| Dynamic Vehicle Graph Interaction for Trajectory Prediction Based on Video Signals | 2542 |
| <i>Jian Chen, Wei Wang, Junxin Chen, Ming Cai</i> | |
| Pre-Trained Model Representations and Their Robustness Against Noise for Speech Emotion Analysis..... | 2547 |
| <i>Vikramjit Mitra, Vasudha Kowtha, Hsiang-Yun Sherry Chien, Erdrin Azemi, Carlos Avendano</i> | |
| Distributed Admm with Limited Communications Via Deep Unfolding | 2552 |
| <i>Yoav Noah, Nir Shlezinger</i> | |
| Cross-Domain Learning with Normalizing Flow | 2557 |
| <i>Chi Wang, Jian Gao, Yang Hua, Hui Wang</i> | |
| A Fusion-Based and Multi-Layer Method for Low Light Image Enhancement..... | 2562 |
| <i>Xueyan Zhou, Jiachen Guo, Hao Liu, Chao Wang</i> | |
| Assessing the Robustness of Deep Learning-Assisted Pathological Image Analysis Under Practical Variables of Imaging System..... | 2567 |
| <i>Yuxuan Sun, Chenglu Zhu, Yunlong Zhang, Honglin Li, Pingyi Chen, Lin Yang</i> | |
| SFR: Semantic-Aware Feature Rendering of Point Cloud..... | 2572 |
| <i>Yaohua Zha, Rongsheng Li, Tao Dai, Jianyu Xiong, Xin Wang, Shu-Tao Xia</i> | |
| LQGNET: Hybrid Model-Based and Data-Driven Linear Quadratic Stochastic Control | 2577 |
| <i>Solomon Goldgraber Casspi, Oliver Hüsser, Guy Revach, Nir Shlezinger</i> | |
| Self-Supervised Guided Hypergraph Feature Propagation for Semi-Supervised Classification with Missing Node Features..... | 2582 |
| <i>Chengxiang Lei, Sichao Fu, Yuetian Wang, Wenhao Qiu, Yachen Hu, Qinmu Peng, Xinge You</i> | |
| Contrastive Learning of Sentence Embeddings in Product Search..... | 2587 |
| <i>Bo-Wen Zhang, Yan Yan, Jiawei Yu</i> | |
| Boosting Bert Subnets with Neural Grafting..... | 2592 |
| <i>Ting Hu, Christoph Meinel, Haojin Yang</i> | |
| Enhanced GM-PHD Filter for Real Time Satellite Multi-Target Tracking..... | 2597 |
| <i>Camilo Aguilar, Mathias Ortner, Josiane Zerubia</i> | |
| A Wavelet Scattering Approach for Load Identification with Limited Amount of Training Data | 2602 |
| <i>Pascal A. Schirmer, Iosif Mporas</i> | |
| Automatic Severity Classification of Dysarthric Speech by Using Self-Supervised Model with Multi-Task Learning..... | 2607 |
| <i>Eun Jung Yeo, Kwanghee Choi, Sunhee Kim, Minhwa Chung</i> | |
| Ternary Weight Networks..... | 2612 |
| <i>Bin Liu, Fengfu Li, Xiaoxing Wang, Bo Zhang, Junchi Yan</i> | |

| | |
|---|------|
| A Slot-Shared Span Prediction-Based Neural Network for Multi-Domain Dialogue State Tracking | 2617 |
| <i>Abibulla Atawulla, Xi Zhou, Yating Yang, Bo Ma, Fengyi Yang</i> | |
| CPA: Compressed Private Aggregation for Scalable Federated Learning Over Massive Networks | 2622 |
| <i>Natalie Lang, Elad Sofer, Nir Shlezinger, Rafael G. L. D'Oliveira, Salim El Rouayheb</i> | |
| BATT: Backdoor Attack with Transformation-Based Triggers | 2627 |
| <i>Tong Xu, Yiming Li, Yong Jiang, Shu-Tao Xia</i> | |
| A Critical Look at Recent Trends in Compression of Channel State Information..... | 2632 |
| <i>Marcus Valtonen Örnhaq, Stefan Adalbjörnsson, Püren Güler, Mojtaba Mahdavi</i> | |
| Cross-Modal Adversarial Contrastive Learning for Multi-Modal Rumor Detection..... | 2637 |
| <i>Ting Zou, Zhong Qian, Peifeng Li, Qiaoming Zhu</i> | |
| AMC-Net: An Effective Network for Automatic Modulation Classification | 2642 |
| <i>Jiawei Zhang, Tiantian Wang, Zhixi Feng, Shuyuan Yang</i> | |
| WHC: Weighted Hybrid Criterion for Filter Pruning on Convolutional Neural Networks | 2647 |
| <i>Shaowu Chen, Weize Sun, Lei Huang</i> | |
| Transformer-Based Deep Hashing Method for Multi-Scale Feature Fusion | 2652 |
| <i>Chao He, Hongxi Wei</i> | |
| STACKMAPS: A Visualization Technique for Diabetic Retinopathy Grading | 2657 |
| <i>Ismail El-Yamany, Abdelrahman Wael, Noha Adly, Marwan Torki</i> | |
| TAPE: An End-To-End Timbre-Aware Pitch Estimator | 2662 |
| <i>Nazif Can Tamer, Yigitcan Özer, Meinard Müller, Xavier Serra</i> | |
| Towards Zero-Shot Personalized Table-to-Text Generation with Contrastive Persona Distillation | 2667 |
| <i>Haolan Zhan, Xuming Lin, Shaobo Cui, Zhongzhou Zhao, Wei Zhou, Haiqing Chen</i> | |
| Associative Learning Network for Coherent Visual Storytelling | 2672 |
| <i>Xin Li, Chunping Liu, Yi Ji</i> | |
| Diabetic Retinopathy Grading with Weakly-Supervised Lesion Priors..... | 2677 |
| <i>Junlin Hou, Fan Xiao, Jilan Xu, Rui Feng, Yuejie Zhang, Haidong Zou, Lina Lu, Wenwen Xue</i> | |
| Gridless Target Localization for FDA-Mimo Radar with Sparse Arrays..... | 2682 |
| <i>Xiaohuan Wu, Yaxin Liu, Xiaoyuan Jia</i> | |
| BECTRA: Transducer-Based End-To-End ASR with Bert-Enhanced Encoder | 2687 |
| <i>Yosuke Higuchi, Tetsuji Ogawa, Tetsunori Kobayashi, Shinji Watanabe</i> | |
| Robust GMM Parameter Estimation Via the K-BM Algorithm | 2692 |
| <i>Ori Kenig, Koby Todros, Tülay Adalı</i> | |
| Intermpl: Momentum Pseudo-Labeling with Intermediate CTC Loss | 2697 |
| <i>Yosuke Higuchi, Tetsuji Ogawa, Tetsunori Kobayashi, Shinji Watanabe</i> | |
| Brainnetformer: Decoding Brain Cognitive States with Spatial-Temporal Cross Attention..... | 2702 |
| <i>Leheng Sheng, Wenhan Wang, Zhiyi Shi, Jichao Zhan, Youyong Kong</i> | |
| Convergence of Stochastic PDMM | 2707 |
| <i>Sebastian O. Jordan, Thomas W. Sherson, Richard Heusdens</i> | |

| | |
|---|------|
| Spatial Similarity Guidance for Few-Shot Segmentation..... | 2712 |
| <i>Xiaoliu Luo, Zhao Duan, Taiping Zhang</i> | |
| Self-Distillation Hashing for Efficient Hamming Space Retrieval..... | 2717 |
| <i>Hongjia Zhai, Hai Li, Hanzhi Zhang, Hujun Bao, Guofeng Zhang</i> | |
| Benchmarking White Blood Cell Classification Under Domain Shift | 2722 |
| <i>Satoshi Tsutsui, Zhengyang Su, Bihan Wen</i> | |
| Exploiting Sparse Recovery Algorithms for Semi-Supervised Training of Deep Neural Networks for Direction-of-Arrival Estimation..... | 2727 |
| <i>Murtiza Ali, Aditya Arie Nugraha, Karan Nathwani</i> | |
| QTROJAN: A Circuit Backdoor Against Quantum Neural Networks..... | 2732 |
| <i>Cheng Chu, Lei Jiang, Martin Swamy, Fan Chen</i> | |
| Deep Root Music Algorithm for Data-Driven Doa Estimation | 2737 |
| <i>Dor H. Shmuel, Julian P. Merkofer, Guy Revach, Ruud J. G. Van Sloun, Nir Shlezinger</i> | |
| SDG-L: A Semiparametric Deep Gaussian Process Based Framework for Battery Capacity Prediction | 2742 |
| <i>Hanbing Liu, Yanru Wu, Yang Li, Ercan E. Kuruoglu, Xuan Zhang</i> | |
| RAT: Radial Attention Transformer for Singing Technique Recognition | 2747 |
| <i>Guan-Yuan Chen, Ya-Fen Yeh, Von-Wun Soo</i> | |
| Emodiff: Intensity Controllable Emotional Text-To-Speech with Soft-Label Guidance | 2752 |
| <i>Yiwei Guo, Chenpeng Du, Xie Chen, Kai Yu</i> | |
| MEET: A Monte Carlo Exploration-Exploitation Trade-Off for Buffer Sampling | 2757 |
| <i>Julius Ott, Lorenzo Servadei, Jose Arjona-Medina, Enrico Rinaldi, Gianfranco Mauro, Daniela Sánchez Lopera, Michael Stephan, Thomas Stadelmayer, Avik Santra, Robert Wille</i> | |
| Lego-Features: Exporting Modular Encoder Features for Streaming and Deliberation ASR..... | 2762 |
| <i>Rami Botros, Rohit Prabhavalkar, Johan Schalkwyk, Ciprian Chelba, Tara N. Sainath, Françoise Beaufays</i> | |
| Robust Subspace Tracking with Contamination Mitigation Via α -Divergence | 2767 |
| <i>Le Trung Thanh, Aref Miri Rekavandi, Abd-Krim Seghouane, Karim Abed-Meraim</i> | |
| CLAP Learning Audio Concepts from Natural Language Supervision..... | 2772 |
| <i>Benjamin Elizalde, Soham Deshmukh, Mahmoud Al Ismail, Huaming Wang</i> | |
| Self-Supervised Learning with Bi-Label Masked Speech Prediction for Streaming Multi-Talker Speech Recognition | 2777 |
| <i>Zili Huang, Zhuo Chen, Naoyuki Kanda, Jian Wu, Yiming Wang, Jinyu Li, Takuya Yoshioka, Xiaofei Wang, Peidong Wang</i> | |
| Rate Region Characterization for Semantics and Bits Based Multiuser Communications..... | 2782 |
| <i>Xidong Mu, Yuanwei Liu</i> | |
| Cascading and Direct Approaches to Unsupervised Constituency Parsing on Spoken Sentences | 2787 |
| <i>Yuan Tseng, Cheng-I Jeff Lai, Hung-Yi Lee</i> | |
| A Nested Ensemble Method to Bilevel Machine Learning | 2792 |
| <i>Lisha Chen, Momin Abbas, Tianyi Chen</i> | |

| | |
|---|------|
| Police: Provably Optimal Linear Constraint Enforcement for Deep Neural Networks | 2797 |
| <i>Randall Balestriero, Yann Lecun</i> | |
| Heterogeneous Graph Learning for Acoustic Event Classification | 2802 |
| <i>Amir Shirian, Mona Ahmadian, Krishna Somandepalli, Tanaya Guha</i> | |
| Epic-Sounds: A Large-Scale Dataset of Actions that Sound..... | 2807 |
| <i>Jaesung Huh, Jacob Chalk, Evangelos Kazakos, Dima Damen, Andrew Zisserman</i> | |
| Ensemble Graph Q-Learning for Large Scale Networks | 2812 |
| <i>Talha Bozkus, Urbashi Mitra</i> | |
| Terminology-Aware Medical Dialogue Generation..... | 2817 |
| <i>Chen Tang, Hongbo Zhang, Tyler Loakman, Chenghua Lin, Frank Guerin</i> | |
| Prefix-Level Detection and Autocorrection of Keyboard Input Errors..... | 2822 |
| <i>Jerome R. Bellegarda</i> | |
| Enhanced Embeddings in Zero-Shot Learning for Environmental Audio | 2827 |
| <i>Ysobel Sims, Alexandre Mendes, Stephan Chalup</i> | |
| SD-PINN: Physics Informed Neural Networks for Spatially Dependent PDES | 2832 |
| <i>Ruixian Liu, Peter Gerstoft</i> | |
| EEG Emotion Recognition Via Ensemble Learning Representations | 2837 |
| <i>Bilal Taha, Dae Yon Hwang, Dimitrios Hatzinakos</i> | |
| Dual-Use Signal Design for MIMO Radcom with Inter-Pulse Index Modulation | 2842 |
| <i>Xue Yao, Guolong Cui, Xianxiang Yu</i> | |
| Unsupervised Vocal Dereverberation with Diffusion-Based Generative Models..... | 2847 |
| <i>Koichi Saito, Naoki Murata, Toshimitsu Uesaka, Chieh-Hsin Lai, Yuhta Takida, Takao Fukui, Yuki Mitsufuji</i> | |
| Towards Trustworthy Phoneme Boundary Detection with Autoregressive Model and Improved Evaluation Metric | 2852 |
| <i>Hyeongju Kim, Hyeong-Seok Choi</i> | |
| Dual-Head Fusion Network for Image Enhancement..... | 2857 |
| <i>Yuhong Zhang, Hengsheng Zhang, Li Song, Rong Xie, Wenjun Zhang</i> | |
| Spatial-Temporal Graph Convolutional Network Boosted Flow-Frame Prediction for Video Anomaly Detection..... | 2862 |
| <i>Kai Cheng, Xinhua Zeng, Yang Liu, Mengyang Zhao, Chengxin Pang, Xing Hu</i> | |
| Nasty-SFDA: Source Free Domain Adaptation from a Nasty Model..... | 2867 |
| <i>Jiajiong Cao, Yufan Liu, Weiming Bai, Jingting Ding, Liang Li</i> | |
| Removing Radio Frequency Interference from Auroral Kilometric Radiation with Stacked Autoencoders..... | 2872 |
| <i>Allen Chang, Mary Knapp, James Labelle, John Swoboda, Ryan Volz, Philip J. Erickson</i> | |
| Shift to Your Device: Data Augmentation for Device-Independent Speaker Verification Anti-Spoofing | 2877 |
| <i>Junhao Wang, Li Lu, Zhongjie Ba, Feng Lin, Kui Ren</i> | |

| | |
|--|------|
| Transformer-Based Tracking Network for Maneuvering Targets | 2882 |
| <i>Yushu Zhang, Gang Li, Xiao-Ping Zhang, You He</i> | |
| Multi-Stage Aggregation Transformer for Medical Image Segmentation | 2887 |
| <i>Xiaoyan Wang, Minghan Shao, Dongyan Guo, Ying Cui, Xiaojie Huang, Ming Xia, Cong Bai</i> | |
| Ultrasound Image Quality Control Using Speech-Assisted Switchable CycleGAN | 2892 |
| <i>Jaeyoung Huh, Shujaat Khan, Eun Sun Lee, Jong Chul Ye</i> | |
| Efficient Quantized Constant Envelope Precoding for Multiuser Downlink Massive MIMO Systems..... | 2897 |
| <i>Zheyu Wu, Ya-Feng Liu, Bo Jiang, Yu-Hong Dai</i> | |
| Image Fusion Via Slice-Based Convolutional Sparse Representation..... | 2902 |
| <i>Jingchen Xu, Yali Zhang, Ze Li, Jinjia Wang</i> | |
| ICCRN: Inplace Cepstral Convolutional Recurrent Neural Network for Monaural Speech Enhancement | 2907 |
| <i>Jinjiang Liu, Xueliang Zhang</i> | |
| Channel State Information-Free Artificial Noise-Aided Location-Privacy Enhancement..... | 2912 |
| <i>Jianxiu Li, Urbashi Mitra</i> | |
| CAENet: Using Collaborative Attention Transformer and Add-Boost Strategy for Single Image Deraining..... | 2917 |
| <i>Shengdi Qin, Shunli Zhang, Yu Zhang, Haoyu Gao</i> | |
| Sparse Aggregation-Based Channel Estimation for Massive MIMO Systems with Decentralized Baseband Processing | 2922 |
| <i>Yanqing Xu, Enbin Song, Qingjiang Shi, Tsung-Hui Chang</i> | |
| A Few Shot Learning of Singing Technique Conversion Based on Cycle Consistency Generative Adversarial Networks..... | 2927 |
| <i>Po-Wei Chen, Von-Wun Soo</i> | |
| Geogcn: Geometric Dual-Domain Graph Convolution Network for Point Cloud Denoising | 2932 |
| <i>Zhaowei Chen, Peng Li, Zeyong Wei, Honghua Chen, Haoran Xie, Mingqiang Wei, Fu Lee Wang</i> | |
| Wireless Location Tracking Via Complex-Domain Super MDS with Time Series Self-Localization Information..... | 2937 |
| <i>Yuya Nishi, Takumi Takahashi, Hiroki Ilmori, Giuseppe Abreu, Shinsuke Ibi, Seiichi Sampei</i> | |
| A Graph Neural Network Multi-Task Learning-Based Approach for Detection and Localization of Cyberattacks in Smart Grids..... | 2942 |
| <i>Abdulrahman Takiddin, Rachad Atat, Muhammad Ismail, Katherine Davis, Erchin Serpedin</i> | |
| Deep Double Self-Expressive Subspace Clustering | 2947 |
| <i>Ling Zhao, Yunpeng Ma, Shanxiong Chen, Jun Zhou</i> | |
| Dual-Stream Siamese Vision Transformer with Mutual Attention for Radar Gait Verification..... | 2952 |
| <i>Ran Ji, Jiarui Li, Wentao He, Jianfeng Ren, Xudong Jiang</i> | |
| Locality Preserving Multiview Graph Hashing for Large Scale Remote Sensing Image Search | 2957 |
| <i>Wenyun Li, Guo Zhong, Xingyu Lu, Chi-Man Pun</i> | |

| | |
|--|------|
| Dual Path Modeling for Semantic Matching by Perceiving Subtle Conflicts | 2962 |
| <i>Chao Xue, Di Liang, Sirui Wang, Jing Zhang, Wei Wu</i> | |
| UNTAG: Learning Generic Features for Unsupervised Type-Agnostic Deepfake Detection | 2967 |
| <i>Nesryne Mejri, Enjie Ghorbel, Djamila Aouada</i> | |
| A Lightweight Fourier Convolutional Attention Encoder for Multi-Channel Speech Enhancement | 2972 |
| <i>Siyu Sun, Jian Jin, Zhe Han, Xianjun Xia, Li Chen, Yijian Xiao, Piao Ding, Shenyi Song, Roberto Togneri, Haijian Zhang</i> | |
| Iterative Shallow Fusion of Backward Language Model for End-To-End Speech Recognition | 2977 |
| <i>Atsunori Ogawa, Takafumi Moriya, Naoyuki Kamo, Naohiro Tawara, Marc Delcroix</i> | |
| Parallel 2D Seismic Ray Tracing Using Cuda on a Jetson Nano | 2982 |
| <i>Ban-Sok Shin, Luis Wientgens, Dmitriy Shutin</i> | |
| Generalized Relative Harmonic Coefficients | 2987 |
| <i>Yonggang Hu, Sharon Gannot, Thushara D. Abhayapala</i> | |
| Model-Matching Principle Applied to the Design of an Array-Based All-Neural Binaural Rendering System for Audio Telepresence | 2992 |
| <i>Yicheng Hsu, Chenghung Ma, Mingsian R. Bai</i> | |
| Learning Task-Aligned Mask Query for Instance Segmentation..... | 2997 |
| <i>Bin Fu, Hongliang He, Pengxu Wei, Jie Chen</i> | |
| PI-Trans: Parallel-Convmlp and Implicit-Transformation Based Gan for Cross-View Image Translation..... | 3002 |
| <i>Bin Ren, Hao Tang, Yiming Wang, Xia Li, Wei Wang, McU Sebe</i> | |
| Self-Transcriber: Few-Shot Lyrics Transcription with Self-Training..... | 3007 |
| <i>Xiaoxue Gao, Xianghu Yue, Haizhou Li</i> | |
| Multi-Level Fusion for Burst Super-Resolution with Deep Permutation-Invariant Conditioning..... | 3012 |
| <i>Martina Cilia, Diego Valsesia, Giulia Fracastoro, Enrico Magli</i> | |
| Source-Free Unsupervised Domain Adaptation for Question Answering | 3017 |
| <i>Zishuo Zhao, Yuexiang Xie, Jingyou Xie, Zhenzhou Lin, Yaliang Li, Ying Shen</i> | |
| Doppler-Coded Joint Division Multiple Access Waveform for Automotive MIMO Radar..... | 3022 |
| <i>Yanhua Wang, Qiubo Pei, Xueyao Hu, Jiamin Long, Hao Yu, Le Zheng</i> | |
| Maximum Likelihood Distillation for Robust Modulation Classification..... | 3027 |
| <i>Javier Maroto, G r me Bovet, Pascal Frossard</i> | |
| Speaker Diaphragm Excursion Prediction: Deep Attention and Online Adaptation..... | 3032 |
| <i>Yuwei Ren, Matt Zivney, Yin Huang, Eddie Choy, Chirag Patel, Hao Xu</i> | |
| A Novel Heart Rate Estimation Method Exploiting Heartbeat Second Harmonic Reconstruction Via Millimeter Wave Radar..... | 3037 |
| <i>Tao Li, Huayu Shou, Yuchen Deng, Yu Zhou, Chenqi Shi, Pengpeng Chen</i> | |
| Centroid Distance Distillation for Effective Rehearsal in Continual Learning | 3042 |
| <i>Daofeng Liu, Fan Lyu, Linyan Li, Zhenping Xia, Fuyuan Hu</i> | |
| On Tracking a Stochastically Time-Varying Subspace..... | 3047 |
| <i>Victor Solo</i> | |

| | |
|---|------|
| Difference Guided VHR Remote Sensing Image Change Detection..... | 3052 |
| <i>Jiukai Sun, Ganchao Liu, Xuelong Li, Yuan Yuan</i> | |
| TT-Net: Dual-Path Transformer Based Sound Field Translation in the Spherical Harmonic Domain | 3057 |
| <i>Yiwen Wang, Zijian Lan, Xihong Wu, Tianshu Qu</i> | |
| Parafac2-Based Coupled Matrix and Tensor Factorizations | 3062 |
| <i>Carla Schenker, Xiulin Wang, Evrim Acar</i> | |
| False Alarm Regulation for Off-Grid Target Detection with the Matched Filter..... | 3067 |
| <i>P Develter, J. Bosse, O. Rabaste, P. Forster, J.-P. Ovarlez</i> | |
| MCKD: Mutually Collaborative Knowledge Distillation for Federated Domain Adaptation and Generalization | 3072 |
| <i>Ziwei Niu, Hongyi Wang, Hao Sun, Shuyi Ouyang, Yen-Wei Chen, Lanfen Lin</i> | |
| A DNN-Based Hearing-Aid Strategy for Real-Time Processing: One Size Fits All | 3077 |
| <i>Fotios Drakopoulos, Arthur Van Den Broucke, Sarah Verhulst</i> | |
| Unsupervised Model-Based Speaker Adaptation of End-To-End Lattice-Free MMI Model for Speech Recognition | 3082 |
| <i>Xurong Xie, Xunying Liu, Hui Chen, Hongan Wang</i> | |
| A Bandit Online Convex Optimization Approach to Distributed Energy Management in Networked Systems..... | 3087 |
| <i>Ioannis Tsetis, Xiaotong Cheng, Setareh Maghsudi</i> | |
| Query-Utterance Attention with Joint Modeuing for Query-Focused Meeting Summarization..... | 3092 |
| <i>Xingxian Liu, Bin Duan, Bo Xiao, Yajing Xu</i> | |
| Optimized Dithering for Quantization Index Modulation | 3097 |
| <i>Shanxiang Lyu</i> | |
| One-Shot Medical Action Recognition with a Cross-Attention Mechanism and Dynamic Time Warping | 3102 |
| <i>Leiyu Xie, Yuxing Yang, Zeyu Fu, Syed Mohsen Naqvi</i> | |
| Defense Against Black-Box Adversarial Attacks Via Heterogeneous Fusion Features | 3107 |
| <i>Jiahuan Zhang, Keisuke Maeda, Takahiro Ogawa, Miki Haseyama</i> | |
| Slot-Triggered Contextual Biasing for Personalized Speech Recognition Using Neural Transducers..... | 3112 |
| <i>Sibo Tong, Philip Harding, Simon Wiesler</i> | |
| Model-Free Online Learning for Waveform Optimization in Integrated Sensing and Communications..... | 3117 |
| <i>Petteri Pulkkinen, Visa Koivunen</i> | |
| Graph Learning from Gaussian and Stationary Graph Signals..... | 3122 |
| <i>Andrei Buciulea, Antonio G. Marques</i> | |
| Hierarchical Softmax for End-To-End Low-Resource Multilingual Speech Recognition..... | 3127 |
| <i>Qianying Liu, Zhuo Gong, Zhengdong Yang, Yuhang Yang, Sheng Li, Chenchen Ding, Nobuaki Minematsu, Hao Huang, Fei Cheng, Chenhui Chu, Sadao Kurohashi</i> | |
| DMSA: Dynamic Multi-Scale Unsupervised Semantic Segmentation Based on Adaptive Affinity | 3132 |
| <i>Kun Yang, Jun Lu</i> | |

| | |
|--|------|
| Image Completion Via Dual-Path Cooperative Filtering..... | 3137 |
| <i>Pourya Shamsolmoali, Masoumeh Zareapoor, Eric Granger</i> | |
| RGB-D Based Pose-Invariant Face Recognition Via Attention Decomposition Module..... | 3142 |
| <i>Wei-Chen Lin, Ching-Te Chiu, Kuan-Chang Shih</i> | |
| Parameter-Efficient Transfer Learning of Pre-Trained Transformer Models for Speaker Verification Using Adapters | 3147 |
| <i>Junyi Peng, Themos Stafylakis, Rongzhi Gu, Oldrich Plchot, Ladislav Mošner, Lukáš Burget, Jan Cernocký</i> | |
| A Dynamic Graph Interactive Framework with Label-Semantic Injection for Spoken Language Understanding | 3152 |
| <i>Zhihong Zhu, Weiyuan Xu, Xuxin Cheng, Tengtao Song, Yuexian Zou</i> | |
| Hyneter: Hybrid Network Transformer for Object Detection..... | 3157 |
| <i>Dong Chen, Duoqian Miao, Xuerong Zhao</i> | |
| Towards Domain Generalisation in ASR with Elitist Sampling and Ensemble Knowledge Distillation | 3162 |
| <i>Rehan Ahmad, Md Asif Jalal, Muhammad Umar Farooq, Anna Ollerenshaw, Thomas Hain</i> | |
| Structural Reparameterization Lightweight Network for Video Action Recognition..... | 3167 |
| <i>Anlei Zhu, Yinghui Wang, Wei Li, Pengjiang Qian</i> | |
| VLKP: Video Instance Segmentation with Visual-Linguistic Knowledge Prompts..... | 3172 |
| <i>Ruixiang Chen, Sheng Liu, Junhao Chen, Bingnan Guo, Feng Zhang</i> | |
| Inverse Reinforcement Learning with Graph Neural Networks for IoT Resource Allocation..... | 3177 |
| <i>Guangchen Wang, Peng Cheng, Zhuo Chen, Wei Xiang, Branka Vucetic, Yonghui Li</i> | |
| Cross-Site Generalization for Imbalanced Epileptic Classification..... | 3182 |
| <i>Tala Abdallah, Nisrine Jrad, Fahed Abdallah, Anne Humeau-Heurtier, Patrick Van Bogaert</i> | |
| Volumetric 3D Reconstruction with Window-Wise Global Feature Aggregation | 3187 |
| <i>Shihao Ren, Yikang Ding, Jinli Liao, Xinghui Li, Jia Guo, Wensen Feng, Xueqian Wang</i> | |
| FFFN: Fashion Feature Fusion Network by Co-Attention Model for Fashion Recommendation..... | 3192 |
| <i>Zhantu Lin, Xiaoyan Zhang</i> | |
| Low-Rank Constrained Memory Autoencoder for Hyperspectral Anomaly Detection..... | 3197 |
| <i>Yuyun Lian, Yongshan Zhang, Xuxiang Feng, Xinwei Jiang, Zhihua Cai</i> | |
| Effect of Lossy Compression Algorithms on Face Image Quality and Recognition | 3202 |
| <i>Torsten Schlett, Sebastian Schachner, Christian Rathgeb, Juan Tapia, Christoph Busch</i> | |
| Human Pose Estimation from Ambiguous Pressure Recordings with Spatio-Temporal Masked Transformers..... | 3207 |
| <i>Vandad Davoodnia, Ali Etemad</i> | |
| Fast Low-Latency Convolution by Low-Rank Tensor Approximation | 3212 |
| <i>Martin Jälmbly, Filip Elvander, Toon Van Waterschoot</i> | |
| TrOMR: Transformer-Based Polyphonic Optical Music Recognition | 3217 |
| <i>Yixuan Li, Huaping Liu, Qiang Jin, Miaomiao Cai, Peng Li</i> | |

| | |
|---|------|
| Adversarial Data Augmentation Using VAE-GAN for Disordered Speech Recognition | 3222 |
| <i>Zengrui Jin, Xurong Xie, Mengzhe Geng, Tianzi Wang, Shujie Hu, Jiajun Deng, Guinan Li, Xunying Liu</i> | |
| Scalable Multi-Task Semantic Communication System with Feature Importance Ranking | 3227 |
| <i>Jiangjing Hu, Fengyu Wang, Wenjun Xu, Hui Gao, Ping Zhang</i> | |
| Design and Performance of the Low-Power Noise Reduction Algorithm of the Med-El Sonnet 2™ Cochlear Implant Audio Processor | 3232 |
| <i>Ernst Aschbacher, Florian Frühauf, Anja Kurz, Peter Nopp</i> | |
| An Evaluation Platform to Scope Performance of Synthetic Environments in Autonomous Ground Vehicles Simulation | 3237 |
| <i>Xiangyu Bai, Le Jiang, Yedi Luo, Aniket Gupta, Pushyami Kaveti, Hanumant Singh, Sarah Ostadabbas</i> | |
| Your Camera Improves Your Point Cloud Compression | 3242 |
| <i>Yuhuan Lin, Tongda Xu, Ziyu Zhu, Yanghao Li, Zhe Wang, Yan Wang</i> | |
| Angle-Of-Arrival Target Tracking Using a Mobile Uav in External Signal-Denied Environment | 3247 |
| <i>Bing Zhu, Sheng Xu, Feng Rice, Kutluyil Dogançay</i> | |
| Inductive Relation Prediction from Relational Paths and Context with Hierarchical Transformers | 3252 |
| <i>Jiaang Li, Quan Wang, Zhendong Mao</i> | |
| MSFORMER: Multi-Scale Transformer with Neighborhood Consensus for Feature Matching..... | 3257 |
| <i>Dongyue Li, Yaping Yan, Dong Liang, Songlin Du</i> | |
| ECG Artifact Removal from Single-Channel Surface EMG Using Fully Convolutional Networks | 3262 |
| <i>Kuan-Chen Wang, Kai-Chun Liu, Sheng-Yu Peng, Yu Tsao</i> | |
| Volumetric Attribute Compression for 3D Point Clouds Using Feedforward Network with Geometric Attention | 3267 |
| <i>Tam Thuc Do, Philip A. Chou, Gene Cheung</i> | |
| Continual Cell Instance Segmentation of Microscopy Images..... | 3272 |
| <i>Tzu-Ting Chuang, Ting-Yun Wei, Yu-Hsing Hsieh, Chu-Song Chen, Huei-Fang Yang</i> | |
| High-Frequency Transformer Network Based on Window Cross-Attention for Pansharpener..... | 3277 |
| <i>Chengjie Ke, Hao Liang, Duidui Li, Xin Tian</i> | |
| An Efficient Beam-Sharing Algorithm for RIS-aided Simultaneous Wireless Information and Power Transfer Applications | 3282 |
| <i>Nguyen Minh Tran, Muhammad Miftahul Amri, Je Hyeon Park, Dong In Kim, Kae Won Choi</i> | |
| Cumulative Attention Based Streaming Transformer ASR with Internal Language Model Joint Training and Rescoring | 3287 |
| <i>Mohan Li, Cong-Thanh Do, Rama Doddipatla</i> | |
| A 3D-Assisted Framework to Evaluate the Quality of Head Motion Replication by Reenactment DEEPFAKE Generators | 3292 |
| <i>Sahar Husseini, Jean-Luc Dugelay, Fabien Aili, Emmanuel Nars</i> | |
| Cross-Modal Audio-Visual Co-Learning for Text-Independent Speaker Verification..... | 3297 |
| <i>Meng Liu, Kong Aik Lee, Longbiao Wang, Hanyi Zhang, Chang Zeng, Jianwu Dang</i> | |

| | |
|--|------|
| Multiple Domain-Adversarial Ensemble Learning for Domain Generalization..... | 3302 |
| <i>Ze-Yu Mi, Kun Long, Yu-Bin Yang</i> | |
| Compose & Embellish: Well-Structured Piano Performance Generation Via a Two-Stage Approach..... | 3307 |
| <i>Shih-Lun Wu, Yi-Hsuan Yang</i> | |
| Multi-Scale Receptive Field Graph Model for Emotion Recognition in Conversations | 3312 |
| <i>Jie Wei, Guanyu Hu, Luu Anh Tuan, Xinyu Yang, Wenjing Zhu</i> | |
| Meta-Dag: Meta Causal Discovery Via Bilevel Optimization..... | 3317 |
| <i>Songtao Lu, Tian Gao</i> | |
| Efficient Monaural Speech Enhancement with Universal Sample Rate Band-Split RNN | 3322 |
| <i>Jianwei Yu, Yi Luo</i> | |
| LA-VOCE: LOW-SNR Audio-Visual Speech Enhancement Using Neural Vocoders | 3327 |
| <i>Rodrigo Mira, Buye Xu, Jacob Donley, Anurag Kumar, Stavros Petridis, Vamsi Krishna Ithapu, Maja Pantic</i> | |
| Outlier-Insensitive Kalman Filtering Using NUV Priors | 3332 |
| <i>Shunit Truzman, Guy Revach, Nir Shlezinger, Itzik Klein</i> | |
| Semantics-Disentangled Contrastive Embedding for Generalized Zero-Shot Learning | 3337 |
| <i>Jian Ni, Yong Liao</i> | |
| Learnable Frontends that Do Not Learn: Quantifying Sensitivity to Filterbank Initialisation | 3342 |
| <i>Mark Anderson, Tomi Kinnunen, Naomi Harte</i> | |
| MMCosine: Multi-Modal Cosine Loss Towards Balanced Audio-Visual Fine-Grained Learning..... | 3347 |
| <i>Ruize Xu, Ruoxuan Feng, Shi-Xiong Zhang, Di Hu</i> | |
| Balanced Mixup Loss for Long-Tailed Visual Recognition | 3352 |
| <i>Haibo Ye, Fangyu Zhou, Xinjie Li, Qingheng Zhang</i> | |
| Local Feature Enhanced Adversarial Network for the Blind Image Quality Assessment | 3357 |
| <i>Xiaomei Shi, Min Zhang, Shouhai Xia, Ruxue Zhang, Jun Feng</i> | |
| A Multi-Signal Perception Network for Textile Composition Identification..... | 3362 |
| <i>Bo Peng, Liren He, Dong Wu, Mingmin Chi, Jintao Chen</i> | |
| Cross-Speaker Emotion Transfer by Manipulating Speech Style Latents | 3367 |
| <i>Suhee Jo, Younggun Lee, Yookyung Shin, Yeongtae Hwang, Taesu Kim</i> | |
| Neural Network Models with Integrated Training and Adaptation for Nonlinear Acoustic System Identification | 3372 |
| <i>Svantje Voit, Gerald Enzner</i> | |
| Fast Online Source Steering Algorithm for Tracking Single Moving Source Using Online Independent Vector Analysis | 3377 |
| <i>Taishi Nakashima, Rintaro Ikeshita, Nobutaka Ono, Shoko Araki, Tomohiro Nakatani</i> | |
| Priv-Aug-Shap-ECGResNet: Privacy Preserving Shapley-Value Attributed Augmented Resnet for Practical Single-Lead Electrocardiogram Classification | 3382 |
| <i>Arijit Ukil, Leandro Marin, Antonio J. Jara</i> | |
| Single-Photon Image Super-Resolution Via Self-Supervised Learning | 3387 |
| <i>Yiwei Chen, Chen Jiang, Yu Pan</i> | |

| | |
|---|------|
| Spatial Correlation Fusion Network for Few-Shot Segmentation | 3392 |
| <i>Xueliang Wang, Wenqi Huang, Wenming Yang, Qingmin Liao</i> | |
| Privacy-Preserving Automatic Speaker Diarization | 3397 |
| <i>Francisco Teixeira, Alberto Abad, Bhiksha Raj, Isabel Trancoso</i> | |
| Local Graph-Homomorphic Processing for Privatized Distributed Systems | 3402 |
| <i>Elsa Rizk, Stefan Vlaski, Ali H. Sayed</i> | |
| ACE-VC: Adaptive and Controllable Voice Conversion Using Explicitly Disentangled Self-Supervised Speech Representations | 3407 |
| <i>Shehzeen Hussain, Paarth Neekhara, Jocelyn Huang, Jason Li, Boris Ginsburg</i> | |
| Frequency-Aware Attentional Feature Fusion for Deepfake Detection | 3412 |
| <i>Cheng Tian, Zhiming Luo, Guimin Shi, Shaozi Li</i> | |
| Towards Improved Room Impulse Response Estimation for Speech Recognition | 3417 |
| <i>Anton Ratnarajah, Ishwarya Ananthabhotla, Vamsi Krishna Ithapu, Pablo Hoffmann, Dinesh Manocha, Paul Calamia</i> | |
| Sora: Scalable Black-Box Reachability Analyser on Neural Networks | 3422 |
| <i>Peipei Xu, Fu Wang, Wenjie Ruan, Chi Zhang, Xiaowei Huang</i> | |
| Data-Driven Quickest Change Detection in Markov Models | 3427 |
| <i>Qi Zhang, Zhongchang Sun, Luis C. Herrera, Shaofeng Zou</i> | |
| 2DSBG: A 2D Semi Bi-Gaussian Filter Adapted for Adjacent and Multi-Scale Line Feature Detection | 3432 |
| <i>Baptiste Magnier, Ghulam Sakhi Shokouh, Louis Berthier, Marcel Pie, Adrien Ruggiero</i> | |
| TreeXGNN: Can Gradient-Boosted Decision Trees Help Boost Heterogeneous Graph Neural Networks? | 3437 |
| <i>Ming-Yi Hong, Shih-Yen Chang, Hao-Wei Hsu, Yi-Hsiang Huang, Chih-Yu Wang, Che Lin</i> | |
| Prompt Makes Mask Language Models Better Adversarial Attackers | 3442 |
| <i>He Zhu, Ce Li, Haitian Yang, Yan Wang, Weiqing Huang</i> | |
| Rate-Distortion Optimized Variable-Node-size Trisoup for Point Cloud Coding | 3447 |
| <i>Kyohei Unno, Kohei Matsuzaki, Satoshi Komorita, Kei Kawamura</i> | |
| I Hear Your True Colors: Image Guided Audio Generation | 3452 |
| <i>Roy Sheffer, Yossi Adi</i> | |
| Windowed Fourier Analysis for Signal Processing on Graph Bundles | 3457 |
| <i>T. Mitchell Roddenberry, Santiago Segarra</i> | |
| Learning Dynamic Graphs Under Partial Observability | 3462 |
| <i>Michele Cirillo, Vincenzo Matta, Ali H. Sayed</i> | |
| NL-DSE: Non-Local Neural Network with Decoder-Squeeze-and-Excitation for Monocular Depth Estimation | 3467 |
| <i>Tsung-Han Tsai, Wei-Chung Wan</i> | |
| Tensor Decomposition Based Latent Feature Clustering for Hyperspectral Band Selection | 3471 |
| <i>Jianwen Qi, Jie Zhang, Yongshan Zhang, Xinwei Jiang, Zhihua Cai</i> | |

| | |
|---|------|
| Column-Based Matrix Approximation with Quasi-Polynomial Structure..... | 3476 |
| <i>Jeongmin Chae, Praneeth Narayanamurthy, Selin Bac, Shaama Mallikarjun Sharada, Urbashi Mitra</i> | |
| Towards Polymorphic Adversarial Examples Generation for Short Text..... | 3481 |
| <i>Yuhang Liang, Zheng Lin, Fengcheng Yuan, Hanwen Zhang, Lei Wang, Weiping Wang</i> | |
| ERBNet: An Effective Representation Based Network for Unbiased Scene Graph Generation..... | 3486 |
| <i>Wenxi Ma, Tianxiang Hou, Qianji Di, Zhongang Qi, Ying Shan, Hanzi Wang</i> | |
| He-Gan: Differentially Private Gan Using Hamiltonian Monte Carlo Based Exponential Mechanism..... | 3491 |
| <i>Usman Hassan, Dongjie Chen, Sen-Ching S. Cheung, Chen-Nee Chuah</i> | |
| Order Reduction of Multi-Channel FIR Filters by Balanced Truncation..... | 3496 |
| <i>Florian Hilgemann, Peter Jax</i> | |
| Hyperbolic Audio Source Separation..... | 3501 |
| <i>Darius Petermann, Gordon Wichern, Aswin Subramanian, Jonathan Le Roux</i> | |
| Learning Robust Self-Attention Features for Speech Emotion Recognition with Label-Adaptive Mixup..... | 3506 |
| <i>Lei Kang, Lichao Zhang, Dazhi Jiang</i> | |
| An Analysis of Degenerating Speech Due to Progressive Dysarthria on ASR Performance..... | 3511 |
| <i>Katrin Tomanek, Katie Seaver, Pan-Pan Jiang, Richard Cave, Lauren Harrell, Jordan R. Green</i> | |
| Enhancing Representation Learning with Deep Classifiers in Presence of Shortcut..... | 3516 |
| <i>Amirhossein Ahmadian, Fredrik Lindsten</i> | |
| Perceptual Analysis of Speaker Embeddings for Voice Discrimination Between Machine and Human Listening..... | 3521 |
| <i>Iordanis Thoidis, Clément Gaultier, Tobias Goehring</i> | |
| An Online Algorithm for Chance Constrained Resource Allocation..... | 3526 |
| <i>Yuwei Chen, Zengde Deng, Yinzhi Zhou, Zaiyi Chen, Yujie Chen, Haoyuan Hu</i> | |
| Learning with Multigraph Convolutional Filters..... | 3531 |
| <i>Landon Butler, Alejandro Parada-Mayorga, Alejandro Ribeiro</i> | |
| A New Probabilistic Distance Metric with Application in Gaussian Mixture Reduction..... | 3536 |
| <i>Ahmad Sajedi, Yuri A. Lawryshyn, Konstantinos N. Plataniotis</i> | |
| Cross-Domain Diffusion Based Speech Enhancement for Very Noisy Speech..... | 3541 |
| <i>Heming Wang, Deliang Wang</i> | |
| Learned Kalman Filtering in Latent Space with High-Dimensional Data..... | 3546 |
| <i>Itay Buchnik, Damiano Steger, Guy Revach, Ruud J. G. Van Sloun, Tirza Routtenberg, Nir Shlezinger</i> | |
| Mendam: Multi-Expert Network with Distribution-Aware Momentum for Long-Tailed Recognition..... | 3551 |
| <i>Qingheng Zhang, Haibo Ye, Kaicheng Yu</i> | |
| Cramér-Rao Bound on Lie Groups with Observations on Lie Groups: Application to SE(2)..... | 3556 |
| <i>Samy Labsir, Alexandre Renaux, Jordi Vilà-Valls, Éric Chaumette</i> | |

| | |
|--|------|
| Signal Processing on Product Spaces | 3561 |
| <i>T. Mitchell Roddenberry, Vincent P. Grande, Florian Frantzen, Michael T. Schaub, Santiago Segarra</i> | |
| Block-Based Color Constancy: The Deviation of Salient Pixels..... | 3566 |
| <i>Oguzhan Ulucan, Diclehan Ulucan, Marc Ebner</i> | |
| Difference Coarrays of Rational Arrays | 3571 |
| <i>Pranav Kulkarni, P. P. Vaidyanathan</i> | |
| Online Edge Flow Prediction Over Expanding Simplicial Complexes | 3576 |
| <i>Maosheng Yang, Bishwadeep Das, Elvin Isufi</i> | |
| Spectral Clustering-Aware Learning of Embeddings for Speaker Diarisation | 3581 |
| <i>Evonne P. C. Lee, Guangzhi Sun, Chao Zhang, Philip C. Woodland</i> | |
| Improvements to Embedding-Matching Acoustic-to-Word ASR Using Multiple-Hypothesis Pronunciation-Based Embeddings..... | 3586 |
| <i>Hao Yen, Woojay Jeon</i> | |
| Various Performance Bounds on the Estimation of Low-Rank Probability Mass Function Tensors from Partial Observations..... | 3591 |
| <i>Tomer HersHKovitz, Martin Haardt, Arie Yeredor</i> | |
| Implicit Bayes Adaptation: A Collaborative Transport Approach | 3596 |
| <i>Bo Jiang, Hamid Krim, Tianfu Wu, Derya Cansever</i> | |
| Adaptive Endpointing with Deep Contextual Multi-Armed Bandits..... | 3601 |
| <i>Do June Min, Andreas Stolcke, Anirudh Raju, Colin Vaz, Di He, Venkatesh Ravichandran, Viet Anh Trinh</i> | |
| Dynamic Speech Endpoint Detection with Regression Targets..... | 3606 |
| <i>Dawei Liang, Hang Su, Tarun Singh, Jay Mahadeokar, Shanil Puri, Jiedan Zhu, Edison Thomaz, Mike Seltzer</i> | |
| NVOC-22: A Low Cost Mel Spectrogram Vocoder for Mobile Devices..... | 3611 |
| <i>Rakesh Iyer</i> | |
| Sparse and Structured Modelling of Underwater Acoustic Channel Impulse Responses..... | 3616 |
| <i>Chaoran Yang, Qing Ling, Xueli Sheng, Mengfei Mu, Andreas Jakobsson</i> | |
| Self-Supervised Representations for Singing Voice Conversion..... | 3621 |
| <i>Tejas Jayashankar, Jilong Wu, Leda Sari, David Kant, Vimal Manohar, Qing He</i> | |
| Improving Speech Prosody of Audiobook Text-To-Speech Synthesis with Acoustic and Textual Contexts..... | 3626 |
| <i>Detai Xin, Sharath Adavanne, Federico Ang, Ashish Kulkarni, Shinnosuke Takamichi, Hiroshi Saruwatari</i> | |
| Utility Polelocalization by Learning from Ambient Traces on Distributed Acoustic Sensing..... | 3631 |
| <i>Zhuocheng Jiang, Yue Tian, Yangmin Ding, Sarper Ozharar, Ting Wang</i> | |
| Passive Acoustic Tracking of Whales in 3-D..... | 3636 |
| <i>Junsu Jang, Florian Meyer, Eric R. Snyder, Sean M. Wiggins, Simone Baumann-Pickering, John A. Hildebrand</i> | |

| | |
|--|------|
| Customized Automatic Face Beautification | 3641 |
| <i>Wang Chen, Peizhen Chen, Weijie Chen, Luojun Lin</i> | |
| A Unified One-Shot Prosody and Speaker Conversion System with Self-Supervised Discrete Speech Units..... | 3646 |
| <i>Li-Wei Chen, Shinji Watanabe, Alexander Rudnicky</i> | |
| Phoneme-Level Bert for Enhanced Prosody of Text-To-Speech with Grapheme Predictions..... | 3651 |
| <i>Yinghao Aaron Li, Cong Han, Xilin Jiang, Nima Mesgarani</i> | |
| Surface-Sampling Based Objective Quality Assessment Metrics for Meshes..... | 3656 |
| <i>Chunyang Fu, Xiang Zhang, Thuong Nguyen-Canh, Xiaozhong Xu, Ge Li, Shan Liu</i> | |
| Generative Modeling Based Manifold Learning for Adaptive Filtering Guidance | 3661 |
| <i>Karim Helwani, Paris Smaragdis, Michael M. Goodwin</i> | |
| Applying Symmetrical Component Transform for Industrial Appliance Classification in Non- Intrusive Load Monitoring | 3666 |
| <i>Anthony Faustine, Lucas Pereira</i> | |
| AURA: Privacy-Preserving Augmentation to Improve Test Set Diversity in Speech Enhancement..... | 3671 |
| <i>Xavier Gitiaux, Aditya Khant, Ebrahim Beyrami, Chandan Reddy, Jayant Gupchup, Ross Cutler</i> | |
| SCA: Streaming Cross-Attention Alignment for Echo Cancellation..... | 3676 |
| <i>Yang Liu, Yangyang Shi, Yun Li, Kaustubh Kalgaonkar, Sriram Srinivasan, Xin Lei</i> | |
| Quantpipe: Applying Adaptive Post-Training Quantization for Distributed Transformer Pipelines in Dynamic Edge Environments..... | 3681 |
| <i>Haonan Wang, Connor Imes, Souvik Kundu, Peter A. Beerel, Stephen P. Crago, John Paul Walters</i> | |
| Quantitative Evidence on Overlooked Aspects of Enrollment Speaker Embeddings for Target Speaker Separation | 3686 |
| <i>Xiaoyu Liu, Xu Li, Joan Serra</i> | |
| Fast Cross-Correlation for TDoA Estimation on Small Aperture Microphone Arrays..... | 3691 |
| <i>François Grondin, Marc-Antoine Maheux, Jean-Samuel Lauzon, Jonathan Vincent, François Michaud</i> | |
| Interpolation Filter Model for Ramanujan Subspace Signals | 3696 |
| <i>Pranav Kulkarni, P. P. Vaidyanathan</i> | |
| Speaker Change Detection for Transformer Transducer ASR | 3701 |
| <i>Jian Wu, Zhuo Chen, Min Hu, Xiong Xiao, Jinyu Li</i> | |
| Wiener Filtering Without Covariance Matrix Inversion | 3706 |
| <i>Pranav U. Damale, Edwin K. P. Chong, Louis L. Scharf</i> | |
| Cough Detection Using Millimeter-Wave Fmcw Radar..... | 3711 |
| <i>Kawon Han, Songcheol Hong</i> | |
| RNN-Based Step-Size Estimation for the RLS Algorithm with Application to Acoustic Echo Cancellation..... | 3716 |
| <i>Ofer Schwartz, Ayal Schwartz</i> | |

| | |
|--|------|
| Bridging Speech and Textual Pre-Trained Models with Unsupervised ASR..... | 3721 |
| <i>Jiatong Shi, Chan-Jan Hsu, Holam Chung, Dongji Gao, Paola Garcia, Shinji Watanabe, Ann Lee, Hung-Yi Lee</i> | |
| FED-3DA: A Dynamic and Personalized Federated Learning Framework..... | 3726 |
| <i>Hui Wang, Jie Sun, Tianyu Wo, Xudong Liu</i> | |
| A New Approach to Extract Fetal Electrocardiogram Using Affine Combination of Adaptive Filters | 3731 |
| <i>Yu Xuan, Xiangyu Zhang, Shuyue Stella Li, Zihan Shen, Xin Xie, Leibny Paola Garcia, Roberto Togneri</i> | |
| Performing Neural Architecture Search Without Gradients | 3736 |
| <i>Pavel Rumiantsev, Mark Coates</i> | |
| Accelerating RNN-T Training and Inference Using CTC Guidance | 3741 |
| <i>Yongqiang Wang, Zhehuai Chen, Chengjian Zheng, Yu Zhang, Wei Han, Parisa Haghani</i> | |
| GANStrument: Adversarial Instrument Sound Synthesis with Pitch-Invariant Instance Conditioning..... | 3746 |
| <i>Gaku Narita, Junichi Shimizu, Taketo Akama</i> | |
| Virtuoso: Massive Multilingual Speech-Text Joint Semi-Supervised Learning for Text-to-Speech | 3751 |
| <i>Takaaki Saeki, Heiga Zen, Zhehuai Chen, Nobuyuki Morioka, Gary Wang, Yu Zhang, Ankur Bapna, Andrew Rosenberg, Bhuvana Ramabhadran</i> | |
| Deep Fusion of Multi-Object Densities Using Transformer..... | 3756 |
| <i>Lechi Li, Chen Dai, Yuxuan Xia, Lennart Svensson</i> | |
| DASA: Difficulty-Aware Semantic Augmentation for Speaker Verification..... | 3761 |
| <i>Yuanyuan Wang, Yang Zhang, Zhiyong Wu, Zhihan Yang, Tao Wei, Kun Zou, Helen Meng</i> | |
| Frequency Reciprocal Action and Fusion for Single Image Super-Resolution | 3766 |
| <i>Shuting Dong, Feng Lu, Chun Yuan</i> | |
| Augmenting Transformer-Transducer Based Speaker Change Detection with Token-Level Training Loss | 3771 |
| <i>Guanlong Zhao, Quan Wang, Han Lu, Yiling Huang, Ignacio Lopez Moreno</i> | |
| Parasympathetic-Sympathetic Causal Interactions and Perceived Workload for Varying Difficulty Affective Computing Tasks | 3776 |
| <i>Pravallika Lavanuru, Sawon Pratiher, Karuna P. Sahoo, Mrinal Acharya, Sreejith S, Nirmalya Ghosh, Amit Patra</i> | |
| Optimizing Quantum Federated Learning Based on Federated Quantum Natural Gradient Descent | 3781 |
| <i>Jun Qi, Xiao-Lei Zhang, Javier Tejedor</i> | |
| DL-NET: Dilation Location Network for Temporal Action Detection..... | 3786 |
| <i>Dianlong You, Houlin Wang, Bingxin Liu, Yang Yu, Zhiming Li</i> | |
| TAPLoss: A Temporal Acoustic Parameter Loss for Speech Enhancement | 3791 |
| <i>Yunyang Zeng, Joseph Konan, Shuo Han, David Bick, Muqiao Yang, Anurag Kumar, Shinji Watanabe, Bhiksha Raj</i> | |
| Adaptive Submanifold-Preserving Sparse Regression for Feature Selection and Multiclass Classification..... | 3796 |
| <i>Rui Xu, Xun Liang</i> | |

| | |
|--|------|
| Improving CTC-Based ASR Models with Gated Interlayer Collaboration | 3801 |
| <i>Yuting Yang, Yuke Li, Binbin Du</i> | |
| Reducing Language Confusion for Code-Switching Speech Recognition with Token-Level Language Diarization | 3806 |
| <i>Hexin Liu, Haihua Xu, Leibny Paola Garcia, Andy W. H. Khong, Yi He, Sanjeev Khudanpur</i> | |
| Training Stronger Spiking Neural Networks with Biomimetic Adaptive Internal Association Neurons | 3811 |
| <i>Haibo Shen, Yihao Luo, Xiang Cao, Liangqi Zhang, Juyi Xiao, Tianjiang Wang</i> | |
| ISmallNet: Densely Nested Network with Label Decoupling for Infrared Small Target Detection | 3816 |
| <i>Zhiheng Hu, Yongzhen Wang, Peng Li, Jie Qin, Haoran Xie, Mingqiang Wei</i> | |
| Hierarchical Network with Decoupled Knowledge Distillation for Speech Emotion Recognition..... | 3821 |
| <i>Ziping Zhao, Huan Wang, Haishuai Wang, Björn Schuller</i> | |
| UAV Remote Sensing Image Dehazing Based on Multi-Dimensional Saliency Awareness Unequal Network..... | 3826 |
| <i>Ruohui Zheng, Libao Zhang</i> | |
| A Two-Branch Network for Video Anomaly Detection with Spatio-Temporal Feature Learning..... | 3831 |
| <i>Guoqiu Li, Shengjie Chen, Yujiu Yang, Zhenhua Guo</i> | |
| Zero-Shot Personalized Lip-To-Speech Synthesis with Face Image Based Voice Control | 3836 |
| <i>Zheng-Yan Sheng, Yang Ai, Zhen-Hua Ling</i> | |
| Speech Reconstruction from Silent Tongue and Lip Articulation by Pseudo Target Generation and Domain Adversarial Training | 3841 |
| <i>Rui-Chen Zheng, Yang Ai, Zhen-Hua Ling</i> | |
| Cross-Domain Object Classification Via Successive Subspace Alignment..... | 3846 |
| <i>Kecheng Chen, Haoliang Li, Hong Yan</i> | |
| Prosody-Aware SpeechT5 for Expressive Neural TTS | 3851 |
| <i>Yan Deng, Long Zhou, Yuanhao Yi, Shujie Liu, Lei He</i> | |
| DEHRFormer: Real-Time Transformer for Depth Estimation and Haze Removal from Varicolored Haze Scenes..... | 3856 |
| <i>Sixiang Chen, Tian Ye, Jun Shi, Yun Liu, Jingxia Jiang, Erkang Chen, Peng Chen</i> | |
| Dual-Path Cross-Modal Attention for Better Audio-Visual Speech Extraction..... | 3861 |
| <i>Zhongweiyang Xu, Xulin Fan, Mark Hasegawa-Johnson</i> | |
| MSP-Former: Multi-Scale Projection Transformer for Single Image Desnowing | 3866 |
| <i>Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Jingxia Jiang, Erkang Chen, Peng Chen</i> | |
| Evaluating Variants of Wav2vec 2.0 on Affective Vocal Burst Tasks..... | 3871 |
| <i>Bagus Tris Atmaja, Akira Sasou</i> | |
| Robust Time Series Recovery and Classification Using Test-Time Noise Simulator Networks | 3876 |
| <i>Eun Som Jeon, Suhas Lohit, Rushil Anirudh, Pavan Turaga</i> | |
| Core: Transferable Long-Range Time Series Forecasting Enhanced by Covariates-Guided Representation | 3881 |
| <i>Xin-Yi Li, Pei-Nan Zhong, Di Chen, Yu-Bin Yang</i> | |

| | |
|--|------|
| PFT-SSR: Parallax Fusion Transformer for Stereo Image Super-Resolution | 3886 |
| <i>Hansheng Guo, Juncheng Li, Guangwei Gao, Zhi Li, Tiejong Zeng</i> | |
| Neural Optimization of Geometry and Fixed Beamformer for Linear Microphone Arrays | 3891 |
| <i>Longfei Yan, Weilong Huang, W. Bastiaan Kleijn, Thushara D. Abhayapala</i> | |
| ECGT2T: Towards Synthesizing Twelve-Lead Electrocardiograms from Two Asynchronous Leads | 3896 |
| <i>Yong-Yeon Jo, Young Sang Choi, Jong-Hwan Jang, Joon-Myoung Kwon</i> | |
| Auxiliary Pooling Layer for Spoken Language Understanding | 3901 |
| <i>Yukun Ma, Trung Hieu Nguyen, Jinjie Ni, Wen Wang, Qian Chen, Chong Zhang, Bin Ma</i> | |
| Bilateral Coarse-To-Fine Network for Point Cloud Completion | 3906 |
| <i>Tran Thanh Phong Nguyen, Son Lam Phung, Vinod Gopaldasani, Jane Whitelaw</i> | |
| SC-Net: Salient Point and Curvature Based Adversarial Point Cloud Generation Network | 3911 |
| <i>Zihao Zhang, Nan Sang, Xupeng Wang, Mumuxin Cai</i> | |
| Light Projection-Based Physical-World Vanishing Attack Against Car Detection | 3916 |
| <i>Huixiang Wen, Shan Chang, Luo Zhou</i> | |
| A Dual-Branch Adaptive Distribution Fusion Framework for Real-World Facial Expression Recognition | 3921 |
| <i>Shu Liu, Yan Xu, Tongming Wan, Xiaoyan Kui</i> | |
| Face Recognition on Point Cloud with Cgan-Top for Denoising | 3926 |
| <i>Junyu Liu, Jianfeng Ren, Hongliang Sun, Xudong Jiang</i> | |
| Self-Paced Partial Domain-Aware Learning for Face Anti-Spoofing | 3931 |
| <i>Zhiyi Chen, Yao Lu, Xinzhe Deng, Jia Meng, Shengchuan Zhang, Liujuan Cao</i> | |
| Adaptive Knowledge Distillation Between Text and Speech Pre-Trained Models | 3936 |
| <i>Jinjie Ni, Yukun Ma, Wen Wang, Qian Chen, Dianwen Ng, Han Lei, Trung Hieu Nguyen, Chong Zhang, Bin Ma, Erik Cambria</i> | |
| A Simple Yet Effective Approach to Structured Knowledge Distillation | 3941 |
| <i>Wenye Lin, Yangming Li, Lemao Liu, Shuming Shi, Hai-Tao Zheng</i> | |
| SQA: Strong Guidance Query with Self-Selected Attention for Human-Object Interaction Detection | 3946 |
| <i>Feng Zhang, Liu Sheng, Bingnan Guo, Ruixiang Chen, Junhao Chen</i> | |
| Picking the Underused Heads: A Network Pruning Perspective of Attention Head Selection for Fusing Dialogue Coreference Information | 3951 |
| <i>Zhengyuan Liu, Nancy F. Chen</i> | |
| LIMI-VC: A Light Weight Voice Conversion Model with Mutual Information Disentanglement | 3956 |
| <i>Liangjie Huang, Tian Yuan, Yunming Liang, Zeyu Chen, Can Wen, Yanlu Xie, Jinsong Zhang, Dengfeng Ke</i> | |
| FCIR: Rethink Aerial Image Super Resolution with Fourier Analysis | 3961 |
| <i>Yan Zhang, Pengcheng Zheng, Jianan Jiang, Pu Xiao, Xinbo Gao</i> | |
| Selective Film Conditioning with CTC-Based ASR Probability for Speech Enhancement | 3966 |
| <i>Da-Hee Yang, Joon-Hyuk Chang</i> | |

| | |
|--|------|
| Improved Deep Speaker Localization and Tracking: Revised Training Paradigm and Controlled Latency | 3971 |
| <i>Alexander Bohlander, Liesbeth Roelens, Nilesh Madhu</i> | |
| Multi-Layer Feature Division Transferable Adversarial Attack | 3976 |
| <i>Zikang Jin, Changchun Yin, Piji Li, Lu Zhou, Liming Fang, Xiangmao Chang, Zhe Liu</i> | |
| Improving Noisy Student Training on Non-Target Domain Data for Automatic Speech Recognition..... | 3981 |
| <i>Yu Chen, Wen Ding, Junjie Lai</i> | |
| Clicker: Attention-Based Cross-Lingual Commonsense Knowledge Transfer..... | 3986 |
| <i>Ruolin Su, Zhongkai Sun, Sixing Lu, Chengyuan Ma, Chenlei Guo</i> | |
| Efficient Data Loading with Quantum Autoencoder | 3991 |
| <i>Siang-Ruei Wu, Chun-Tse Li, Hao-Chung Cheng</i> | |
| Spteaec: A Soft Prompt Transfer Model for Zero-Shot Cross-Lingual Event Argument Extraction..... | 3996 |
| <i>Huipeng Ma, Qiu Tang, Ni Zhang, Rui Xu, Yanhua Shao, Wei Yan, Yaojun Wang</i> | |
| Strategies for Enhanced Signal Modulation Classifications Under Unknown Symbol Rates and Noise Conditions | 4001 |
| <i>Ruixuan Wang, Yue Qi, Mojtaba Vaezi, Xun Jiao, Moeness Amin</i> | |
| Knowledge Distillation with Active Exploration and Self-Attention Based Inter-Class Variation Transfer for Image Segmentation | 4006 |
| <i>Yifan Zhang, Shaojie Li, Xuan Yang</i> | |
| Oct Image Blind Despeckling Based on Gradient Guided Filter with Speckle Statistical Prior..... | 4011 |
| <i>Sanqian Li, Muxing Xiong, Bing Yang, Xiaoqing Zhang, Risa Higashita, Jiang Liu</i> | |
| Convolution-Based Channel-Frequency Attention for Text-Independent Speaker Verification..... | 4016 |
| <i>Jingyu Li, Yusheng Tian, Tan Lee</i> | |
| Streaming Multi-Channel Speech Separation with Online Time-Domain Generalized Wiener Filter | 4021 |
| <i>Yi Luo</i> | |
| Invariant Adversarial Imitation Learning from Visual Inputs..... | 4026 |
| <i>Haoran Zhang, Yinghong Tian, Liang Yuan, Yue Lu</i> | |
| A Distributed Adaptive Algorithm for Non-Smooth Spatial Filtering Problems..... | 4031 |
| <i>Charles Hovine, Alexander Bertrand</i> | |
| Performance of Social Machine Learning Under Limited Data | 4036 |
| <i>Ping Hu, Virginia Bordignon, Mert Kayaalp, Ali H. Sayed</i> | |
| Single-Branch Network for Multimodal Training | 4041 |
| <i>Muhammad Saad Saeed, Shah Nawaz, Muhammad Haris Khan, Muhammad Zaigham Zaheer, Karthik Nandakumar, Muhammad Haroon Yousaf, Arif Mahmood</i> | |
| Structure-Aware Multi-Feature Co-Learning for Dual Branch Face Super Resolution..... | 4046 |
| <i>Kangli Zeng, Zhongyuan Wang, Tao Lu, Jianyu Chen</i> | |
| Measuring Deviation from Stochasticity in Time-Series Using Autoencoder Based Time-Invariant Representation: Application to Black Hole Data..... | 4051 |
| <i>Chakka Sai Pradeep, Neelam Sinha, Banibrata Mukhopadhyay</i> | |

| | |
|---|------|
| Multi-Modal Domain Generalization for Cross-Scene Hyperspectral Image Classification..... | 4056 |
| <i>Yuxiang Zhang, Mengmeng Zhang, Wei Li, Ran Tao</i> | |
| Multi-Speaker Expressive Speech Synthesis Via Multiple Factors Decoupling..... | 4061 |
| <i>Xinfa Zhu, Yi Lei, Kun Song, Yongmao Zhang, Tao Li, Lei Xie</i> | |
| Multi-Speaker and Wide-Band Simulated Conversations as Training Data for End-to-End Neural Diarization..... | 4066 |
| <i>Federico Landini, Mireia Diez, Alicia Lozano-Diez, Lukáš Burget</i> | |
| Spatio-Temporal Structure Consistency for Semi-Supervised Medical Image Classification..... | 4071 |
| <i>Wentao Lei, Lei Liu, Li Liu</i> | |
| Hypothesis Test for Leakage Detection in Water Pipelines with High-Dimensional Sensor Signals | 4076 |
| <i>Liusha Yang, Matthew R. McKay, Xun Wang</i> | |
| Procter: Pronunciation-Aware Contextual Adapter for Personalized Speech Recognition in Neural Transducers..... | 4081 |
| <i>Rahul Pandey, Roger Ren, Qi Luo, Jing Liu, Ariya Rastrow, Ankur Gandhe, Denis Filimonov, Grant Strimel, Andreas Stolcke, Ivan Bulyko</i> | |
| Bebert: Efficient and Robust Binary Ensemble Bert | 4086 |
| <i>Jiayi Tian, Chao Fang, Haonan Wang, Zhongfeng Wang</i> | |
| End-To-End Unsupervised Sketch to Image Generation | 4091 |
| <i>Xingming Lv, Lei Wu, Zhenwei Cheng, Xiangxu Meng</i> | |
| Regularized Deep Generative Model Learning for Real-Time Massive MIMO Channel Tracking | 4096 |
| <i>Lixiang Lian, Ben Wang</i> | |
| Attention Localness in Shared Encoder-Decoder Model for Text Summarization | 4101 |
| <i>Li Huang, Hongmei Wu, Qiang Gao, Guisong Liu</i> | |
| PU-Edgeformer: Edge Transformer for Dense Prediction in Point Cloud Upsampling | 4106 |
| <i>Dohoon Kim, Minwoo Shin, Joonki Paik</i> | |
| LSSSED: A Robust Segmentation Network for Inflamed Appendix from CT Images..... | 4111 |
| <i>Wing W Y. Ng, Peixin Zheng, Ting Wang, Jianjun Zhang, Yinhao Liang, Hui Zhou, Dan Liang, Guangming Li, Xinhua Wei</i> | |
| Tree-Like Interaction Learning for Bundle Recommendation..... | 4116 |
| <i>Haole Ke, Lin Li, Peipei Wang, Jingling Yuan, Xiaohui Tao</i> | |
| IfUNet++: Iterative Feedback UNet++ for Infrared Small Target Detection..... | 4121 |
| <i>Zhangying Weng, Peng Li, Xin Zhuang, Xuefeng Yan, Lina Gong, Haoran Xie, Mingqiang Wei</i> | |
| MmSense: Detecting Concealed Weapons with a Miniature Radar Sensor..... | 4126 |
| <i>Kevin Mitchell, Khaled Kassem, Chaitanya Kaul, Valentin Kapitanov, Philip Binner, Andrew Ramsay, Daniele Faccio, Roderick Murray-Smith</i> | |
| Improving Learning Objectives for Speaker Verification from the Perspective of Score Comparison | 4131 |
| <i>Min Hyun Han, Sung Hwan Mun, Minchan Kim, Myeonghun Jeong, Sung Hwan Ahn, Nam Soo Kim</i> | |
| Adapter Tuning with Task-Aware Attention Mechanism..... | 4136 |
| <i>Jinliang Lu, Feihu Jin, Jiajun Zhang</i> | |

| | |
|--|------|
| Twitter Stance Detection Via Neural Production Systems..... | 4141 |
| <i>Bowen Zhang, Daijun Ding, Guangning Xu, Jinjin Guo, Zhichao Huang, Xu Huang</i> | |
| Zero-Shot Sound Event Classification Using a Sound Attribute Vector with Global and Local Feature Learning..... | 4146 |
| <i>Yi-Han Lin, Xunquan Chen, Ryoichi Takashima, Tetsuya Takiguchi</i> | |
| Optimization for Robustness Evaluation Beyond ℓ_p Metrics..... | 4151 |
| <i>Hengyue Liang, Buyun Liang, Ying Cui, Tim Mitchell, Ju Sun</i> | |
| Semi-Supervised Contrastive Learning with Soft Mask Attention for Facial Action Unit Detection..... | 4156 |
| <i>Zhongling Liu, Rujie Liu, Ziqiang Shi, Liu Liu, Xiaoyu Mi, Kentaro Murase</i> | |
| Speech Enhancement with Intelligent Neural Homomorphic Synthesis | 4161 |
| <i>Shulin He, Wei Rao, Jinjiang Liu, Jun Chen, Yukai Ju, Xueliang Zhang, Yannan Wang, Shidong Shang</i> | |
| DeepSpace: Dynamic Spatial and Source CUE Based Source Separation for Dialog Enhancement..... | 4166 |
| <i>Aaron Master, Lie Lu, Jonas Samuelsson, Heidi-Maria Lehtonen, Scott Norcross, Nathan Swedlow, Audrey Howard</i> | |
| Multiple Contrastive Learning for Multimodal Sentiment Analysis | 4171 |
| <i>Xiaocui Yang, Shi Feng, Daling Wang, Pengfei Hong, Soujanya Poria</i> | |
| Towards Real-Time Person Search with Invariant Feature Learning | 4176 |
| <i>Chengyou Jia, Minnan Luo, Zhuohang Dang, Xiaojun Chang, Qinghua Zheng</i> | |
| Unsupervised Fine-Tuning Data Selection for ASR Using Self-Supervised Speech Models..... | 4181 |
| <i>Reem Gody, David Harwath</i> | |
| PCF: ECAPA-TDNN with Progressive Channel Fusion for Speaker Verification | 4186 |
| <i>Zhenduo Zhao, Zhuo Li, Wenchao Wang, Pengyuan Zhang</i> | |
| Short-Segment Speaker Verification Using ECAPA-TDNN with Multi-Resolution Encoder..... | 4191 |
| <i>Sangwook Han, Youngdo Ahn, Kyeongmuk Kang, Jong Won Shin</i> | |
| AST-SED: An Effective Sound Event Detection Method Based on Audio Spectrogram Transformer | 4196 |
| <i>Kang Li, Yan Song, Li-Rong Dai, Ian McLoughlin, Xin Fang, Lin Liu</i> | |
| Ψ -Net: Point Structural Information Network for No-Reference Point Cloud Quality Assessment..... | 4201 |
| <i>Jian Xiong, Sifan Wu, Wang Luo, Jinli Suo, Hao Gao</i> | |
| Inter-Subnet: Speech Enhancement with Subband Interaction..... | 4206 |
| <i>Jun Chen, Wei Rao, Zilin Wang, Jiuxin Lin, Zhiyong Wu, Yannan Wang, Shidong Shang, Helen Meng</i> | |
| Audio Cross Verification Using Dual Alignment Likelihood Ratio Test..... | 4211 |
| <i>Heidi Lei, Arm Wonghirundacha, Irmak Bukey, T J Tsai</i> | |
| Gender-Cartoon: Image Cartoonization Method Based on Gender Classification | 4216 |
| <i>Long Feng, Guohua Geng, Chen Guo, Longquan Yan, Xingrui Ma, Zhan Li, Kang Li</i> | |
| CADET: Control-Aware Dynamic Edge Computing for Real-Time Target Tracking in UAV Systems..... | 4221 |
| <i>Luis Felipe Florenzan Reyes, Francesco Smarra, Alessandro D'Innocenzo, Marco Levorato</i> | |

| | |
|--|------|
| Select the Best: Enhancing Graph Representation with Adaptive Negative Sample Selection | 4226 |
| <i>Xiangping Zheng, Xun Liang, Bo Wu</i> | |
| Regression to Classification: Waveform Encoding for Neural Field-Based Audio Signal Representation | 4231 |
| <i>Taesoo Kim, Daniel Rho, Gahui Lee, Jaehan Park, Jong Hwan Ko</i> | |
| Exploiting Virtual Array Diversity for Accurate Radar Detection..... | 4236 |
| <i>Junfeng Guan, Sohrab Madani, Waleed Ahmed, Samah Hussein, Saurabh Gupta, Haitham Hassanieh</i> | |
| Nowcasting of Extreme Precipitation Using Deep Generative Models..... | 4241 |
| <i>Haoran Bi, Maksym Kyryliuk, Zhiyi Wang, Cristian Meo, Yanbo Wang, Ruben Imhoff, Remko Uijlenhoet, Justin Dauwels</i> | |
| Deep Spatio-Temporal Multiplex Graph Learning for Cardiac Imaging Classification..... | 4246 |
| <i>Jaume Banus, Augustin Ogier, Roger Hullin, Philippe Meyer, Ruud B. Van Heeswijk, Jonas Richiardi</i> | |
| JNDMix: Jnd-Based Data Augmentation for No-Reference Image Quality Assessment | 4251 |
| <i>Jiamu Sheng, Jiayuan Fan, Peng Ye, Jianjian Cao</i> | |
| Deep Generative Fixed-Filter Active Noise Control | 4256 |
| <i>Zhengding Luo, Dongyuan Shi, Xiaoyi Shen, Junwei Ji, Woon-Seng Gan</i> | |
| Leveraging Neural Koopman Operators to Learn Continuous Representations of Dynamical Systems from Scarce Data..... | 4261 |
| <i>Anthony Frion, Lucas Drumetz, Mauro Dalla Mura, Guillaume Tochon, Abdeldjalil Aïssa-El-Bey</i> | |
| Decoding Musical Pitch from Human Brain Activity with Automatic Voxel-Wise Whole-Brain FMRI Feature Selection | 4266 |
| <i>Vincent K. M. Cheung, Yueh-Po Peng, Jing-Hua Lin, Li Su</i> | |
| RDO Candidate Selection for Maximizing Coding Efficiency in a Practical HEVC Encoder..... | 4271 |
| <i>Joose Sainio, Alexandre Mercat, Jarno Vanne</i> | |
| Residual Hybrid Attention Network for Compression Artifact Reduction | 4276 |
| <i>Bingchun Luo, Wei Yu</i> | |
| Robust Binary Component Decompositions | 4281 |
| <i>Christos Kolomvakis, Nicolas Gillis</i> | |
| A Computationally Efficient Algorithm for Distributed Adaptive Signal Fusion Based on Fractional Programs..... | 4286 |
| <i>Cem Ates Musluoglu, Alexander Bertrand</i> | |
| SL-MoE: A Two-Stage Mixture-of-Experts Sequence Learning Framework for Forecasting Rapid Intensification of Tropical Cyclone | 4291 |
| <i>Jian Xu, Yang Lei, Guangqi Zhu, Yunling Feng, Bo Xiao, Qifeng Qian, Yajing Xu</i> | |
| Personalized Federated Learning on Long-Tailed Data Via Adversarial Feature Augmentation..... | 4296 |
| <i>Yang Lu, Pinxin Qian, Gang Huang, Hanzi Wang</i> | |
| Autotts: End-To-End Text-To-Speech Synthesis Through Differentiable Duration Modeling..... | 4301 |
| <i>Bac Nguyen, Fabien Cardinaux, Stefan Uhlich</i> | |

| | |
|---|------|
| Semi-Supervised Learning with Per-Class Adaptive Confidence Scores for Acoustic Environment Classification with Imbalanced Data | 4306 |
| <i>Luan Vinicius Fiorio, Boris Karanov, Johan David, Wim Van Houtum, Frans Widdershoven, Ronald M. Aarts</i> | |
| Central Nodes Detection from Partially Observed Graph Signals..... | 4311 |
| <i>Yiran He, Hoi-To Wai</i> | |
| Compressive Channel Estimation for IRS-Aided Millimeter-Wave Systems Via Two-Stage Lamp Network..... | 4316 |
| <i>Wen-Chiao Tsai, Chi-Wei Chen, An-Yeu Andy Wu</i> | |
| A Discriminative Multi-Channel Noise Feature Representation Model for Image Manipulation Localization | 4321 |
| <i>Yang Zhou, Hongxia Wang, Qiang Zeng, Rui Zhang, Sijiang Meng</i> | |
| A Practical Distributed Active Noise Control Algorithm Overcoming Communication Restrictions | 4326 |
| <i>Junwei Ji, Dongyuan Shi, Zhengding Luo, Xiaoyi Shen, Woon-Seng Gan</i> | |
| Group-Wise Co-Salient Object Detection with Siamese Transformers Via Brownian Distance Covariance Matching..... | 4331 |
| <i>Yang Wu, Hao Zhang, Lingyan Liang, Yaqian Zhao, Kaihua Zhang</i> | |
| A Content-Based Multi-Scale Network for Single Image Super-Resolution..... | 4336 |
| <i>Jiahuan Ji, Baojiang Zhong, Kai-Kuang Mu</i> | |
| Cross-Modal Fusion Techniques for Utterance-Level Emotion Recognition from Text and Speech | 4341 |
| <i>Jiachen Luo, Huy Phan, Joshua Reiss</i> | |
| Lightvessel: Exploring Lightweight Coronary Artery Vessel Segmentation Via Similarity Knowledge Distillation..... | 4346 |
| <i>Hao Dang, Yuekai Zhang, Xingqun Qi, Wanting Zhou, Muyi Sun</i> | |
| Enhancing Robustness and Imperceptibility of Blind Watermarking with Improved Message Processor | 4351 |
| <i>Yufeng Wu, Baowei Wang, Changyu Dai, Yi Yuan, Bin Li, Weiqian Zheng, Hao Wu</i> | |
| Prior-Enhanced Temporal Action Localization Using Subject-Aware Spatial Attention..... | 4356 |
| <i>Yifan Liu, Youbao Tang, Ning Zhang, Rwei-Sung Lin, Haoqian Wang</i> | |
| Dynamic Independent Component Extraction with Blending Mixing Vector: Lower Bound on Mean Interference-To-Signal Ratio | 4361 |
| <i>Jaroslav Cmejla, Zbynek Koldovský, Václav Kautský, Tülay Adalı</i> | |
| Integrating Syntactic and Semantic Knowledge in AMR Parsing with Heterogeneous Graph Attention Network..... | 4366 |
| <i>Yikemaiti Sataer, Chuanqi Shi, Miao Gao, Yunlong Fan, Bin Li, Zhiqiang Gao</i> | |
| Contrastive Learning at the Relation and Event Level for Rumor Detection | 4371 |
| <i>Yingrui Xu, Jingyuan Hu, Jingguo Ge, Yulei Wu, Tong Li, Hui Li</i> | |
| Multimodal Propaganda Detection Via Anti-Persuasion Prompt Enhanced Contrastive Learning | 4376 |
| <i>Jian Cui, Lin Li, Xin Zhang, Jingling Yuan</i> | |
| Multitask Detection of Speaker Changes, Overlapping Speech and Voice Activity Using Wav2vec 2.0..... | 4381 |
| <i>Marie Kunešová, Zbynek Zajic</i> | |

| | |
|--|------|
| Blood Oxygen Saturation Estimation from Facial Video Via DC and AC Components of Spatio-Temporal Map | 4386 |
| <i>Yusuke Akamatsu, Yoshifumi Onishi, Hitoshi Imaoka</i> | |
| Ensemble of Deep Neural Network Models for MOS Prediction | 4391 |
| <i>Marie Kunešová, Jindrich Matoušek, Jan Lehečka, Jan Švec, Josef Michálek, Daniel Tihelka, Martin Bulín, Zdenek Hanzlíček, Markéta Rezáčková</i> | |
| Using Auxiliary Tasks in Multimodal Fusion of Wav2vec 2.0 and Bert for Multimodal Emotion Recognition | 4396 |
| <i>Dekai Sun, Yancheng He, Jiqing Han</i> | |
| A2S-NAS: Asymmetric Spectral-Spatial Neural Architecture Search for Hyperspectral Image Classification | 4401 |
| <i>Lin Zhan, Jiayuan Fan, Peng Ye, Jianjian Cao</i> | |
| NCL: Textual Backdoor Defense Using Noise-Augmented Contrastive Learning | 4406 |
| <i>Shengfang Zhai, Qingni Shen, Xiaoyi Chen, Weilong Wang, Cong Li, Yuejian Fang, Zhonghai Wu</i> | |
| Prefallkd: Pre-Impact Fall Detection Via CNN-ViT Knowledge Distillation..... | 4411 |
| <i>Tin-Han Chi, Kai-Chun Liu, Chia-Yeh Hsieh, Yu Tsao, Chia-Tai Chan</i> | |
| Towards Reducing Patient Effort for the Automatic Prediction of Speech Intelligibility in Head and Neck Cancers..... | 4416 |
| <i>Sebastião Quintas, Alberto Abad, Julie Mauclair, Virginie Woisard, Julien Piquier</i> | |
| Integrated Sensing and Full-Duplex Communication: Joint Transceiver Beamforming and Power Allocation | 4421 |
| <i>Zhenyao He, Wei Xu, Hong Shen, Derrick Wing Kwan Ng, Yonina C. Eldar, Xiaohu You</i> | |
| Ripple Sparse Self-Attention for Monaural Speech Enhancement..... | 4426 |
| <i>Qiquan Zhang, Hongxu Zhu, Qi Song, Xinyuan Qian, Zhaoheng Ni, Haizhou Li</i> | |
| Spectral Super-Resolution on the Unit Circle Via Gradient Descent..... | 4431 |
| <i>Xunmeng Wu, Zai Yang, Jian-Feng Cai, Zongben Xu</i> | |
| Interweaved Graph and Attention Network for 3D Human Pose Estimation | 4436 |
| <i>Ti Wang, Hong Liu, Runwei Ding, Wenhao Li, Yingxuan You, Xia Li</i> | |
| Scoreformer: Score Fusion-Based Transformers for Weakly-Supervised Violence Detection | 4441 |
| <i>Yang Xiao, Liejun Wang, Tongguan Wang, Huicheng Lai</i> | |
| Skillnet-NLG: General-Purpose Natural Language Generation with a Sparsely Activated Approach | 4446 |
| <i>Junwei Liao, Duyu Tang, Fan Zhang, Shuming Shi</i> | |
| Semantic Memory Guided Image Representation for Polyp Segmentation | 4451 |
| <i>Zijin Yin, Runpu Wei, Kongming Liang, Yiyang Lin, Wei Liu, Zhanyu Ma, Min Min, Jun Guo</i> | |
| OAFFormer: Learning Occlusion Distinguishable Feature for Amodal Instance Segmentation | 4456 |
| <i>Zhixuan Li, Ruohua Shi, Tiejun Huang, Tingting Jiang</i> | |
| Rethink Long-Tailed Recognition with Vision Transforms..... | 4461 |
| <i>Zhengzhuo Xu, Shuo Yang, Xingjun Wang, Chun Yuan</i> | |

| | |
|---|------|
| Multi-Task Transformer with Relation-Attention and Type-Attention for Named Entity Recognition..... | 4466 |
| <i>Ying Mo, Hongyin Tang, Jiahao Liu, Qifan Wang, Zenglin Xu, Jingang Wang, Wei Wu, Zhoujun Li</i> | |
| Regularized Neural Detection for Millimeter Wave Massive MIMO Communication Systems with One-Bit ADCs | 4471 |
| <i>Aditya Sant, Bhaskar D. Rao</i> | |
| Decoupled Non-Parametric Knowledge Distillation for end-to-End Speech Translation | 4476 |
| <i>Hao Zhang, Nianwen Si, Yaqi Chen, Wenlin Zhang, Xukui Yang, Dan Qu, Zhen Li</i> | |
| Smoothing Complex-Valued Signals on Graphs with Monte-Carlo | 4481 |
| <i>Hugo Jaquard, Michaël Fanuel, Pierre-Olivier Amblard, Rémi Bardenet, Simon Barthelmé, Nicolas Tremblay</i> | |
| Constrained Dynamical Neural ODE for Time Series Modelling: A Case Study on Continuous Emotion Prediction..... | 4486 |
| <i>Ting Dang, Antoni Dimitriadis, Jingyao Wu, Vidhyasaharan Sethu, Eliathamby Ambikairajah</i> | |
| Parameter Efficient Transfer Learning for Various Speech Processing Tasks | 4491 |
| <i>Shinta Otake, Rei Kawakami, Nakamasa Inoue</i> | |
| Distortion-Aware Convolutional Neural Network-Based Interpolation Filter for AVS3 | 4496 |
| <i>Ying Zhang, Liang Wen, Lizhong Wang, Yinji Piao, Weijing Shi, Kwang Pyo Choi</i> | |
| YOLO-Based Lightweight Object Detection with Structure Simplification and Attention Enhancement | 4501 |
| <i>Shuqi Sun, Xiaohui Yang, Jingliang Peng</i> | |
| UFO2: A Unified Pre-Training Framework for Online and Offline Speech Recognition..... | 4506 |
| <i>Li Fu, Siqi Li, Qingtao Li, Liping Deng, Fangzhu Li, Lu Fan, Meng Chen, Xiaodong He</i> | |
| Lightweight and High-Fidelity End-to-End Text-to-Speech with Multi-Band Generation and Inverse Short-Time Fourier Transform..... | 4511 |
| <i>Masaya Kawamura, Yuma Shirahata, Ryuichi Yamamoto, Kentaro Tachibana</i> | |
| Self-Adaptive Incremental Machine Speech Chain for Lombard TTS with High-Granularity ASR Feedback in Dynamic Noise Condition..... | 4516 |
| <i>Sashi Novitasari, Sakriani Sakti, Satoshi Nakamura</i> | |
| Cooperative Five Degrees of Freedom Motion Estimation for a Swarm of Autonomous Vehicles..... | 4521 |
| <i>Nikos Piperigkos, Aris S. Lalos, Kostas Berberidis, Christos Anagnostopoulos</i> | |
| LongFNT: Long-Form Speech Recognition with Factorized Neural Transducer..... | 4526 |
| <i>Xun Gong, Yu Wu, Jinyu Li, Shujie Liu, Rui Zhao, Xie Chen, Yanmin Qian</i> | |
| RL-IFF: Indoor Localization Via Reinforcement Learning-Based Information Fusion | 4531 |
| <i>Mohammad Salimibeni, Arash Mohammadi</i> | |
| Feature Space Recovery for Incomplete Multi-View Clustering..... | 4536 |
| <i>Zhen Long, Ce Zhu, Pierre Comon, Yipeng Liu</i> | |
| On the Joint Estimation of Phase Noise and time-Varying Channels for OFDM Under High-Mobility Conditions..... | 4541 |
| <i>Francesco Linsalata, Nassar Ksairi</i> | |

| | |
|--|------|
| 6G Integrated Sensing and Communication - Sensing Assisted Environmental Reconstruction and Communication | 4546 |
| <i>Zhi Zhou, Xianjin Li, Jia He, Xiaoyan Bi, Yan Chen, Guangjian Wang, Peiyong Zhu</i> | |
| A Novel Approach Based on Voronoï Cells to Classify Spectrogram Zeros of Multicomponent Signals | 4551 |
| <i>N. Laurent, S. Meignen, M. A. Colominas, J. M. Miramont, F. Auger</i> | |
| DTTR: Detecting Text with Transformers..... | 4556 |
| <i>Jing Yang, Zhiqiang You, Zhiwei Zhong, Peng Liu, Langqi Mei, Shenguang Huang</i> | |
| Learning to Explain: A Gradient-Based Attribution Method for Interpreting Super-Resolution Networks | 4561 |
| <i>Anni Yu, Yu-Bin Yang</i> | |
| Estimation of Cardiac Fibre Direction Based on Activation Maps..... | 4566 |
| <i>Johannes W. De Vries, Miao Sun, Natasja M. S. De Groot, Richard C. Hendriks</i> | |
| Bimodal Fusion Network for Basic Taste Sensation Recognition from Electroencephalography and Electromyography | 4571 |
| <i>Han Gao, Shuo Zhao, Huiyan Li, Li Liu, You Wang, Ruifen Hu, Jin Zhang, Guang Li</i> | |
| Pushing the Limits of Self-Supervised Speaker Verification Using Regularized Distillation Framework..... | 4576 |
| <i>Yafeng Chen, Siqi Zheng, Hui Wang, Luyao Cheng, Qian Chen</i> | |
| A Multi-Scale Feature Aggregation Based Lightweight Network for Audio-Visual Speech Enhancement | 4581 |
| <i>Haitao Xu, Liangfa Wei, Jie Zhang, Jianming Yang, Yannan Wang, Tian Gao, Xin Fang, Lirong Dai</i> | |
| Multi-Lingual Pronunciation Assessment with Unified Phoneme Set and Language-Specific Embeddings | 4586 |
| <i>Binghuai Lin, Liyuan Wang</i> | |
| Robust Multi-Modal Speech Emotion Recognition with ASR Error Adaptation | 4591 |
| <i>Binghuai Lin, Liyuan Wang</i> | |
| Matrix Recovery Using Deep Generative Priors with Low-Rank Deviations..... | 4596 |
| <i>Pengbin Yu, Jianjun Wang, Chen Xu</i> | |
| Spectro-Temporal Post-Filtering Via Short-Time Target Cancellation for Directional Speech Enhancement in a Dual-Microphone Hearing AID | 4601 |
| <i>Marcos A. Cantu, Volker Hohmann</i> | |
| OTW: Optimal Transport Warping for Time Series..... | 4606 |
| <i>Fabian Latorre, Chenghao Liu, Doyen Sahoo, Steven C. H. Hoi</i> | |
| Narrow Down Before Selection: A Dynamic Exclusion Model for Multiple-Choice QA..... | 4611 |
| <i>Xiyan Liu, Yidong Shi, Ruifang Liu, Ge Bai, Yanyi Chen</i> | |
| Voice Conversion Using Feature Specific Loss Function Based Self-Attentive Generative Adversarial Network | 4616 |
| <i>Sandipan Dhar, Padmanabha Banerjee, Nanda Dulal Jana, Swagatam Das</i> | |
| Improving Text-Audio Retrieval by Text-Aware Attention Pooling and Prior Matrix Revised Loss | 4621 |
| <i>Yifei Xin, Dongchao Yang, Yuexian Zou</i> | |

| | |
|---|------|
| OPT: One-Shot Pose-Controllable Talking Head Generation..... | 4626 |
| <i>Jin Liu, Xi Wang, Xiaomeng Fu, Yesheng Chai, Cai Yu, Jiao Dai, Jizhong Han</i> | |
| Exploration of Language Dependency for Japanese Self-Supervised Speech Representation Models | 4631 |
| <i>Takanori Ashihara, Takafumi Moriya, Kohei Matsuura, Tomohiro Tanaka</i> | |
| Multi-Modal ASR Error Correction with Joint ASR Error Detection | 4636 |
| <i>Binghuai Lin, Liyuan Wang</i> | |
| Joint Modeling for ASR Correction and Dialog State Tracking | 4641 |
| <i>Deyuan Wang, Tiantian Zhang, Caixia Yuan, Xiaojie Wang</i> | |
| Optimising Different Feature Types for Inpainting-Based Image Representations | 4646 |
| <i>Ferdinand Jost, Vassillen Chizhov, Joachim Weickert</i> | |
| Structured-Anchor Projected Clustering for Hyperspectral Images | 4651 |
| <i>Guozhu Jiang, Jie Zhang, Yongshan Zhang, Xinwei Jiang, Zhihua Cai</i> | |
| Target Velocity Estimation for Quantization-Based Cooperative MIMO Radar and Communications System | 4656 |
| <i>Zhen Wang, Xuedan Yan, Qian He, Rick S. Blum</i> | |
| Unified Prompt Learning Makes Pre-Trained Language Models Better Few-Shot Learners..... | 4661 |
| <i>Feihu Jin, Jinliang Lu, Jiajun Zhang</i> | |
| Self-Supervised Facial Action Unit Detection with Region and Relation Learning..... | 4666 |
| <i>Juan Song, Zhilei Liu</i> | |
| Self-Supervised Learning with Explorative Knowledge Distillation | 4671 |
| <i>Tongtong Su, Jinsong Zhang, Gang Wang, Xiaoguang Liu</i> | |
| Robust and Globally Sparse Pca Via Majorization-Minimization and Variable Splitting | 4676 |
| <i>Hugo Brehier, Arnaud Breloy, Mohammed Nabil El Korso, Sandeep Kumar</i> | |
| Boosting No-Reference Super-Resolution Image Quality Assessment with Knowledge Distillation and Extension | 4681 |
| <i>Haiyu Zhang, Shaolin Su, Yu Zhu, Jinqiu Sun, Yanning Zhang</i> | |
| Cutting Through the Noise: An Empirical Comparison of Psycho-Acoustic and Envelope-based Features for Machinery Fault Detection | 4686 |
| <i>Peter Wißbrock, Yvonne Richter, David Pelkmann, Zhao Ren, Gregory Palmer</i> | |
| Backdoor Attack Against Automatic Speaker Verification Models in Federated Learning | 4691 |
| <i>Dan Meng, Xue Wang, Jun Wang</i> | |
| Adapting Exploratory Behaviour in Active Inference for Autonomous Driving..... | 4696 |
| <i>Sheida Nozari, Ali Krayani, Pablo Marin, Lucio Marcenaro, David Martin, Carlo Regazzoni</i> | |
| Multi-Microphone Speaker Separation by Spatial Regions..... | 4701 |
| <i>Julian Wechsler, Srikanth Raj Chetupalli, Wolfgang Mack, Emanuël A. P. Habets</i> | |
| GTN-Bailando: Genre Consistent long-Term 3D Dance Generation Based on Pre-Trained Genre Token Network | 4706 |
| <i>Haolin Zhuang, Shun Lei, Long Xiao, Weiqin Li, Liyang Chen, Sicheng Yang, Zhiyong Wu, Shiyin Kang, Helen Meng</i> | |

| | |
|--|------|
| MSNet: A Deep Architecture Using Multi-Sentiment Semantics for Sentiment-Aware Image Style Transfer | 4711 |
| <i>Shikun Sun, Jia Jia, Haozhe Wu, Zijie Ye, Junliang Xing</i> | |
| CNN Filter for Super-Resolution with RPR Functionality in VVC | 4716 |
| <i>Shimin Huang, Cheolkon Jung, Yang Liu, Ming Li</i> | |
| Mimo Radar Transmit Beampattern Matching Via Manifold Optimization..... | 4721 |
| <i>Weijie Xiong, Jinfeng Hu, Kai Zhong</i> | |
| Interpretable Nonnegative Incoherent Deep Dictionary Learning for FMRI Data Analysis | 4726 |
| <i>Manuel Morante, Jan Østergaard, Sergios Theodoridis</i> | |
| Multi-Local Attention for Speech-Based Depression Detection..... | 4731 |
| <i>Fuxiang Tao, Xuri Ge, Wei Ma, Anna Esposito, Alessandro Vinciarelli</i> | |
| Daily Mental Health Monitoring from Speech: A Real-World Japanese Dataset and Multitask Learning Analysis..... | 4736 |
| <i>Meishu Song, Andreas Triantafyllopoulos, Ziji Yang, Hiroki Takeuchi, Toru Nakamura, Akifumi Kishi, Tetsuro Ishizawa, Kazuhiro Yoshiuchi, Xin Jing, Vincent Karas, Zhonghao Zhao, Kun Qian, Bin Hu, Björn W Schuller, Yoshiharu Yamamoto</i> | |
| Audio-Driven Talking Head Video Generation with Diffusion Model..... | 4741 |
| <i>Yizhe Zhua, Chunhui Zhanga, Qiong Liub, Xi Zhou</i> | |
| Joint Pre-Training with Speech and Bilingual Text for Direct Speech to Speech Translation..... | 4746 |
| <i>Kun Wei, Long Zhou, Ziqiang Zhang, Liping Chen, Shujie Liu, Lei He, Jinyu Li, Furu Wei</i> | |
| Tracking Targets in Hyper-Scale Cameras Using Movement Predication..... | 4751 |
| <i>Jiaping Yu, Tongqing Zhou, Zhiping Cai, Wenyuan Kuang</i> | |
| CNN Filter for RPR-Based SR in VVC with Wavelet Decomposition | 4756 |
| <i>Hui Lan, Cheolkon Jung, Yang Liu, Ming Li</i> | |
| Sketch Less Face Image Retrieval: A New Challenge..... | 4761 |
| <i>Dawei Dai, Yutang Li, Liang Wang, Shiyu Fu, Shuyin Xia, Guoyin Wang</i> | |
| Comparing Decentralized Gradient Descent Approaches and Guarantees | 4766 |
| <i>Shana Moothedath, Namrata Vaswani</i> | |
| IQGAN: Robust Quantum Generative Adversarial Network for Image Synthesis on NISQ Devices..... | 4771 |
| <i>Cheng Chu, Grant Skipper, Martin Swamy, Fan Chen</i> | |
| Fine-Grained Emotional Control of Text-to-Speech: Learning to Rank Inter- and Intra-Class Emotion Intensities..... | 4776 |
| <i>Shijun Wang, Jón Guðnason, Damian Borth</i> | |
| Sharing Low Rank Conformer Weights for Tiny Always-On Ambient Speech Recognition Models | 4781 |
| <i>Steven M. Hernandez, Ding Zhao, Shaojin Ding, Antoine Bruguier, Rohit Prabhavalkar, Tara N. Sainath, Yanzhang He, Ian McGraw</i> | |
| Anomaly Detection in Optical Spectra VIA Joint Optimization..... | 4786 |
| <i>Antonino Maria Rizzo, Luca Magri, Pietro Invernizzi, Enrico Sozio, Stefano Piciaccia, Alberto Tanzi, Stefano Binetti, Cesare Alippi, Giacomo Boracchi</i> | |
| Fast Single-Person 2D Human Pose Estimation Using Multi-Task Convolutional Neural Networks..... | 4791 |
| <i>Christos Papaioannidis, Ioannis Mademlis, Ioannis Pitas</i> | |

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

| | |
|--|------|
| Optimal Transport in Diffusion Modeling for Conversion Tasks in Audio Domain..... | 4881 |
| <i>Vadim Popov, Amantur Amatov, Mikhail Kudinov, Vladimir Gogoryan, Tasnima Sadekova, Ivan Vovk</i> | |
| Longshortnet: Exploring Temporal and Semantic Features Fusion in Streaming Perception..... | 4886 |
| <i>Chenyang Li, Zhi-Qi Cheng, Jun-Yan He, Pengyu Li, Bin Luo, Hanyuan Chen, Yifeng Geng, Jin-Peng Lan, Xuansong Xie</i> | |
| Wordreg: Mitigating the Gap Between Training and Inference with Worst-Case Drop Regularization | 4891 |
| <i>Jun Xia, Ge Wang, Bozhen Hu, Cheng Tan, Jiangbin Zheng, Yongjie Xu, Stan Z. Li</i> | |
| SS-ADMM: Stationary and Sparse Granger Causal Discovery for Cortico-Muscular Coupling | 4896 |
| <i>Farwa Abbas, Verity McClelland, Zoran Cvetkovic, Wei Dai</i> | |
| LP-IOANet: Efficient High Resolution Document Shadow Removal | 4901 |
| <i>Konstantinos Georgiadis, M. Kerim Yucel, Evangelos Skartados, Valia Dimaridou, Anastasios Drosou, Albert Saà-Garriga, Bruno Manganeli</i> | |
| Joint Ann-SNN Co-training for Object Localization and Image Segmentation | 4906 |
| <i>Marc Baltes, Nidal Abuhajar, Ye Yue, Charles D. Smith, Jundong Liu</i> | |
| Frame-Wise and Overlap-Robust Speaker Embeddings for Meeting Diarization | 4911 |
| <i>Tobias Cord-Landwehr, Christoph Boeddeker, Catalin Zorila, Rama Doddipatla, Reinhold Haeb-Umbach</i> | |
| Effective Graph-Based Modeling of Articulation Traits for Mispronunciation Detection and Diagnosis..... | 4916 |
| <i>Bi-Cheng Yan, Hsin-Wei Wang, Yi-Cheng Wang, Berlin Chen</i> | |
| Tensorized Neural Layer Decomposition for 2-D DOA Estimation..... | 4921 |
| <i>Hang Zheng, Chengwei Zhou, Sergiy A. Vorobyov, Zhiguo Shi</i> | |
| Non-Convex Approaches for Low-Rank Tensor Completion Under Tubal Sampling | 4926 |
| <i>Zheng Tan, Longxiu Huang, Hanqin Cai, Yifei Lou</i> | |
| Data Driven Joint Sensor Fusion and Regression Based on Geometric Mean Squared Error..... | 4931 |
| <i>Carlos Alejandro Lopez, Jaume Riba</i> | |
| Time-Resolved fMRI Shared Response Model Using Gaussian Process Factor Analysis | 4936 |
| <i>Mohammadreza Ebrahimi, Navona Calarco, Colin Hawco, Aristotle Voineskos, Ashish Khisti</i> | |
| Model Fingerprinting with Benign Inputs..... | 4941 |
| <i>Thibault Maho, Teddy Furon, Erwan Le Merrer</i> | |
| Automatic Camera Pose Estimation by Key-Point Matching of Reference Objects | 4946 |
| <i>Jinchen Zeng, Rick Butler, John J. Van Den Dobbela, Benno H. W. Hendriks, Maarten Van Der Elst, Justin Dauwels</i> | |
| To Regularize Or Not to Regularize: The Role of Positivity in Sparse Array Interpolation with a Single Snapshot | 4951 |
| <i>Mehmet Can Hüçümenoglu, Pulak Sarangi, Robin Rajamäki, Piya Pal</i> | |
| Style Modeling for Multi-Speaker Articulation-to-Speech..... | 4956 |
| <i>Miseul Kim, Zhenyu Piao, Jihyun Lee, Hong-Goo Kang</i> | |

| | |
|--|------|
| DRL Path Planning for UAV-Aided V2X Networks: Comparing Discrete to Continuous Action Spaces..... | 4961 |
| <i>Leonardo Spampinato, Alessia Tarozzi, Chiara Buratti, Riccardo Marini</i> | |
| On Super-Resolution with Separation Prior | 4966 |
| <i>Xingyun Mao, Heng Qiao</i> | |
| A Patient Invariant Model Towards the Prediction of Freezing of Gait..... | 4971 |
| <i>Nasimuddin Ahmed, Shivam Singhal, Aniruddha Sinha, Avik Ghose</i> | |
| Multi-Speaker Speech Synthesis from Electromyographic Signals by Soft Speech Unit Prediction | 4976 |
| <i>Kevin Scheck, Tanja Schultz</i> | |
| Deep Subband Network for Joint Suppression of Echo, Noise and Reverberation in Real-Time Fullband Speech Communication..... | 4981 |
| <i>Feifei Xiong, Minya Dong, Kechenying Zhou, Houwei Zhu, Jinwei Feng</i> | |
| Joint Neural Representation for Multiple Light Fields | 4986 |
| <i>Guillaume Le Guludec, Christine Guillemot</i> | |
| Fan-Net: Fourier-Based Adaptive Normalization for Cross-Domain Stroke Lesion Segmentation | 4991 |
| <i>Weiyi Yu, Yiming Lei, Hongming Shan</i> | |
| Structured Errors-In-Variables Modelling for Cortico-Muscular Coherence Enhancement..... | 4996 |
| <i>Zhenghao Guo, Verity M. McClelland, Wei Dai, Zoran Cvetkovic</i> | |
| Multiscale Audio Spectrogram Transformer for Efficient Audio Classification..... | 5001 |
| <i>Wentao Zhu, Mohamed Omar</i> | |
| CB-Conformer: Contextual Biasing Conformer for Biased Word Recognition | 5006 |
| <i>Yaoxun Xu, Baiji Liu, Qiaochu Huang, Xingchen Song, Zhiyong Wu, Shiyin Kang, Helen Meng</i> | |
| SENER: Sentiment Element Named Entity Recognition for Aspect-Based Sentiment Analysis | 5011 |
| <i>Sun-Kyung Lee, Jong-Hwan Kim</i> | |
| Exploiting Spatial Information with the Informed Complex-Valued Spatial Autoencoder for Target Speaker Extraction..... | 5016 |
| <i>Annika Briegleb, Mhd Modar Halimeh, Walter Kellermann</i> | |
| Enrollment Rate Prediction in Clinical Trials Based on CDF Sketching and Tensor Factorization Tools | 5021 |
| <i>Magda Amiridi, Cheng Qian, Nicholas D. Sidiropoulos, Lucas M. Glass</i> | |
| Domain Adaptation with External Off-Policy Acoustic Catalogs for Scalable Contextual End-to-End Automated Speech Recognition | 5026 |
| <i>David M. Chan, Shalini Ghosh, Ariya Rastrow, Björn Hoffmeister</i> | |
| Towards Robust Data-Driven Underwater Acoustic Localization: A Deep CNN Solution with Performance Guarantees for Model Mismatch..... | 5031 |
| <i>Amir Weiss, Andrew C. Singer, Gregory W. Wornell</i> | |
| Raising the Limit of Image Rescaling Using Auxiliary Encoding..... | 5036 |
| <i>Chenzhong Yin, Zhihong Pan, Xin Zhou, Le Kang, Paul Bogdan</i> | |

| | |
|---|------|
| LSTM-Based Video Quality Prediction Accounting for Temporal Distortions in Videoconferencing Calls..... | 5041 |
| <i>Gabriel Mittag, Babak Naderi, Vishak Gopal, Ross Cutler</i> | |
| Extended Kalman Filter for Graph Signals in Nonlinear Dynamic Systems..... | 5046 |
| <i>Guy Sagi, Nir Shlezinger, Tirza Routtenberg</i> | |
| Quantum Deep Recurrent Reinforcement Learning | 5051 |
| <i>Samuel Yen-Chi Chen</i> | |
| Prune Then Distill: Dataset Distillation with Importance Sampling | 5056 |
| <i>Anirudh S Sundar, Gokce Keskin, Chander Chandak, I-Fan Chen, Pegah Ghahremani, Shalini Ghosh</i> | |
| Pair DETR: Toward Faster Convergent DETR | 5061 |
| <i>Seyed Mehdi Iranmanesh, Sherry X. Chen, Kuo-Chin Lien</i> | |
| Enhanced Coprime Array Configuration for DoA Estimation of Non-Circular Signals..... | 5066 |
| <i>Nabil Mohsen, Ammar Hawbani, Xingfu Wang, Benjamin Bairrington, Liang Zhao, Saeed Alsamhi</i> | |
| Audio Barlow Twins: Self-Supervised Audio Representation Learning | 5071 |
| <i>Jonah Anton, Harry Coppock, Pancham Shukla, Björn W. Schuller</i> | |
| Interpretable, Unrolled Deep Radar Beampattern Design | 5076 |
| <i>Kareem Metwaly, Junho Kweon, Khaled Alhujaili, Maria Greco, Fulvio Gini, Vishal Monga</i> | |
| Adaptive Multi-Corpora Language Model Training for Speech Recognition | 5081 |
| <i>Yingyi Ma, Zhe Liu, Xuedong Zhang</i> | |
| Hardware-Limited Non-Uniform Task-Based Quantizers | 5086 |
| <i>Neil Irwin Bernardo, Jingge Zhu, Yonina C. Eldar, Jamie Evans</i> | |
| Expectation Propagation on Factor Graphs Based on Matrix Decomposition | 5091 |
| <i>Adam Mekhiche, Antonio Maria Cipriano, Charly Poulliat</i> | |
| An End-To-End Neural Network for Image-To-Audio Transformation | 5096 |
| <i>Liu Chen, Michael Deisher, Munir Georges</i> | |
| Lit the Darkness: Three-Stage Zero-Shot Learning for Low-Light Enhancement with Multi-Neighbor Enhancement Factors..... | 5101 |
| <i>Mariam Saeed, Marwan Torki</i> | |
| HuBERT-AGG: Aggregated Representation Distillation of Hidden-Unit Bert for Robust Speech Recognition | 5106 |
| <i>Wei Wang, Yanmin Qian</i> | |
| Single-Anchor UWB Localization Using Channel Impulse Response Distributions | 5111 |
| <i>Sitian Li, Alexios Balatsoukas-Stimming, Andreas Burg</i> | |
| Efficient Siamese Network for UAV Tracking | 5116 |
| <i>Xiaohan Zhang, Dong Wang, Xiaohong Ma</i> | |
| Optimal Carrier Frequency Design for Frequency Diverse Array Mimo Radar..... | 5121 |
| <i>Jie Cheng, Maria Juhlin, Wen-Qin Wang, Andreas Jakobsson</i> | |

| | |
|---|------|
| Learning Audio-Visual Dereverberation..... | 5126 |
| <i>Changan Chen, Wei Sun, David Harwath, Kristen Grauman</i> | |
| Can Spoofing Countermeasure and Speaker Verification Systems Be Jointly Optimised? | 5131 |
| <i>Wanying Ge, Hemlata Tak, Massimiliano Todisco, Nicholas Evans</i> | |
| Exploring Vision Transformer Layer Choosing for Semantic Segmentation | 5136 |
| <i>Fangjian Lin, Yizhe Ma, Shengwei Tian</i> | |
| Dataset Balancing Can Hurt Model Performance..... | 5141 |
| <i>R. Channing Moore, Daniel P. W. Ellis, Eduardo Fonseca, Shawn Hershey, Aren Jansen, Manoj Plakal</i> | |
| Mcrood: Multi-Class Radar Out-Of-Distribution Detection..... | 5146 |
| <i>Sabri Mustafa Kahya, Muhammet Sami Yavuz, Eckehard Steinbach</i> | |
| Data Augmentation Based on Invariant Shape Blending for Deep Learning Classification..... | 5151 |
| <i>Emna Ghorbel, Mahmoud Ghorbel, Slim Mhiri</i> | |
| Supercm: Revisiting Clustering for Semi-Supervised Learning..... | 5156 |
| <i>Durgesh Singh, Ahcène Boubekki, Robert Jenssen, Michael C. Kampffmeyer</i> | |
| Multi-Rate Adaptive Transform Coding for Video Compression..... | 5161 |
| <i>Lyndon R. Duong, Bohan Li, Cheng Chen, Jingning Han</i> | |
| Towards Explainable Recommendation Via Bert-Guided Explanation Generator | 5166 |
| <i>Huijing Zhan, Ling Li, Shaohua Li, Weide Liu, Manas Gupta, Alex C. Kot</i> | |
| FedAudio: A Federated Learning Benchmark for Audio Tasks..... | 5171 |
| <i>Tuo Zhang, Tiantian Feng, Samiul Alam, Sunwoo Lee, Mi Zhang, Shrikanth S. Narayanan, Salman Avestimehr</i> | |
| Dynamic Signed Graph Learning | 5176 |
| <i>Abdullah Karaaslanli, Selin Aviyente</i> | |
| Equivalence of Aperture Reduction in Element Space and Constrained Combination of DFT Beams in Beamspace..... | 5181 |
| <i>Damir Rakhimov, Martin Haardt</i> | |
| Neurally Augmented State Space Model for Simultaneous Communication and Tracking with Low Complexity Receivers | 5186 |
| <i>Fernando Pedraza, Giuseppe Caire</i> | |
| Continilm: A Continual Learning Scheme for Non-Intrusive Load Monitoring..... | 5191 |
| <i>Stavros Sykiotis, Maria Kaselimi, Anastasios Doulamis, Nikolaos Doulamis</i> | |
| Efficiently Fusing Sparse Lidar for Enhanced Self-Supervised Monocular Depth Estimation | 5196 |
| <i>Yue Wang, Mingrong Gong, Lei Xia, Qieshi Zhang, Jun Cheng</i> | |
| Learned Generative Misspecified Lower Bound | 5201 |
| <i>Hai Victor Habi, Hagit Messer, Yoram Bresler</i> | |
| Robust and Parallelizable Tensor Completion Based on Tensor Factorization and Maximum Correntropy Criterion | 5206 |
| <i>Yicong He, George K. Atia</i> | |

| | |
|--|------|
| Visual Prompting for Adversarial Robustness | 5211 |
| <i>Aochuan Chen, Peter Lorenz, Yuguang Yao, Pin-Yu Chen, Sijia Liu</i> | |
| Sparse Convolution Based Octree Feature Propagation for Lidar Point Cloud Compression | 5216 |
| <i>Muhammad Asad Lodhi, Jiahao Pang, Dong Tian</i> | |
| A Non-Contact SpO2 Estimation Using Video Magnification and Infrared Data | 5221 |
| <i>Thomas Stogiannopoulos, Grigorios-Aris Cheimariotis, Nikolaos Mitianoudis</i> | |
| Speech Separation with Large-Scale Self-Supervised Learning..... | 5226 |
| <i>Zhuo Chen, Naoyuki Kanda, Jian Wu, Yu Wu, Xiaofei Wang, Takuya Yoshioka, Jinyu Li, Sunit Sivasankaran, Sefik Emre Eskimez</i> | |
| A Variational Inequality Model for Learning Neural Networks | 5231 |
| <i>Patrick L. Combettes, Jean-Christophe Pesquet, Audrey Repetti</i> | |
| High-Dimensional Confidence Regions in Sparse MRI..... | 5236 |
| <i>Frederik Hoppe, Felix Kraher, Claudio Mayrlink Verdun, Marion I. Menzel, Holger Rauhut</i> | |
| Jointly Visual- And Semantic-Aware Graph Memory Networks for Temporal Sentence Localization in Videos..... | 5241 |
| <i>Daizong Liu, Pan Zhou</i> | |
| Temporal Contrastive Learning with Curriculum..... | 5246 |
| <i>Shuvendu Roy, Ali Etemad</i> | |
| Fine-Grained Textual Knowledge Transfer to Improve RNN Transducers for Speech Recognition and Understanding..... | 5251 |
| <i>Vishal Sunder, Samuel Thomas, Hong-Kwang J. Kuo, Brian Kingsbury, Eric Fosler-Lussier</i> | |
| Transformer-Based Bioacoustic Sound Event Detection on Few-Shot Learning Tasks | 5256 |
| <i>Liwen You, Erika Pelaez Coyotl, Suren Gunturu, Maarten Van Segbroeck</i> | |
| RIS-Aided Wideband DFRC with Reconfigurable Holographic Surface..... | 5261 |
| <i>Tong Wei, Linlong Wu, Kumar Vijay Mishra, M. R. Bhavani Shankar</i> | |
| HARQ Delay Minimization of 5G Wireless Network with Imperfect Feedback | 5266 |
| <i>Weihang Ding, Mohammad Shikh-Bahaei</i> | |
| Guided Speech Enhancement Network | 5271 |
| <i>Yang Yang, Shao-Fu Shih, Hakan Erdogan, Jamie Menjay Lin, Chehung Lee, Yunpeng Li, George Sung, Matthias Grundmann</i> | |
| JEIT: Joint End-To-End Model and Internal Language Model Training for Speech Recognition..... | 5276 |
| <i>Zhong Meng, Weiran Wang, Rohit Prabhavalkar, Tara N. Sainath, Tongzhou Chen, Ehsan Variani, Yu Zhang, Bo Li, Andrew Rosenberg, Bhuvana Ramabhadran</i> | |
| Bayesian Methods for Optical Flow Estimation Using a Variational Approximation, with Applications to Ultrasound..... | 5281 |
| <i>Jan Dorazil, Bernard H. Fleury, Franz Hlawatsch</i> | |
| Variable Attention Masking for Configurable Transformer Transducer Speech Recognition | 5286 |
| <i>Pawel Swietojanski, Stefan Braun, Dogan Can, Thiago Fraga Da Silva, Arnab Ghoshal, Takaaki Hori, Roger Hsiao, Henry Mason, Erik McDermott, Honza Silovsky, Ruchir Travadi, Xiaodan Zhuang</i> | |

| | |
|---|------|
| Enhancing Multimodal Alignment with Momentum Augmentation for Dense Video Captioning..... | 5291 |
| <i>Yiwei Wei, Shaozu Yuan, Meng Chen, Longbiao Wang</i> | |
| Robust Hypothesis Testing with Moment Constrained Uncertainty Sets | 5296 |
| <i>Akshayaa Magesh, Zhongchang Sun, Venugopal V. Veeravalli, Shaofeng Zou</i> | |
| Adaptive Step-Size Methods for Compressed SGD | 5301 |
| <i>Adarsh M. Subramaniam, Akshayaa Magesh, Venugopal V. Veeravalli</i> | |
| Study on the Fairness of Speaker Verification Systems Across Accent and Gender Groups..... | 5306 |
| <i>Mariel Estevez, Luciana Ferrer</i> | |
| Improving Electric Load Demand Forecasting with Anchor-Based Forecasting Method | 5311 |
| <i>Maria Tzelepi, Paraskevi Nousi, Anastasios Tefas</i> | |
| Resource-Efficient Transfer Learning from Speech Foundation Model Using Hierarchical Feature Fusion..... | 5316 |
| <i>Zhouyuan Huo, Khe Chai Sim, Bo Li, Dongseong Hwang, Tara N. Sainath, Trevor Strohman</i> | |
| BEANS: The Benchmark of Animal Sounds..... | 5321 |
| <i>Masato Hagiwara, Benjamin Hoffman, Jen-Yu Liu, Maddie Cusimano, Felix Effenberger, Katie Zacarian</i> | |
| Towards Accurate and Real-Time End-of-Speech Estimation..... | 5326 |
| <i>Yifeng Fan, Colin Vaz, Di He, Jahn Heymann, Viet Anh Trinh, Zhe Zhang, Venkatesh Ravichandran</i> | |
| Gaussian Process Dynamical Modeling for Adaptive Inference Over Graphs | 5331 |
| <i>Qin Lu, Konstantinos D. Polyzos</i> | |
| Pop2Piano : Pop Audio-Based Piano Cover Generation | 5336 |
| <i>Jongho Choi, Kyogu Lee</i> | |
| Locale Encoding for Scalable Multilingual Keyword Spotting Models..... | 5341 |
| <i>Pai Zhu, Hyun Jin Park, Alex Park, Angelo Scorza Scarpatti, Ignacio Lopez Moreno</i> | |
| Bayesian Optimization with Ensemble Learning Models and Adaptive Expected Improvement | 5346 |
| <i>Konstantinos D. Polyzos, Qin Lu, Georgios B. Giannakis</i> | |
| SINCO: A Novel Structural Regularizer for Image Compression Using Implicit Neural Representations | 5351 |
| <i>Harry Gao, Weijie Gan, Zhixin Sun, Ulugbek S. Kamilov</i> | |
| Physics-Informed Transfer Learning for Voltage Stability Margin Prediction | 5356 |
| <i>Manish K. Singh, Konstantinos D. Polyzos, Panagiotis A. Traganitis, Sairaj V. Dhople, Georgios B. Giannakis</i> | |
| Adapting a Self-Supervised Speech Representation for Noisy Speech Emotion Recognition by Using Contrastive Teacher-Student Learning | 5361 |
| <i>Seong-Gyun Leem, Daniel Fulford, Jukka-Pekka Onnela, David Gard, Carlos Busso</i> | |
| Evidence of Vocal Tract Articulation in Self-Supervised Learning of Speech | 5366 |
| <i>Cheol Jun Cho, Peter Wu, Abdelrahman Mohamed, Gopala K. Anumanchipalli</i> | |
| Active Subsampling Using Deep Generative Models by Maximizing Expected Information Gain..... | 5371 |
| <i>Koen C. E. Van De Camp, Hamdi Joudeh, Duarte J. Antunes, Ruud J. G. Van Sloun</i> | |

| | |
|--|------|
| A Synthetic Corpus Generation Method for Neural Vocoder Training | 5376 |
| <i>Zilin Wang, Peng Liu, Jun Chen, Sipan Li, Jinfeng Bai, Gang He, Zhiyong Wu, Helen Meng</i> | |
| Alternating Constrained Minimization Based Approximate Message Passing..... | 5381 |
| <i>Christo Kurisummoottil Thomas, Dirk Slock</i> | |
| Progressive Diversifying Policy for Multi-Agent Reinforcement Learning..... | 5386 |
| <i>Shaoqi Sun, Yuanzhao Zhai, Kele Xu, Dawei Feng, Bo Ding</i> | |
| Towards Real-Time Single-Channel Speech Separation in Noisy and Reverberant Environments | 5391 |
| <i>Julian Neri, Sebastian Braun</i> | |
| Accidental Learners: Spoken Language Identification in Multilingual Self-Supervised Models..... | 5396 |
| <i>Travis M. Bartley, Fei Jia, Krishna C. Puvvada, Samuel Kriman, Boris Ginsburg</i> | |
| Implementing Continuous HRTF Measurement in Near-Field..... | 5401 |
| <i>Ee-Leng Tan, Santi Peksi, Woon-Seng Gan</i> | |
| Efficient Protein Structural Class Prediction Via Chaos Game Representation and Recurrent Neural Networks | 5406 |
| <i>Michaela Areti Zervou, Effrosyni Doutsis, Panagiotis Tsakalides</i> | |
| Efficient Similarity-Based Passive Filter Pruning for Compressing CNNs | 5411 |
| <i>Arshdeep Singh, Mark D. Plumbley</i> | |
| Leveraging Multiple Sources in Automatic African American English Dialect Detection for Adults and Children | 5416 |
| <i>Alexander Johnson, Vishwas M. Shetty, Mari Ostendorf, Abeer Alwan</i> | |
| MTDL-NET: Morphological and Temporal Discriminative Learning for Heartbeat Classification..... | 5421 |
| <i>Can Han, Suncheng Xiang, Dahong Qian</i> | |
| Blind Source Counting and Separation with Relative Harmonic Coefficients | 5426 |
| <i>Huiyuan Sun, Prasanga Samarasinghe, Thushara Abhayapala</i> | |
| Global Matching-Optimization Network for Stereo Depth Estimation | 5431 |
| <i>Yidi Zhang, Wenqi Huang, Wenming Yang</i> | |
| U-Beat: A Multi-Scale Beat Tracking Model Based on Wave-U-Net..... | 5436 |
| <i>Tian Cheng, Masataka Goto</i> | |
| Low-Bitrate Redundancy Coding of Speech Using a Rate-Distortion-Optimized Variational Autoencoder | 5441 |
| <i>Jean-Marc Valin, Jan Bütke, Ahmed Mustafa</i> | |
| Zero-Shot Domain Adaptation of Anomalous Samples for Semi-Supervised Anomaly Detection | 5446 |
| <i>Tomoya Nishida, Takashi Endo, Yohei Kawaguchi</i> | |
| Estimation of High-Dimensional Differential Graphs from Multi-Attribute Data | 5451 |
| <i>Jitendra K. Tugnait</i> | |
| GAPter: Gray-Box Data Protector for Deep Learning Inference Services at User Side..... | 5456 |
| <i>Hao Wu, Bo Yang, Xiaopeng Ke, Siyi He, Fengyuan Xu, Sheng Zhong</i> | |
| Text-To-Speech Synthesis Based on Latent Variable Conversion Using Diffusion Probabilistic Model and Variational Autoencoder | 5461 |
| <i>Yusuke Yasuda, Tomoki Toda</i> | |

| | |
|--|------|
| RD-NAS: Enhancing One-Shot Supernet Ranking Ability Via Ranking Distillation from Zero-Cost Proxies..... | 5466 |
| <i>Peijie Dong, Xin Niu, Lujun Li, Zhiliang Tian, Xiaodong Wang, Zimian Wei, Hengyue Pan, Dongsheng Li</i> | |
| ADHD Classification with Biomarker Identification Using a Triplet Loss Attention Auto-Encoding Network..... | 5471 |
| <i>Yibin Tang, Ying Chen, Yuan Gao, Aimin Jiang, Lin Zhou</i> | |
| Phase Retrieval for Rydberg Quantum Arrays..... | 5476 |
| <i>Peter Vouras, Kumar Vijay Mishra, Alexandra Artusio-Glimpse</i> | |
| Tempo Vs. Pitch: Understanding Self-Supervised Tempo Estimation..... | 5481 |
| <i>Giovana Morais, Matthew E. P. Davies, Marcelo Queiroz, Magdalena Fuentes</i> | |
| Perceptual Quality Assessment for Digital Human Heads | 5486 |
| <i>Zicheng Zhang, Yingjie Zhou, Wei Sun, Xiongkuo Min, Yuzhe Wu, Guangtao Zhai</i> | |
| Leveraging Language Embeddings for Cross-Lingual Self-Supervised Speech Representation Learning | 5491 |
| <i>Tomohiro Tanaka, Ryo Masumura, Mana Ihori, Hiroshi Sato, Taiga Yamane, Takanori Ashihara, Kohei Matsuura, Takafumi Moriya</i> | |
| Progressive Meta-Pooling Learning for Lightweight Image Classification Model..... | 5496 |
| <i>Peijie Dong, Xin Niu, Zhiliang Tian, Lujun Li, Xiaodong Wang, Zimian Wei, Hengyue Pan, Dongsheng Li</i> | |
| Cross-Training: A Semi-Supervised Training Scheme for Speech Recognition..... | 5501 |
| <i>Soheil Khorram, Anshuman Tripathi, Jaeyoung Kim, Han Lu, Qian Zhang, Rohit Prabhavalkar, Hasim Sak</i> | |
| Multi-View Millimeter-Wave Imaging Over Wireless Cellular Network..... | 5506 |
| <i>Xin Tong, Zhaoyang Zhang, Zhaohui Yang</i> | |
| A Mutual Implicit Sentiment Analysis Model with Bundle-Aware Contrastive Learning | 5511 |
| <i>Siqi Cai, Jingling Yuan, Lin Li</i> | |
| T5lephone: Bridging Speech and Text Self-Supervised Models for Spoken Language Understanding Via Phoneme Level T5..... | 5516 |
| <i>Chan-Jan Hsu, Ho-Lam Chung, Hung-Yi Lee, Yu Tsao</i> | |
| Speech Summarization of Long Spoken Document: Improving Memory Efficiency of Speech/Text Encoders | 5521 |
| <i>Takatomo Kano, Atsunori Ogawa, Marc Delcroix, Roshan Sharma, Kohei Matsuura, Shinji Watanabe</i> | |
| Spatial-Domain Object Detection Under Mimo-Fmcw Automotive Radar Interference | 5526 |
| <i>Sian Jin, Pu Wang, Petros Boufounos, Ryuhei Takahashi, Sumit Roy</i> | |
| SADI: A Self-Adaptive Decomposed Interpretable Framework for Electric Load Forecasting Under Extreme Events..... | 5531 |
| <i>Hengbo Liu, Ziqing Ma, Linxiao Yang, Tian Zhou, Rui Xia, Yi Wang, Qingsong Wen, Liang Sun</i> | |
| UNeXt: A Low-Dose CT Denoising UNet Model with the Modified ConvNeXt Block..... | 5536 |
| <i>Farzan Niknejad Mazandarani, Paul Babyn, Javad Alirezaie</i> | |

| | |
|--|------|
| Flow-Guided Deformable Alignment Network with Self-Supervision for Video Inpainting | 5541 |
| <i>Zhiliang Wu, Kang Zhang, Changchang Sun, Hanyu Xuan, Yan Yan</i> | |
| A Novel State Connection Strategy for Quantum Computing to Represent and Compress Digital Images | 5546 |
| <i>Md Ershadul Haque, Manoranjan Paul, Anwar Ulhaq, Tanmoy Debnath</i> | |
| Light-Weight Sequential SBL Algorithm: An Alternative to OMP | 5551 |
| <i>Rohan R. Pote, Bhaskar D. Rao</i> | |
| Robust Video Object Segmentation with Restricted Attention..... | 5556 |
| <i>Huaizheng Zhang, Pinxue Guo, Zhongwen Le, Wenqiang Zhang</i> | |
| MmWave Wi-Fi Trajectory Estimation with Continuous-Time Neural Dynamic Learning | 5561 |
| <i>Cristian J. Vaca-Rubio, Pu Wang, Toshiaki Koike-Akino, Ye Wang, Petros Boufounos, Petar Popovski</i> | |
| Multi-Stream Facial Adaptive Network for Expression Recognition from a Single Image | 5566 |
| <i>Baichuan Zhang, Fanyang Meng, Runwei Ding, Mengyuan Liu</i> | |
| On the Role of Visual Context in Enriching Music Representations..... | 5571 |
| <i>Kleanthis Avramidis, Shanti Stewart, Shrikanth Narayanan</i> | |
| TINYCOD: Tiny and Effective Model for Camouflaged Object Detection | 5576 |
| <i>Haozhe Xing, Shuyong Gao, Hao Tang, Tsui Qin Mok, Yanlan Kang, Wenqiang Zhang</i> | |
| Soft Dynamic Time Warping for Multi-Pitch Estimation and Beyond..... | 5581 |
| <i>Michael Krause, Christof Weiß, Meinard Müller</i> | |
| Classification of the Cervical Vertebrae Maturation (CVM) Stages Using the Tripod Network..... | 5586 |
| <i>Salih Atici, Hongyi Pan, Mohammed H. Elnagar, Veerasathpurush Allareddy, Omar Suhaym, Rashid Ansari, Ahmet Enis Cetin</i> | |
| Multi-Agent Reinforcement Learning for Covert Semantic Communications Over Wireless Networks | 5591 |
| <i>Yining Wang, Ye Hu, Hongyang Du, Tao Luo, Dusit Niyato</i> | |
| Framework Wavegan: High Speed Adversarial Vocoder in Time Domain with Very Low Computational Complexity | 5596 |
| <i>Ahmed Mustafa, Jean-Marc Valin, Jan Bütthe, Paris Smaragdis, Mike Goodwin</i> | |
| Promptts: Controllable Text-To-Speech with Text Descriptions | 5601 |
| <i>Zhifang Guo, Yichong Leng, Yihan Wu, Sheng Zhao, Xu Tan</i> | |
| Intermediate Fine-Tuning Using Imperfect Synthetic Speech for Improving Electrolaryngeal Speech Recognition | 5606 |
| <i>Lester Phillip Violeta, Ding Ma, Wen-Chin Huang, Tomoki Toda</i> | |
| Semantic Preprocessor for Image Compression for Machines..... | 5611 |
| <i>Mingyi Yang, Luis Herranz, Fei Yang, Luka Murn, Marc Gorriz Blanch, Shuai Wan, Fuzheng Yang, Marta Mrak</i> | |
| Privacy Preserving Face Recognition with Lensless Camera..... | 5616 |
| <i>Chris Henry, M. Salman Asif, Zhu Li</i> | |
| High Quality Audio Coding with Mdcnet | 5621 |
| <i>Grant Davidson, Mark Vinton, Per Ekstrand, Cong Zhou, Lars Villemoes, Lie Lu</i> | |

| | |
|--|------|
| Search for Efficient Deep Visual-Inertial Odometry Through Neural Architecture Search..... | 5626 |
| <i>Yu Chen, Mingyu Yang, Hun-Seok Kim</i> | |
| An Antispoofing Approach in Biometric Authentication System for a Smartcard | 5631 |
| <i>Han-Sol Lee, Moon-Kyu Song, Junseo Lee, Yeolmin Seong, Ducksoo Kim, Kwanghyuk Bae, Seongwook Song</i> | |
| Decoupled Visual Causality for Robust Detection | 5636 |
| <i>Ping Jiang, Xiaoheng Deng, Shichao Zhang</i> | |
| Factorized Blank Thresholding for Improved Runtime Efficiency of Neural Transducers | 5641 |
| <i>Duc Le, Frank Seide, Yuhao Wang, Yang Li, Kjell Schubert, Ozlem Kalinli, Michael L. Seltzer</i> | |
| Nonparallel Emotional Voice Conversion for Unseen Speaker-Emotion Pairs Using Dual Domain Adversarial Network & Virtual Domain Pairing | 5646 |
| <i>Nirmesh Shah, Mayank Singh, Naoya Takahashi, Naoyuki Onoe</i> | |
| Stacking-Based Attention Temporal Convolutional Network for Action Segmentation | 5651 |
| <i>Liu Yang, Yu Jiang, Junkun Hong, Zhenjie Wu, Zhan Yang, Jun Long</i> | |
| Lightweight Fisher Vector Transfer Learning for Video Deduplication | 5656 |
| <i>Chris Henry, Rijun Liao, Ruiyuan Lin, Zhebin Zhang, Hongyu Sun, Zhu Li</i> | |
| Iterative Water-Filling Power and Subcarrier Allocation for Multicarrier NOMA Downlink | 5661 |
| <i>Chin Choy Chai, Xiao-Ping Zhang</i> | |
| Self-Adaptive Reasoning on Sub-Questions for Multi-Hop Question Answering | 5666 |
| <i>Zekai Li, Wei Peng</i> | |
| Optimal Compression for Minimizing Classification Error Probability: An Information-Theoretic Approach | 5671 |
| <i>Jingchao Gao, Ao Tang, Weiyu Xu</i> | |
| Quickest Change Detection with Leave-One-Out Density Estimation..... | 5676 |
| <i>Yuchen Liang, Venugopal V. Veeravalli</i> | |
| Self-Supervised Learning-Based Source Separation for Meeting Data..... | 5681 |
| <i>Yuang Li, Xianrui Zheng, Philip C. Woodland</i> | |
| The R3VIVAL Dataset: Repository of Room Responses and 360 Videos of a Variable Acoustics Lab..... | 5686 |
| <i>Florian Klein, Sebastià V. Amengual Garí</i> | |
| Unrestricted Anchor Graph Based GCN for Incomplete Multi-View Clustering | 5691 |
| <i>Liang Zhao, Zihao Wang, Yukun Yuan, Feng Ding</i> | |
| DecomFormer: Decompose Self-Attention Via Fourier Transform for VHR Aerial Image Scene Classification | 5696 |
| <i>Yan Zhang, Xiyuan Gao, Xiao Pu, Tao Wang, Xinbo Gao</i> | |
| Nonparallel High-Quality Audio Super Resolution with Domain Adaptation and Resampling CycleGANs | 5701 |
| <i>Reo Yoneyama, Ryuichi Yamamoto, Kentaro Tachibana</i> | |
| COVID-19 Detection from Speech in Noisy Conditions | 5706 |
| <i>Shuo Liu, Adria Mallol-Ragolta, Björn W. Schuller</i> | |

| | |
|--|------|
| Sparse Black-Box Inversion Attack with Limited Information | 5711 |
| <i>Yixiao Xu, Xiaolei Liu, Teng Hu, Bangzhou Xin, Run Yang</i> | |
| Small-Footprint Slimmable Networks for Keyword Spotting | 5716 |
| <i>Zuhaib Akhtar, Mohammad Omar Khursheed, Dongsu Du, Yuzong Liu</i> | |
| BAUENet: Boundary-Aware Uncertainty Enhanced Network for Infrared Small Target Detection..... | 5721 |
| <i>Tianxiang Chen, Qi Chu, Zhentao Tan, Bin Liu, Nenghai Yu</i> | |
| A Novel Extrapolation Technique to Accelerate WMMSE | 5726 |
| <i>Kaiwen Zhou, Zhilin Chen, Guochen Liu, Zhitang Chen</i> | |
| Bagging R-CNN: Ensemble for Object Detection in Complex Traffic Scenes | 5731 |
| <i>Pengteng Li, Ying He, Dongfu Yin, F Richard Yu, Pinhao Song</i> | |
| Possibilistic Bernoulli Filter for Extended Target Tracking | 5736 |
| <i>Zhijin Chen, Branko Ristic, Du Yong Kim</i> | |
| Does Human Speech Follow Benford's Law? | 5741 |
| <i>Leo Hsu, Visar Berisha</i> | |
| Image Adversarial Steganography Based on Joint Distortion | 5746 |
| <i>Zexin Fan, Kejiang Chen, Chuan Qin, Kai Zeng, Weiming Zhang, Nenghai Yu</i> | |
| End-To-End Neural Audio Coding in the MDCT Domain | 5751 |
| <i>Hyungseob Lim, Jihyun Lee, Byeong Hyeon Kim, Inseon Jang, Hong-Goo Kang</i> | |
| Multimodal Knowledge Distillation for Arbitrary-Oriented Object Detection in Aerial Images | 5756 |
| <i>Zhanchao Huang, Wei Li, Ran Tao</i> | |
| Simultaneous Reconstruction and Uncertainty Quantification for Tomography | 5761 |
| <i>Agnimitra Dasgupta, Carlo Graziani, Zichao Wendy Di</i> | |
| Joint Waveform and Passive Beamformer Design in Multi-IRS-Aided Radar..... | 5766 |
| <i>Zahra Esmailbeig, Arian Eamaz, Kumar Vijay Mishra, Mojtaba Soltanalian</i> | |
| SingNet: A Real-Time Singing Voice Beat and Downbeat Tracking System | 5771 |
| <i>Mojtaba Heydari, Ju-Chiang Wang, Zhiyao Duan</i> | |
| RCDPT: Radar-Camera Fusion Dense Prediction Transformer | 5776 |
| <i>Chen-Chou Lo, Patrick Vandewalle</i> | |
| Local to Global Prior Learning for Blind Unsupervised Image Super Resolution | 5781 |
| <i>Kazuhiro Yamawaki, Xian-Hua Han</i> | |
| Reducing the Computational Complexity of Learning with Random Convolutional Features | 5786 |
| <i>M. A. Omid, B. Seyfe, S. Valaee</i> | |
| On Unsupervised Uncertainty-Driven Speech Pseudo-Label Filtering and Model Calibration | 5791 |
| <i>Nauman Dawalatabad, Sameer Khurana, Antoine Laurent, James Glass</i> | |
| Scalable Weight Reparametrization for Efficient Transfer Learning..... | 5796 |
| <i>Byeongeun Kim, Jun-Tae Lee, Seunghan Yang, Simyung Chang</i> | |
| Eigen-Decomposition-Free Directed Graph Sampling Via Gershgorin Disc Alignment..... | 5801 |
| <i>Yuejiang Li, H. Vicky Zhao, Gene Cheung</i> | |

| | |
|---|------|
| Stay in the Middle: A Semi-Supervised Model for CT Metal Artifact Reduction | 5806 |
| <i>Tao Wang, Hui Yu, Zexin Lu, Zhongzhou Zhang, Jiliu Zhou, Yi Zhang</i> | |
| A Bayesian Perspective for Determinant Minimization Based Robust Structured Matrix Factorization | 5811 |
| <i>Gokcan Tatli, Alper T. Erdogan</i> | |
| Unsupervised Domain Adaptation Via Subspace Interpolating Deep Dictionary Learning: A Case Study in Machine Inspection | 5816 |
| <i>Kriti Kumar, Angshul Majumdar, A Anil Kumar, M Girish Chandra</i> | |
| Dual-Attention Neural Transducers for Efficient Wake Word Spotting in Speech Recognition..... | 5821 |
| <i>Saumya Y. Sahai, Jing Liu, Thejaswi Muniyappa, Kanthashree M. Sathyendra, Anastasios Alexandridis, Grant P. Strimel, Ross McGowan, Ariya Rastrow, Feng-Ju Chang, Athanasios Mouchtaris, Siegfried Kunzmann</i> | |
| LiNuQA: Lightweight No-Reference Image Quality Assessment Based on Non-Uniform Weighting | 5826 |
| <i>Wook-Hyung Kim, Cheul-Hee Hahm, Anant Bajjal, Namuk Kim, Ilhyun Cho, Jayoon Koo</i> | |
| MFAT: A Multi-Level Feature Aggregated Transformer for Person Re-Identification..... | 5831 |
| <i>Bowen Tan, Linfeng Xu, Zihuan Qiu, Qingbo Wu, Fanman Meng</i> | |
| Single-Particle Tracking by Graph Transformer..... | 5836 |
| <i>Satoshi Kamiya, Kazuhiro Hotta, Taka-Aki Tsunoyama, Akihiro Kusumi</i> | |
| Large-Scale Contrastive Language-Audio Pretraining with Feature Fusion and Keyword-to-Caption Augmentation | 5841 |
| <i>Yusong Wu, Ke Chen, Tianyu Zhang, Yuchen Hui, Taylor Berg-Kirkpatrick, Shlomo Dubnov</i> | |
| Active Noise Control Over 3D Space: A Realistic Error Microphone Geometry Design | 5846 |
| <i>Huiyuan Sun, Prasanga Samarasinghe, Thushara Abhayapala</i> | |
| MTFD: Multi-Teacher Fusion Distillation for Compressed Video Action Recognition..... | 5851 |
| <i>Jinxin Guo, Jiaqiang Zhang, Shaojie Li, Xiaojing Zhang, Ming Ma</i> | |
| Comparison of Soft and Hard Target RNN-T Distillation for Large-Scale ASR..... | 5856 |
| <i>Dongseong Hwang, Khe Chai Sim, Yu Zhang, Trevor Strohman</i> | |
| Mask Guided Selective Context Decoding for Handwritten Chinese Text Recognition | 5861 |
| <i>Tao Li, Shilian Wu, Zengfu Wang</i> | |
| A Multi-Stage Hierarchical Relational Graph Neural Network for Multimodal Sentiment Analysis..... | 5866 |
| <i>Peizhu Gong, Jin Liu, Xiliang Zhang, Xingye Li</i> | |
| Towards Practical Edge Inference Attacks Against Graph Neural Networks | 5871 |
| <i>Kailai Li, Jiawei Sun, Ruoxin Chen, Wei Ding, Kexue Yu, Jie Li, Chentao Wu</i> | |
| Halluaudio: Hallucinate Frequency as Concepts for Few-Shot Audio Classification | 5876 |
| <i>Zhongjie Yu, Shuyang Wang, Lin Chen, Zhongwei Cheng</i> | |
| AVES: Animal Vocalization Encoder Based on Self-Supervision | 5881 |
| <i>Masato Hagiwara</i> | |
| Boosting Signal Modulation Few-Shot Learning with Pre-Transformation..... | 5886 |
| <i>Peng Sun, Jie Su, Zhenyu Wen, Yejian Zhou, Zhen Hong, Shanqing Yu, Huaji Zhou</i> | |

| | |
|--|------|
| Leveraging Heteroscedastic Uncertainty in Learning Complex Spectral Mapping for Single-Channel Speech Enhancement..... | 5891 |
| <i>Kuan-Lin Chen, Daniel D. E. Wong, Ke Tan, Buye Xu, Anurag Kumar, Vamsi Krishna Ithapu</i> | |
| Metric Learning for User-Defined Keyword Spotting..... | 5896 |
| <i>Jaemin Jung, Youkyum Kim, Jihwan Park, Youshin Lim, Byeong-Yeol Kim, Youngjoon Jang, Joon Son Chung</i> | |
| Multi-View Graph Regularized Deep Autoencoder-Like NMF Framework | 5901 |
| <i>Liang Zhao, Zihao Wang, Ziyue Wang, Zhikui Chen</i> | |
| A Spatial-Temporal ECG Emotion Recognition Model Based on Dynamic Feature Fusion | 5906 |
| <i>Shuo Xiao, Xiaojing Qiu, Chaogang Tang, Zhenzhen Huang</i> | |
| A Flow-Guided Non-Local Alignment Network for Video Compressive Sensing Reconstruction..... | 5911 |
| <i>Chao Zhou, Can Chen, Dengyin Zhang</i> | |
| Compressed-Sensing-Based 3D Localization with Distributed Passive Reconfigurable Intelligent Surfaces | 5916 |
| <i>Jiguang He, Aymen Fakhreddine, Henk Wymeersch, George C. Alexandropoulos</i> | |
| Hierarchical Pronunciation Assessment with Multi-Aspect Attention | 5921 |
| <i>Heejin Do, Yunsu Kim, Gary Geunbae Lee</i> | |
| Generic Dependency Modeling for Multi-Party Conversation..... | 5926 |
| <i>Weizhou Shen, Xiaojun Quan, Ke Yang</i> | |
| Certified Robustness of Quantum Classifiers Against Adversarial Examples Through Quantum Noise..... | 5931 |
| <i>Jih-Cing Huang, Yu-Lin Tsai, Chao-Han Huck Yang, Cheng-Fang Su, Chia-Mu Yu, Pin-Yu Chen, Sy-Yen Kuo</i> | |
| Statistical Analysis of Speech Disorder Specific Features to Characterise Dysarthria Severity Level..... | 5936 |
| <i>Amlu Anna Joshy, P. N. Parameswaran, Siddharth R. Nair, Rajeev Rajan</i> | |
| Binary Sequence Set Optimization for CDMA Applications Via Mixed-Integer Quadratic Programming | 5941 |
| <i>Alan Yang, Tara Mina, Grace Gao</i> | |
| Zero-Shot Anomalous Sound Detection in Domestic Environments Using Large-Scale Pretrained Audio Pattern Recognition Models | 5946 |
| <i>Alessandro Ilic Mezza, Giulio Zanetti, Maximo Cobos, Fabio Antonacci</i> | |
| Compressive Sensing with Tensorized Autoencoder | 5951 |
| <i>Rakib Hyder, M. Salman Asif</i> | |
| SRTNET: Time Domain Speech Enhancement Via Stochastic Refinement..... | 5956 |
| <i>Zhibin Qiu, Mengfan Fu, Yinfeng Yu, Lili Yin, Fuchun Sun, Hao Huang</i> | |
| Symbol Level Precoding in the RF Domain for Low Hardware Complexity RIS-Assisted MU-MISO Systems..... | 5961 |
| <i>Christos G. Tsinos, Theodoros A. Tsiftsis, Robert Schober</i> | |
| Hint-Dynamic Knowledge Distillation..... | 5966 |
| <i>Yiyang Liu, Chenxin Li, Xiaotong Tu, Xinghao Ding, Yue Huang</i> | |

| | |
|--|------|
| A Benchmark for Evaluating Robustness of Spoken Language Understanding Models in Slot Filling | 5971 |
| <i>Meizheng Peng, Xu Jia, Min Peng</i> | |
| Maskdul: Data Uncertainty Learning in Masked Face Recognition..... | 5976 |
| <i>Libo Zhang, Weiming Xiong, Ku Zhao, Kehan Chen, Mingyang Zhong</i> | |
| Topological Signal Processing Over Weighted Simplicial Complexes | 5981 |
| <i>Claudio Battiloro, Stefania Sardellitti, Sergio Barbarossa, Paolo Di Lorenzo</i> | |
| Selecting Language Models Features VIA Software-Hardware Co-Design..... | 5986 |
| <i>Vlad Pandelea, Edoardo Ragusa, Paolo Gastaldo, Erik Cambria</i> | |
| A Perceptual Neural Audio Coder with a Mean-Scale Hyperprior..... | 5991 |
| <i>Joon Byun, Seungmin Shin, Youngcheol Park, Jongmo Sung, Seungkwon Beack</i> | |
| Single Domain Dynamic Generalization for Iris Presentation Attack Detection..... | 5996 |
| <i>Yachun Li, Jingjing Wang, Yuhui Chen, Di Xie, Shiliang Pu</i> | |
| Token2vec: A Joint Self-Supervised Pre-Training Framework Using Unpaired Speech and Text | 6001 |
| <i>Xianghu Yue, Junyi Ao, Xiaoxue Gao, Haizhou Li</i> | |
| Convergence Analysis of Graphical Game-Based Nash Q-Learning Using the Interaction Detection Signal of N-Step Return | 6006 |
| <i>Yunkai Zhuang, Shangdong Yang, Wenbin Li, Yang Gao</i> | |
| Convex Optimization of Deep Polynomial and ReLU Activation Neural Networks | 6011 |
| <i>Burak Bartan, Mert Pilanci</i> | |
| Joint Generative-Contrastive Representation Learning for Anomalous Sound Detection..... | 6016 |
| <i>Xiao-Min Zeng, Yan Song, Zhu Zhuo, Yu Zhou, Yu-Hong Li, Hui Xue, Li-Rong Dai, Ian McLoughlin</i> | |
| Tg-Critic: A Timbre-Guided Model for Reference-Independent Singing Evaluation..... | 6021 |
| <i>Xiaoheng Sun, Yuejie Gao, Hanyao Lin, Huaping Liu</i> | |
| Self-Supervised Learning of Audio Representations Using Angular Contrastive Loss..... | 6026 |
| <i>Shanshan Wang, Soumya Tripathy, Annamaria Mesaros</i> | |
| Learning Expressive and Generalizable Motion Features for Face Forgery Detection | 6031 |
| <i>Jingyi Zhang, Peng Zhang, Jingjing Wang, Di Xie, Shiliang Pu</i> | |
| NSV-TTS: Non-Speech Vocalization Modeling and Transfer in Emotional Text-To-Speech | 6036 |
| <i>Haitong Zhang, Xinyuan Yu, Yue Lin</i> | |
| On Adversarial Robustness of Audio Classifiers..... | 6041 |
| <i>Kangkang Lu, Manh Cuong Nguyen, Xun Xu, Chuan Sheng Foo</i> | |
| Mitigating Domain Dependency for Improved Speech Enhancement Via SNR Loss Boosting | 6046 |
| <i>Lili Yin, Di Wu, Zhibin Qiu, Hao Huang</i> | |
| An End-To-End Framework for Partial View-Aligned Clustering with Graph Structure..... | 6051 |
| <i>Liang Zhao, Qiongjie Xie, Sontao Wu, Shubin Ma</i> | |
| A Topic-Enhanced Approach for Emotion Distribution Forecasting in Conversations | 6056 |
| <i>Xin Lu, Weixiang Zhao, Yanyan Zhao, Bing Qin, Zhentao Zhang, Junjie Wen</i> | |

| | |
|--|------|
| Learning 3D Human Pose and Shape Estimation Using Uncertainty-Aware Body Part Segmentation | 6061 |
| <i>Ziming Wang, Han Yu, Xiaoguang Zhu, Zengwen Li, Changxue Chen, Liang Song</i> | |
| Vitasd: Robust Vision Transformer Baselines for Autism Spectrum Disorder Facial Diagnosis..... | 6066 |
| <i>Xu Cao, Wenqian Ye, Elena Sizikova, Xue Bai, Megan Coffee, Hongwu Zeng, Jianguo Cao</i> | |
| Real-Time Target Sound Extraction | 6071 |
| <i>Bandhav Veluri, Justin Chan, Malek Itani, Tuochao Chen, Takuya Yoshioka, Shyamnath Gollakota</i> | |
| Underwater Image Restoration with Light-Aware Progressive Network | 6076 |
| <i>Jian Yang, Chen Li, Xuelong Li</i> | |
| Modaldrop: Modality-Aware Regularization for Temporal-Spectral Fusion in Human Activity Recognition | 6081 |
| <i>Xin Zeng, Yiqiang Chen, Benfeng Xu, Tengxiang Zhang</i> | |
| A Simulation-Based Framework for Urban Traffic Accident Detection..... | 6086 |
| <i>Haohan Luo, Feng Wang</i> | |
| Context-Aware Fine-Tuning of Self-Supervised Speech Models | 6091 |
| <i>Suwon Shon, Felix Wu, Kwangyoun Kim, Prashant Sridhar, Karen Livescu, Shinji Watanabe</i> | |
| DAIS: The Delft Database of EEG Recordings of Dutch Articulated and Imagined Speech..... | 6096 |
| <i>Bo Dekker, Alfred C. Schouten, Odette Scharenborg</i> | |
| Diversifying Message Aggregation in Multi-Agent Communication Via Normalized Tensor Nuclear Norm Regularization | 6101 |
| <i>Yuanzhao Zhai, Kele Xu, Bo Ding, Dawei Feng, Zijian Gao, Huaimin Wang</i> | |
| Learning to Auto-Correct for High-Quality Spectrograms..... | 6106 |
| <i>Zhiyang Zhou, Shihui Liu</i> | |
| Saliency-Driven Hierarchical Learned Image Coding for Machines..... | 6111 |
| <i>Kristian Fischer, Fabian Brand, Christian Blum, André Kaup</i> | |
| Time-Varying Signals Recovery Via Graph Neural Networks | 6116 |
| <i>Jhon A. Castro-Correa, Jhony H. Giraldo, Anindya Mondal, Mohsen Badiely, Thierry Bouwmans, Fragkiskos D. Malliaros</i> | |
| MHLAT: Multi-Hop Label-Wise Attention Model for Automatic ICD Coding | 6121 |
| <i>Junwen Duan, Han Jiang, Ying Yu</i> | |
| Improving Phase-Vocoder-Based Time Stretching by Time-Directional Spectrogram Squeezing..... | 6126 |
| <i>Natsuki Akaishi, Kohei Yatabe, Yasuhiro Oikawa</i> | |
| Dual-Uncertainty Guided Curriculum Learning and Part-Aware Feature Refinement for Domain Adaptive Person Re-Identification | 6131 |
| <i>Zhangping Liu, Bin Liu, Zhiwei Zhao, Qi Chu, Nenghai Yu</i> | |
| Weight Averaging: A Simple Yet Effective Method to Overcome Catastrophic Forgetting in Automatic Speech Recognition | 6136 |
| <i>Steven Vander Eeckt, Hugo Van Hamme</i> | |
| Change Point Detection with Neural Online Density-Ratio Estimator..... | 6141 |
| <i>Xiuheng Wang, Ricardo Augusto Borsoi, Cédric Richard, Jie Chen</i> | |

| | |
|--|------|
| Finding Optimal Numerical Format for Sub-8-Bit Post-Training Quantization of Vision Transformers..... | 6146 |
| <i>Janghwan Lee, Youngdeok Hwang, Jungwook Choi</i> | |
| Audio-Visual Inpainting: Reconstructing Missing Visual Information with Sound..... | 6151 |
| <i>Valentina Sanguineti, Sanket Thakur, Pietro Morerio, Alessio Del Bue, Vittorio Murino</i> | |
| Achievable Error Exponents for Almost Fixed-Length M-Ary Hypothesis Testing..... | 6156 |
| <i>Jun Diao, Lin Zhou, Lin Bai</i> | |
| A Memory-Free Evolving Bipolar Neural Network for Efficient Multi-Label Stream Learning..... | 6161 |
| <i>Sourav Mishra, Suresh Sundaram</i> | |
| Transwnet: Integrating Transformers into CNNs Via Row and Column Attention for Abdominal Multi-Organ Segmentation..... | 6166 |
| <i>Yazhen Xie, Yanglin Huang, Yuan Zhang, Xuanya Li, Xiongjun Ye, Kai Hu</i> | |
| UAV Local Path Planning Based on Improved Proximal Policy Optimization Algorithm..... | 6171 |
| <i>Jiahao Xu, Xufeng Yan, Cui Peng, Xinquan Wu, Lipeng Gu, Yanbiao Niu</i> | |
| Tracking Objects and Activities with Attention for Temporal Sentence Grounding..... | 6176 |
| <i>Zeyu Xiong, Daizong Liu, Pan Zhou, Jiahao Zhu</i> | |
| Does Your Model Think Like an Engineer? Explainable AI for Bearing Fault Detection with Deep Learning | 6181 |
| <i>Thomas Decker, Michael Lebacher, Volker Tresp</i> | |
| Incorporating Reliability in Graph Information Propagation by Fluid Dynamics Diffusion: A Case of Multimodal Semisupervised Deep Learning..... | 6186 |
| <i>Andrea Marinoni, Marine Mercier, Qian Shi, Sivasakthy Selvakumaran, Mark Girolami</i> | |
| Fast Yet Effective Speech Emotion Recognition with Self-Distillation | 6191 |
| <i>Zhao Ren, Thanh Tam Nguyen, Yi Chang, Björn W. Schuller</i> | |
| Heart Rate Estimation and Performance Analysis Using MIMO Radar with Dispersed Antennas..... | 6196 |
| <i>Peichao Wang, Qian He</i> | |
| Multi-Layer Seasonal Perception Network for Time Series Forecasting..... | 6201 |
| <i>Ruoshu Wang, Shengfa Miao, Di Liu, Xin Jin, Weisheng Zhang</i> | |
| Measure and Countermeasure of the Capsulation Attack Against Backdoor-Based Deep Neural Network Watermarks | 6206 |
| <i>Fang-Qi Li, Shi-Lin Wang, Yun Zhu</i> | |
| A Context-Aware Computational Approach for Measuring Vocal Entrainment in Dyadic Conversations..... | 6211 |
| <i>Rimta Lahiri, Md Nasir, Catherine Lord, So Hyun Kim, Shrikanth Narayanan</i> | |
| Federated Intelligent Terminals Facilitate Stuttering Monitoring..... | 6216 |
| <i>Yongzi Yu, Wanyong Qiu, Chen Quan, Kun Qian, Zhihua Wang, Yu Ma, Bin Hu, Björn W. Schuller, Yoshiharu Yamamoto</i> | |
| Automatic Segmentation of Nasopharyngeal Carcinoma in CT Images Using Dual Attention and Edge Detection | 6221 |
| <i>Qizhi Wang, Wei Huang, Yuan Zhang, Xuanya Li, Xiongjun Ye, Kai Hu</i> | |

| | |
|---|------|
| A Study of Audio Mixing Methods for Piano Transcription in Violin-Piano Ensembles | 6226 |
| <i>Hyemi Kim, Jiyun Park, Taegyun Kwon, Dasaem Jeong, Juhan Nam</i> | |
| Intent Does Matter! Propagating High-Order Relations for Exploring Interest Preferences | 6231 |
| <i>Xiangping Zheng, Xun Liang, Bo Wu, Junlan Feng, Yuhui Guo, Sensen Zhang</i> | |
| Learning Scene Flow from 3D Point Clouds with Cross-Transformer and Global Motion Cues..... | 6236 |
| <i>Mingliang Zhai, Kang Ni, Jiucheng Xie, Hao Gao</i> | |
| Bit Error and Block Error Rate Training for ML-Assisted Communication | 6241 |
| <i>Reinhard Wiesmayr, Gian Marti, Chris Dick, Haochuan Song, Christoph Studer</i> | |
| Disentangled and Robust Representation Learning for Bragging Classification in Social Media | 6246 |
| <i>Xiang Li, Yucheng Zhou</i> | |
| Whether Contribution of Features Differ Between Video-Mediated and In-Person Meetings in Important Utterance Estimation..... | 6251 |
| <i>Fumio Nihei, Ryo Ishii, Yukiko I. Nakano, Atsushi Fukayama, Takao Nakamura</i> | |
| Spike-Based Optical Flow Estimation Via Contrastive Learning..... | 6256 |
| <i>Mingliang Zhai, Kang Ni, Jiucheng Xie, Hao Gao</i> | |
| LABANet: Lead-Assisting Backbone Attention Network for Oral Multi-Pathology Segmentation..... | 6261 |
| <i>Huabao Chen, Xiaolong Huang, Qiankun Li, Jianqing Wang, Bo Fang, Junxin Chen</i> | |
| Hybrid Neural Network with Cross- And Self-Module Attention Pooling for Text-Independent Speaker Verification | 6266 |
| <i>Jahangir Alam, Woo Hyun Kang, Abderrahim Fathan</i> | |
| Feature-Rich Audio Model Inversion for Data-Free Knowledge Distillation Towards General Sound Classification | 6271 |
| <i>Zuheng Kang, Yayun He, Jianzong Wang, Junqing Peng, Xiaoyang Qu, Jing Xiao</i> | |
| Cross-Modal Optical Flow Estimation Via Modality Compensation and Alignment..... | 6276 |
| <i>Mingliang Zhai, Kang Ni, Jiucheng Xie, Hao Gao</i> | |
| Joint Data Association, NLOS Mitigation, and Clutter Suppression for Networked Device-Free Sensing in 6G Cellular Network..... | 6281 |
| <i>Qin Shi, Liang Liu, Shuowen Zhang</i> | |
| Product Graph Learning from Multi-Attribute Graph Signals with Inter-Layer Coupling..... | 6286 |
| <i>Chen-Yue Zhang, Yiran He, Hoi-To Wai</i> | |
| CommDre: Document-Level Relation Extraction with Self-Supervised Commonsense Learning | 6291 |
| <i>Rongzhen Li, Jiang Zhong, Zhongxuan Xue, Qizhu Dai, Chen Wang, Xue Li</i> | |
| TeAw: Text-Aware Few-Shot Remote Sensing Image Scene Classification | 6296 |
| <i>Kaihui Cheng, Chule Yang, Zunlin Fan, Dayan Wu, Naiyang Guan</i> | |
| A Bidirectional Joint Model for Spoken Language Understanding..... | 6301 |
| <i>Nguyen Anh Tu, Duong Xuan Hieu, Tu Minh Phuong, Ngo Xuan Bach</i> | |
| Learning from Label Proportion with Online Pseudo-Label Decision by Regret Minimization..... | 6306 |
| <i>Shinnosuke Matsuo, Ryoma Bise, Seiichi Uchida, Daiki Suehiro</i> | |
| Blind Estimation of Audio Processing Graph..... | 6311 |
| <i>Sungho Lee, Jaehyun Park, Seungryeol Paik, Kyogu Lee</i> | |

| | |
|--|------|
| Show Me the Instruments: Musical Instrument Retrieval from Mixture Audio | 6316 |
| <i>Kyungsu Kim, Minju Park, Haesun Joung, Yunkee Chae, Yeongbeom Hong, Seonghyeon Go, Kyogu Lee</i> | |
| Movienet-PS: A Large-Scale Person Search Dataset in the Wild | 6321 |
| <i>Jie Qin, Peng Zheng, Yichao Yan, Rong Quan, Xiaogang Cheng, Bingbing Ni</i> | |
| LEAPT: Learning Adaptive Prefix-To-Prefix Translation for Simultaneous Machine Translation | 6326 |
| <i>Lei Lin, Shuangtao Li, Xiaodong Shi</i> | |
| Graph Signal Processing for Narrowband Direction of Arrival Estimation | 6331 |
| <i>Disheng Li, Wei Liu, Yuriy Zakharov, Paul D Mitchell</i> | |
| Discriminative Speaker Representation Via Contrastive Learning with Class-Aware Attention in Angular Space | 6336 |
| <i>Zhe Li, Man-Wai Mak, Helen Mei-Ling Meng</i> | |
| SDTN: Speaker Dynamics Tracking Network for Emotion Recognition in Conversation..... | 6341 |
| <i>Jiawei Chen, Peijie Huang, Guotai Huang, Qianer Li, Yuhong Xu</i> | |
| Long-Tailed Image Recognition with Dynamic Re-Weighting | 6346 |
| <i>Xinyuan Li, Yu Wang, Jien Kato</i> | |
| Adaptive Simulated Annealing Through Alternating Rényi Divergence Minimization | 6351 |
| <i>Thomas Guilmeau, Emilie Chouzenoux, Victor Elvira</i> | |
| Neural Diarization with Non-Autoregressive Intermediate Attractors | 6356 |
| <i>Yusuke Fujita, Tatsuya Komatsu, Robin Scheibler, Yusuke Kida, Tetsuji Ogawa</i> | |
| Monocular 3D Human Pose Estimation Based on Global Temporal-Attentive and Joints-Attention in Video | 6361 |
| <i>Ruhan He, Shanshan Xiang, Peng Tao, Yongsheng Yu</i> | |
| Event-Based Visual Microphone | 6366 |
| <i>Matthew Howard, Keigo Hirakawa</i> | |
| Domain and Language Adaptation Using Heterogeneous Datasets for Wav2vec2.0-Based Speech Recognition of Low-Resource Language | 6371 |
| <i>Kak Soky, Sheng Li, Chenhui Chu, Tatsuya Kawahara</i> | |
| Visual Graph Reasoning Network | 6376 |
| <i>Dingbang Li, Xin Lin, Haibin Cai, Wenzhou Chen</i> | |
| A Dynamic Cross-Scale Transformer with Dual-Compound Representation for 3D Medical Image Segmentation | 6381 |
| <i>Ruixia Zhang, Zhiqiong Wang, Zhongyang Wang, Junchang Xin</i> | |
| Multilevel Transformer for Multimodal Emotion Recognition | 6386 |
| <i>Junyi He, Meimei Wu, Meng Li, Xiaobo Zhu, Feng Ye</i> | |
| Preformer: Predictive Transformer with Multi-Scale Segment-Wise Correlations for Long-Term Time Series Forecasting | 6391 |
| <i>Dazhao Du, Bing Su, Zhewei Wei</i> | |
| Spatial Cross-Attention for Transformer-Based Image Captioning..... | 6396 |
| <i>Khoa Anh Ngo, Kyuhong Shim, Byonghyo Shim</i> | |

| | |
|---|------|
| Speaker-Aware Hierarchical Transformer for Personality Recognition in Multiparty Dialogues | 6401 |
| <i>Wenjing Han, Yirong Chen, Xiaofen Xing, Guohua Zhou, Xiangmin Xu</i> | |
| Ensemble Knowledge Distillation of Self-Supervised Speech Models | 6406 |
| <i>Kuan -Po Huang, Tzu-Hsun Feng, Yu-Kuan Fu, Tsu-Yuan Hsu, Po-Chieh Yen, Wei-Cheng Tseng, Kai-Wei Chang, Hung-Yi Lee</i> | |
| Enhancing and Adversarial: Improve ASR with Speaker Labels | 6411 |
| <i>Wei Zhou, Haotian Wu, Jingjing Xu, Mohammad Zeineldeen, Christoph Lüscher, Ralf Schlüter, Hermann Ney</i> | |
| LDTSF: A Label-Decoupling Teacher-Student Framework for Semi-Supervised Echocardiography Segmentation | 6416 |
| <i>Jiapeng Zhang, Yongxiong Wang, Zhiqun Pan, Zhenhui Tang, Lijun Chen, Jinlong Liu</i> | |
| Lexicon-Injected Semantic Parsing for Task-Oriented Dialog | 6421 |
| <i>Xiaojun Meng, Wenlin Dai, Yasheng Wang, Baojun Wang, Zhiyong Wu, Xin Jiang, Qun Liu</i> | |
| MASKED-AP: Attention Pyramid Convolutional Neural Network with Mask for Cervical Cell Classification | 6426 |
| <i>Yu Jin, Juan Liu, Hua Chen, Wensi Duan, Dehua Cao, Baochuan Pang</i> | |
| On the Value of Stochastic Side Information in Online Learning | 6431 |
| <i>Junzhang Jia, Xuetong Wu, Jamie Evans, Jingge Zhu</i> | |
| Frequency-Selective Hybrid Beamforming for Mmwave Full-Duplex | 6436 |
| <i>Andrea Guamo-Morocho, Roberto López-Valcarce</i> | |
| Towards Hyperbolic Regularizers for Point Cloud Part Segmentation | 6441 |
| <i>Antonio Montanaro, Diego Valsesia, Enrico Magli</i> | |
| Self-Sufficient Framework for Continuous Sign Language Recognition | 6446 |
| <i>Youngjoon Jang, Youngtaek Oh, Jae Won Cho, Myungchul Kim, Dong-Jin Kim, In So Kweon, Joon Son Chung</i> | |
| Leveraging Large Text Corpora for End-To-End Speech Summarization..... | 6451 |
| <i>Kohei Matsuura, Takanori Ashihara, Takafumi Moriya, Tomohiro Tanaka, Atsunori Ogawa, Marc Delcroix, Ryo Masumura</i> | |
| Topology Uncertainty Modeling for Imbalanced Node Classification on Graphs..... | 6456 |
| <i>Jiayi Gao, Jiaxing Li, Ke Zhang, Youyong Kong</i> | |
| A Dataset for Audio-Visual Sound Event Detection in Movies..... | 6461 |
| <i>Rajat Hebbar, Digbalay Bose, Krishna Somandepalli, Veena Vijai, Shrikanth Narayanan</i> | |
| DDN: Dynamic Aggregation Enhanced Dual-Stream Network for Medical Image Classification..... | 6466 |
| <i>Lang Wang, Juan Liu, Peng Jiang, Dehua Cao, Baochuan Pang</i> | |
| Enhancing Speech-To-Speech Translation with Multiple TTS Targets | 6471 |
| <i>Jiatong Shi, Yun Tang, Ann Lee, Hirofumi Inaguma, Changhan Wang, Juan Pino, Shinji Watanabe</i> | |
| CFFMixer: Multi-Dimensional Feature Fusion for Object Detection | 6476 |
| <i>Hao Xie, Weizhe Yuan, Bin Kang, Songlin Du</i> | |

| | |
|---|------|
| Real-Time Multichannel Speech Separation and Enhancement Using a Beamspace-Domain-Based Lightweight CNN | 6481 |
| <i>Marco Olivieri, Luca Comanducci, Mirco Pezzoli, Davide Balsarri, Luca Menescardi, Michele Buccoli, Simone Pecorino, Antonio Grosso, Fabio Antonacci, Augusto Sarti</i> | |
| Neural Mode Estimation..... | 6486 |
| <i>Peng Sun, Zhenyu Wen, Yejian Zhou, Zhen Hong, Tao Lin</i> | |
| Dynamic TF-TDNN: Dynamic Time Delay Neural Network Based on Temporal-Frequency Attention for Dialect Recognition | 6491 |
| <i>Chao Liao, Jinwen Huang, Huan Yuan, Peng Yao, Jianchao Tan, Dawei Zhang, Feng Deng, Xiaorui Wang, Chengru Song</i> | |
| Multi-Resolution Convolutional Dictionary Learning for Riverbed Dynamics Modeling..... | 6496 |
| <i>E. Kobayashi, H. Yasuda, K. Hayasaka, Y. Otake, S. Ono, S. Muramatsu</i> | |
| Grad-CAM-Inspired Interpretation of Nearfield Acoustic Holography Using Physics-Informed Explainable Neural Network | 6501 |
| <i>Hagar Kafri, Marco Olivieri, Fabio Antonacci, Mordehay Moradi, Augusto Sarti, Sharon Gannot</i> | |
| Double Compression Detection Based on the De-Blocking Filtering of HEVC Videos | 6506 |
| <i>Xiangui Kang, Pengcheng Su, Zisheng Huang, Yifang Chen, Jie Wang</i> | |
| Contrastive Learning of Functionality-Aware Code Embeddings | 6511 |
| <i>Yiyang Li, Hongqiu Wu, Hai Zhao</i> | |
| Smoothing Point Adjustment-Based Evaluation of Time Series Anomaly Detection | 6516 |
| <i>Mingyu Liu, Yijie Wang, Hongzuo Xu, Xiaohui Zhou, Bin Li, Yongjun Wang</i> | |
| Label-Efficient and Robust Learning from Multiple Experts..... | 6521 |
| <i>Bojan Kolosnjaji, Apostolis Zarras</i> | |
| General Or Specific? Investigating Effective Privacy Protection in Federated Learning for Speech Emotion Recognition..... | 6526 |
| <i>Chao Tan, Yang Cao, Sheng Li, Masatoshi Yoshikawa</i> | |
| Brain Network Features Differentiate Intentions from Different Emotional Expressions of the Same Text..... | 6531 |
| <i>Zhongjie Li, Bin Zhao, Gaoyan Zhang, Jianwu Dang</i> | |
| Diffusion-Based Sound Source Localization Using Networks of Planar Microphone Arrays | 6536 |
| <i>Davide Albertini, Gioele Greco, Alberto Bernardini, Augusto Sarti</i> | |
| Robust Knowledge Distillation from RNN-T Models with Noisy Training Labels Using Full-Sum Loss | 6541 |
| <i>Mohammad Zeineldeen, Kartik Audhkhasi, Murali Karthick Baskar, Bhuvana Ramabhadran</i> | |
| MGAT: Multi-Granularity Attention Based Transformers for Multi-Modal Emotion Recognition | 6546 |
| <i>Weiquan Fan, Xiaofen Xing, Bolun Cai, Xiangmin Xu</i> | |
| From English to More Languages: Parameter-Efficient Model Reprogramming for Cross-Lingual Speech Recognition | 6551 |
| <i>Chao-Han Huck Yang, Bo Li, Yu Zhang, Nanxin Chen, Rohit Prabhavalkar, Tara N. Sainath, Trevor Strohman</i> | |

| | |
|--|------|
| Wireless Sensing for Simultaneous Human Vocal Sound and Heart Sound Recognition..... | 6556 |
| <i>Yu Rong, Kumar Vijay Mishra, Daniel W. Bliss</i> | |
| Adaptive Gaussian Nested Filter for Parameter Estimation and State Tracking in Dynamical Systems..... | 6561 |
| <i>Sara Pérez-Vieites, Víctor Elvira</i> | |
| -Complexity Low-Rank Approximation SVD for Massive Matrix in Tensor Train Format..... | 6566 |
| <i>Jung-Chun Chi, Chiao-En Chen, Yuan-Hao Huang</i> | |
| Fast-U2++: Fast and Accurate End-to-End Speech Recognition in Joint CTC/Attention Frames | 6571 |
| <i>Chengdong Liang, Xiao-Lei Zhang, Binbin Zhang, Di Wu, Shengqiang Li, Xingchen Song, Zhendong Peng, Fuping Pan</i> | |
| Wekws: A Production First Small-Footprint End-to-End Keyword Spotting Toolkit..... | 6576 |
| <i>Jie Wang, Menglong Xu, Jingyong Hou, Binbin Zhang, Xiao-Lei Zhang, Lei Xie, Fuping Pan</i> | |
| Image Inpainting with Semantic-Aware Transformer..... | 6581 |
| <i>Shiyu Chen, Wenxin Yu, Qi Wang, Jun Gong, Peng Chen</i> | |
| Vararray Meets T-Sot: Advancing the State of the Art of Streaming Distant Conversational Speech Recognition | 6586 |
| <i>Naoyuki Kanda, Jian Wu, Xiaofei Wang, Zhuo Chen, Jinyu Li, Takuya Yoshioka</i> | |
| Meta Learning with Adaptive Loss Weight for Low-Resource Speech Recognition | 6591 |
| <i>Qiulin Wang, Wenxuan Hu, Lin Li, Qingyang Hong</i> | |
| An Asynchronous Updating Reinforcement Learning Framework for Task-Oriented Dialog System..... | 6596 |
| <i>Sai Zhang, Yuwei Hu, Xiaojie Wang, Caixia Yuan</i> | |
| Wave-U-Net Discriminator: Fast and Lightweight Discriminator for Generative Adversarial Network-Based Speech Synthesis | 6601 |
| <i>Takuhiro Kaneko, Hirokazu Kameoka, Kou Tanaka, Shogo Seki</i> | |
| An Application of Quantum Mechanics to Attention Methods in Computer Vision | 6606 |
| <i>Juntao Zhang, Yihao Luo, Peng Cheng, Zehan Li, Hao Wu, Kun Yu, Wenbo An, Jun Zhou</i> | |
| Robust Dominant Periodicity Detection for Time Series with Missing Data | 6611 |
| <i>Qingsong Wen, Linxiao Yang, Liang Sun</i> | |
| An Empirical Study of Backdoor Attacks on Masked Auto Encoders | 6616 |
| <i>Shuli Zhuang, Pengfei Xia, Bin Li</i> | |
| Joint Training of Hierarchical GANs and Semantic Segmentation for Expression Translation | 6621 |
| <i>Rumaysa Bodur, Binod Bhattarai, Tae-Kyun Kim</i> | |
| Deep Reinforcement Learning for Green UAV-Assisted Data Collection..... | 6626 |
| <i>Abhishek Mondal, Deepak Mishra, Ganesh Prasad, Ashraf Hossain</i> | |
| Delay-Aware Backpressure Routing Using Graph Neural Networks | 6631 |
| <i>Zhongyuan Zhao, Bojan Radojcic, Gunjan Verma, Ananthram Swami, Santiago Segarra</i> | |
| Deep Feature Aggregation for Lightweight Single Image Super-Resolution | 6636 |
| <i>Yanchun Li, Xinan He, Shujuan Tian, Zhetao Li, Saiqin Long</i> | |
| Class-Guided Triple Head Prediction Network for Long-Tail Object Detection..... | 6641 |
| <i>Xuyang Liu, Yuan Zheng</i> | |

| | |
|--|------|
| Sparse Bayesian Learning Based Three-Dimensional Imaging for Antenna Array Radar | 6646 |
| <i>Yuhan Li, Jesper Rindom Jensen, Maozhong Fu, Zhenmiao Deng, Mads Græsbøll Christensen</i> | |
| MCNET: Fuse Multiple Cues for Multichannel Speech Enhancement | 6651 |
| <i>Yujie Yang, Changsheng Quan, Xiaofei Li</i> | |
| Structural Optimization of Factor Graphs for Symbol Detection Via Continuous Clustering and Machine Learning..... | 6656 |
| <i>Lukas Rapp, Luca Schmid, Andrej Rode, Laurent Schmalen</i> | |
| TOLD: A Novel Two-Stage Overlap-Aware Framework for Speaker Diarization | 6661 |
| <i>Jiaming Wang, Zhihao Du, Shiliang Zhang</i> | |
| Domain Adaptation Without Catastrophic Forgetting on a Small-Scale Partially-Labeled Corpus for Speech Emotion Recognition | 6666 |
| <i>Zhi Zhu, Yoshinao Sato</i> | |
| Codes Correcting Burst and Arbitrary Erasures for Reliable and Low-Latency Communication..... | 6671 |
| <i>Serge Kas Hanna, Zhiyuan Tan, Wen Xu, Antonia Wachter-Zeh</i> | |
| Logo-Former: Local-Global Spatio-Temporal Transformer for Dynamic Facial Expression Recognition | 6676 |
| <i>Fuyan Ma, Bin Sun, Shutao Li</i> | |
| A Dual-Path Transformer Network for Scene Text Detection | 6681 |
| <i>Jingyu Lin, Yan Yan, Hanzi Wang</i> | |
| Capacity Maximization for Active RIS Assisted Outdoor-To-Indoor Communication System | 6686 |
| <i>Chen He, Weisheng Gong, Yangrui Dong, Xie Xie, Z. Jane Wang</i> | |
| A Multi-Channel Aggregation Framework for Object Detection in Large-Scale SAR Image..... | 6691 |
| <i>Chule Yang, Chao Zhang, Zunlin Fan, Zeting Yu, Qianchong Sun, Mengyuan Dai</i> | |
| Speech-Based Emotion Recognition with Self-Supervised Models Using Attentive Channel-Wise Correlations and Label Smoothing | 6696 |
| <i>Sofoklis Kakouros, Themis Stafylakis, Ladislav Mošner, Lukáš Burget</i> | |
| JSV-VC: Jointly Trained Speaker Verification and Voice Conversion Models | 6701 |
| <i>Shogo Seki, Hirokazu Kameoka, Kou Tanaka, Takuhiro Kaneko</i> | |
| How to Push the Fastest Model 50x Faster: Streaming Non-Autoregressive Speech Synthesis on Resouce-Limited Devices..... | 6706 |
| <i>Van-Thinh Nguyen, Hung-Cuong Pham, Dang-Khoa Mac</i> | |
| Hierarchical Spatiotemporal Feature Fusion Network for Video Saliency Prediction | 6711 |
| <i>Yunzuo Zhang, Tian Zhang, Cunyu Wu, Yuxin Zheng</i> | |
| Weavspeech: Data Augmentation Strategy for Automatic Speech Recognition Via Semantic-Aware Weaving..... | 6716 |
| <i>Kyusung Seo, Joonhyung Park, Jaeyun Song, Eunho Yang</i> | |
| An Automotive Radar Dataset for Object Classification..... | 6721 |
| <i>Akshad Shyam, Kusum Komalavally, Monika Gautam, Vamshikrishna Kancharla, Vennela Gudisa, Virendra Patil, Aanandh Balasubramanian, Sumohana Channappayya</i> | |

| | |
|---|------|
| Joint Microstrip Selection and Beamforming Design for MmWave Systems with Dynamic Metasurface Antennas | 6726 |
| <i>Wei Huang, Haiyang Zhang, Nir Shlezinger, Yonina C. Eldar</i> | |
| Subspace Modeling Enabled High-Sensitivity X-Ray Chemical Imaging | 6731 |
| <i>Jizhou Li, Bin Chen, Guibin Zan, Guannan Qian, Piero Pianetta, Yijin Liu</i> | |
| Multi-Carrier Wideband OCDM-Based THZ Automotive Radar..... | 6736 |
| <i>Sangeeta Bhattacharjee, Kumar Vijay Mishra, Ramesh Annavajjala, Chandra R. Murthy</i> | |
| Body Prior Guided Graph Convolutional Neural Network for Skeleton-Based Action Recognition..... | 6741 |
| <i>Qianshuo Hu, Hong Liu, Hua-Qiu Wang, Mengyuan Liu</i> | |
| Real-Time MRI Video Synthesis from Time Aligned Phonemes with Sequence-to-Sequence Networks | 6746 |
| <i>Sathvik Udupa, Prasanta Kumar Ghosh</i> | |
| A Highly Interpretable Deep Equilibrium Network for Hyperspectral Image Deconvolution | 6751 |
| <i>Alexandros Gkillas, Dimitris Ampeliotis, Kostas Berberidis</i> | |
| Improved Acoustic-To-Articulatory Inversion Using Representations from Pretrained Self-Supervised Learning Models | 6756 |
| <i>Sathvik Udupa, Siddarth C, Prasanta Kumar Ghosh</i> | |
| MPE4G : Multimodal Pretrained Encoder for Co-Speech Gesture Generation | 6761 |
| <i>Gwantae Kim, Seonghyeok Noh, Insung Ham, Hanseok Ko</i> | |
| Dynamic Split Computing for Efficient Deep EDGE Intelligence..... | 6766 |
| <i>Arian Bakhtiarnia, Nemanja Milošević, Qi Zhang, Dragana Bajovic, Alexandros Iosifidis</i> | |
| Multi-Agent Adversarial Training Using Diffusion Learning | 6771 |
| <i>Ying Cao, Elsa Rizk, Stefan Vlaski, Ali H. Sayed</i> | |
| Efficient Super-Resolution for Compression of Gaming Videos..... | 6776 |
| <i>Yifan Wang, Luka Murn, Luis Herranz, Fei Yang, Marta Mrak, Wei Zhang, Shuai Wan, Marc Gorriz Blanch</i> | |
| Peak-First CTC: Reducing the Peak Latency of CTC Models by Applying Peak-First Regularization | 6781 |
| <i>Zhengkun Tian, Hongyu Xiang, Min Li, Feifei Lin, Ke Ding, Guanglu Wan</i> | |
| From Easy to Hard: Two-Stage Selector and Reader for Multi-Hop Question Answering | 6786 |
| <i>Xin-Yi Li, Wei-Jun Lei, Yu-Bin Yang</i> | |
| Sparse Asynchronous Samples from Networks of Tems for Reconstruction of Classes of Non-Bandlimited Signals | 6791 |
| <i>Marek Hilton, Pier Luigi Dragotti</i> | |
| Time-Weighted Frequency Domain Audio Representation with GMM Estimator for Anomalous Sound Detection | 6796 |
| <i>Jian Guan, Youde Liu, Qiaoxi Zhu, Tieran Zheng, Jiqing Han, Wenwu Wang</i> | |
| DyLiteRADHAR: Dynamic Lightweight Slowfast Network for Human Activity Recognition Using MMWAVE Radar | 6801 |
| <i>Biyun Sheng, Yan Bao, Fu Xiao, Linqing Gui</i> | |

| | |
|---|------|
| Stereoscopic Video Retargeting Based on Camera Motion Classification | 6806 |
| <i>Linghui Cai, Zhenhua Tang</i> | |
| Pseudo Multi-Source Domain Extension and Selective Pseudo-Labeling for Unsupervised Domain Adaptive Medical Image Segmentation..... | 6811 |
| <i>Xiaokang Liu, Zhiqiang Wang, Kai Hu, Xieping Gao</i> | |
| Joint Unsupervised and Supervised Learning for Context-Aware Language Identification | 6816 |
| <i>Jinseok Park, Hyung Yong Kim, Jihwan Park, Byeong-Yeol Kim, Shukjae Choi, Yunkyu Lim</i> | |
| Adapted Multimodal Bert with Layer-Wise Fusion for Sentiment Analysis | 6821 |
| <i>Odysseas S. Chlapanis, Georgios Paraskevopoulos, Alexandros Potamianos</i> | |
| Knowledge-Augmented Frame Semantic Parsing with Hybrid Prompt-Tuning | 6826 |
| <i>Rui Zhang, Yajing Sun, Jingyuan Yang, Wei Peng</i> | |
| Energy Efficiency Maximization in RIS-Aided Networks with Global Reflection Constraints..... | 6831 |
| <i>Robert K. Foteck, Alessio Zappone, Marco Di Renzo</i> | |
| Joint Robust Representation and Generalization Enhancement for Cross-Modality Person Re-Identification | 6836 |
| <i>Heqing Cheng, Yong Feng, Mingliang Zhou, Xiancai Xiong, Yongheng Wang, Baohua Qiang</i> | |
| Subspace-Based Detector for Distributed Mmwave Mimo Radar Sensors | 6841 |
| <i>Moein Ahmadi, Mohammad Alae-Kerahroodi, Bhavani Shankar M. R., Björn Ottersten</i> | |
| Unsupervised Feature Selection with self-Weighted and $\ell_{2,0}$ -Norm Constraint | 6846 |
| <i>Yongjin Yuan, Zheng Wang, Feiping Nie, Xuelong Li</i> | |
| Revisit Sampling Theory of Bandlimited Graph Signals: One Bridge Between GSP and DSP | 6851 |
| <i>Fen Wang, Taihao Li, Xue Zhang</i> | |
| Multistage Spatial Context Models for Learned Image Compression | 6856 |
| <i>Fangzheng Lin, Heming Sun, Jinming Liu, Jiro Katto</i> | |
| Electric Network Frequency Detection Using Least Absolute Deviations | 6861 |
| <i>Christos Korgialas, Constantine Kotropoulos</i> | |
| D-3DLD: Depth-Aware Voxel Space Mapping for Monocular 3D Lane Detection with Uncertainty | 6866 |
| <i>Nayeon Kim, Moonsub Byeon, Daehyun Ji, Dokwan Oh</i> | |
| Unsupervised Pre-Training for Data-Efficient Text-to-Speech on Low Resource Languages | 6871 |
| <i>Seongyeon Park, Myungseo Song, Bohyung Kim, Tae-Hyun Oh</i> | |
| Sensor Selection for Angle of Arrival Estimation Based on the Two-Target Cramér-Rao Bound..... | 6876 |
| <i>Costas A. Kokke, Mario Coutino, Laura Anitori, Richard Heusdens, Geert Leus</i> | |
| Deep Autoencoding One-Class Time Series Anomaly Detection..... | 6881 |
| <i>Xudong Mou, Rui Wang, Tiejun Wang, Jie Sun, Bo Li, Tianyu Wo, Xudong Liu</i> | |
| Local-Global Siamese Network with Efficient Inter-Scale Feature Learning for Change Detection in VHR Remote Sensing Images | 6886 |
| <i>Yue Zhang, Tao Lei, Shaoxiong Han, Yetong Xu, Asoke K. Nandi</i> | |
| Signal Reconstruction for FMCW Radar Interference Mitigation Using Deep Unfolding | 6891 |
| <i>J. Overdevest, A. G. C. Koppelaar, M. J. G. Bekooij, J. Youn, R. J. G. Van Sloun</i> | |

| | |
|---|------|
| ShuffleAugment: A Data Augmentation Method Using Time Shuffling | 6896 |
| <i>Yoshinao Sato, Narumitsu Ikeda, Hirokazu Takahashi</i> | |
| F0 Estimation from Telephone Speech Using Deep Feature Loss..... | 6901 |
| <i>Supritha M Shetty, Shraddha Revankar, Nalini C Iyer, K T Deepak</i> | |
| Multiresolution Signal Processing of Financial Market Objects | 6906 |
| <i>Ioana Boier</i> | |
| Learning Hypergraphs from Signals with Dual Smoothness Prior..... | 6911 |
| <i>Bohan Tang, Siheng Chen, Xiaowen Dong</i> | |
| Wavsyncswap: End-To-End Portrait-Customized Audio-Driven Talking Face Generation | 6916 |
| <i>Weihong Bao, Liyang Chen, Chaoyong Zhou, Sicheng Yang, Zhiyong Wu</i> | |
| Duration-Aware Pause Insertion Using Pre-Trained Language Model for Multi-Speaker Text-To-Speech | 6921 |
| <i>Dong Yang, Tomoki Koriyama, Yuki Saito, Takaaki Saeki, Detai Xin, Hiroshi Saruwatari</i> | |
| WUDA: Unsupervised Domain Adaptation Based on Weak Source Domain Labels | 6926 |
| <i>Shengjie Liu, Chuang Zhu, Yuan Li, Wenqi Tang</i> | |
| Fine-Grained Private Knowledge Distillation | 6931 |
| <i>Yuntong Li, Shaowei Wang, Yingying Wang, Jin Li, Yuqiu Qian, Bangzhou Xin, Wei Yang</i> | |
| Two-Stage UNet with Multi-Axis Gated Multilayer Perceptron for Monaural Noisy-Reverberant Speech Enhancement..... | 6936 |
| <i>Zehua Zhang, Shiyun Xu, Xuyi Zhuang, Lianyu Zhou, Heng Li, Mingjiang Wang</i> | |
| Frequency and Scale Perspectives of Feature Extraction | 6941 |
| <i>Liangqi Zhang, Yihao Luo, Xiang Cao, Haibo Shen, Tianjiang Wang</i> | |
| Temporal Modeling Matters: A Novel Temporal Emotional Modeling Approach for Speech Emotion Recognition..... | 6946 |
| <i>Jiaxin Ye, Xin-Cheng Wen, Yujie Wei, Yong Xu, Kunhong Liu, Hongming Shan</i> | |
| Once-For-All Sequence Compression for Self-Supervised Speech Models..... | 6951 |
| <i>Hsuan-Jui Chen, Yen Meng, Hung-Yi Lee</i> | |
| Multitrack Music Transformer..... | 6956 |
| <i>Hao-Wen Dong, Ke Chen, Shlomo Dubnov, Julian McAuley, Taylor Berg-Kirkpatrick</i> | |
| Receptive Field Reliant Zero-Cost Proxies for Neural Architecture Search..... | 6961 |
| <i>Prateek Keserwani, Srinivas Soumitri Miriyala, Vikram N. Rajendiran, Pradeep N. Shivamurthappa</i> | |
| SMCL: Saliency Masked Contrastive Learning for Long-Tailed Visual Recognition | 6966 |
| <i>Sanglee Park, Seung-Won Hwang, Jungmin So</i> | |
| Multi-Dimensional and Multi-Scale Modeling for Speech Separation Optimized by Discriminative Learning | 6971 |
| <i>Zhaoxi Mu, Xinyu Yang, Wenjing Zhu</i> | |
| RIS Reflection and Placement Optimisation for Underlay D2D Communications in Cognitive Cellular Networks..... | 6976 |
| <i>Sarbani Ghose, Deepak Mishra, Santi P. Maity, George C. Alexandropoulos</i> | |

| | |
|---|------|
| Autonomous Navigation of a Robotic Swarm in Space Exploration Missions | 6981 |
| <i>Siwei Zhang, Tobias Baumgartner, Emanuel Staudinger, Robert Pöhlmann, Fabio Broghammer, Armin Dammann</i> | |
| Prosody-Controllable Spontaneous TTS with Neural HMMS | 6986 |
| <i>Harm Lameris, Shivam Mehta, Gustav Eje Henter, Joakim Gustafson, Éva Székely</i> | |
| Misspecified Cramér-Rao Bound of RIS-Aided Localization Under Geometry Mismatch | 6991 |
| <i>Pinjun Zheng, Hui Chen, Tarig Ballal, Henk Wymeersch, Tareq Y. Al-Naffouri</i> | |
| Noise-Disentanglement Metric Learning for Robust Speaker Verification | 6996 |
| <i>Yao Sun, Hanyi Zhang, Longbiao Wang, Kong Aik Lee, Meng Liu, Jianwu Dang</i> | |
| Gated Enhanced RPN and Hybrid-View for Few-Shot Object Detection | 7001 |
| <i>Xujun Wei, Zechu Zhou, Pinxue Guo, Wenqiang Zhang</i> | |
| Self-Supervised Audio-Visual Speaker Representation with Co-Meta Learning | 7006 |
| <i>Hui Chen, Hanyi Zhang, Longbiao Wang, Kong Aik Lee, Meng Liu, Jianwu Dang</i> | |
| Training Sound Event Detection with Soft Labels from Crowdsourced Annotations | 7011 |
| <i>Irene Martín-Morató, Manu Harju, Paul Ahokas, Annamaria Mesaros</i> | |
| DialogMI: A Dialogue Model Based on Enhancing Dialogue Mutual Information | 7016 |
| <i>Yibo Zhang, Ping Gong, Zelin Wang, Zhe Li, Xuanyuan Yang</i> | |
| Inter-Scale Sure-Let Denoise with Structured Deep Image Prior: Interpretable Self-Supervised Learning | 7021 |
| <i>Jikai Li, Shogo Muramatsu</i> | |
| Multichannel Time-Encoding of Finite-Rate-of-Innovation Signals | 7026 |
| <i>Abijith Jagannath Kamath, Chandra Sekhar Seelamantula</i> | |
| Autovocoder: Fast Waveform Generation from a Learned Speech Representation Using Differentiable Digital Signal Processing | 7031 |
| <i>Jacob J Webber, Cassia Valentini-Botinhao, Evelyn Williams, Gustav Eje Henter, Simon King</i> | |
| An Experimental Study on Sound Event Localization and Detection Under Realistic Testing Conditions | 7036 |
| <i>Shutong Niu, Jun Du, Qing Wang, Li Chai, Huaxin Wu, Zhaoxu Nian, Lei Sun, Yi Fang, Jia Pan, Chin-Hui Lee</i> | |
| Feature Selection and Text Embedding for Detecting Dementia from Spontaneous Cantonese | 7041 |
| <i>Xiaoquan Ke, Man-Wai Mak, Helen M. Meng</i> | |
| Transmit Energy Focusing for Parameter Estimation in Transmit Beamspace Slow-Time MIMO Radar | 7046 |
| <i>Tingting Zhang, Feng Xu, Sergiy A. Vorobyov</i> | |
| Vehicle View Synthesis by Generative Adversarial Network | 7051 |
| <i>Chan-Shuo Hu, Sung-Wei Tseng, Xin-Yun Fan, Chen-Kuo Chiang</i> | |
| Meeting Action Item Detection with Regularized Context Modeling | 7056 |
| <i>Jiaqing Liu, Chong Deng, Qinglin Zhang, Qian Chen, Wen Wang</i> | |
| A Momentum Two-Gradient Direction Algorithm with Variable Step Size Applied to Solve Practical Output Constraint Issue for Active Noise Control | 7061 |
| <i>Xiaoyi Shen, Dongyuan Shi, Zhengding Luo, Junwei Ji, Woon-Seng Gan</i> | |

| | |
|---|------|
| General Category Network: Handwritten Mathematical Expression Recognition with Coarse-Grained Recognition Task | 7066 |
| <i>Xinyu Zhang, Han Ying, Ye Tao, Youlu Xing, Guihuan Feng</i> | |
| Attention Mixup: An Accurate Mixup Scheme Based on Interpretable Attention Mechanism for Multi-Label Audio Classification | 7071 |
| <i>Wuyang Liu, Yanzhen Ren, Jingru Wang</i> | |
| MCNeT: Measurement-Consistent Networks Via a Deep Implicit Layer for Solving Inverse Problems..... | 7076 |
| <i>Rahul Mourya, João F. C. Mota</i> | |
| Efficient Privacy Preserving Graph Neural Network for Node Classification | 7081 |
| <i>Xinjun Pei, Xiaoheng Deng, Shengwei Tian, Kaiping Xue</i> | |
| Subgradient Descent Learning with Over-The-Air Computation..... | 7086 |
| <i>Tamir L. S. Gez, Kobi Cohen</i> | |
| Speech and Noise Dual-Stream Spectrogram Refine Network with Speech Distortion Loss for Robust Speech Recognition..... | 7091 |
| <i>Haoyu Lu, Nan Li, Tongtong Song, Longbiao Wang, Jianwu Dang, Xiaobao Wang, Shiliang Zhang</i> | |
| Joint Unmixing and Demosaicing Methods for Snapshot Spectral Images..... | 7096 |
| <i>Kinan Abbas, Matthieu Puigt, Gilles Delmaire, Gilles Roussel</i> | |
| SR-Init: An Interpretable Layer Pruning Method | 7101 |
| <i>Hui Tang, Yao Lu, Qi Xuan</i> | |
| On Cross-Layer Alignment for Model Fusion of Heterogeneous Neural Networks..... | 7106 |
| <i>Dang Nguyen, Trang Nguyen, Khai Nguyen, Dinh Phung, Hung Bui, Nhat Ho</i> | |
| Zone Plate Virtual Lenses for Memory-Constrained NLOS Imaging | 7111 |
| <i>Pablo Luesia-Lahoz, Diego Gutierrez, Adolfo Muñoz</i> | |
| Sparse Non-Contact Multiple People Localization and Vital Signs Monitoring Via FMCW Radar | 7116 |
| <i>Yonathan Eder, Zhuoyang Liu, Yonina C. Eldar</i> | |
| Knowledge Transfer for on-Device Speech Emotion Recognition with Neural Structured Learning | 7121 |
| <i>Yi Chang, Zhao Ren, Thanh Tam Nguyen, Kun Qian, Björn W. Schuller</i> | |
| Convolutional Recurrent Neural Networks for the Classification of Cetacean Bioacoustic Patterns..... | 7126 |
| <i>Dimitris N. Makropoulos, Antigoni Tsiami, Aristides Prospathopoulos, Dimitris Kassis, Alexandros Frantzis, Emmanuel Skarsoulis, George Piperakis, Petros Maragos</i> | |
| Joint Discriminator and Transfer Based Fast Domain Adaptation for End-To-End Speech Recognition | 7131 |
| <i>Hang Shao, Tian Tan, Wei Wang, Xun Gong, Yanmin Qian</i> | |
| Neural Source Coding for Bandwidth-Efficient Brain-Computer Interfacing with Wireless Neuro-Sensor Networks..... | 7136 |
| <i>Thomas Strypsteen, Alexander Bertrand</i> | |
| Clustered Greedy Algorithm for Large-Scale Sensor Selection | 7141 |
| <i>Kaushani Majumder, Sibi Raj B. Pillai, Satish Mulleti</i> | |

| | |
|---|------|
| End-To-End Classification of Cell-Cycle Stages with Center-Cell Focus Tracker Using Recurrent Neural Networks..... | 7146 |
| <i>Abin Jose, Rijo Roy, Dennis Eschweiler, Ina Laube, Reza Azad, Daniel Moreno-Andrés, Johannes Stegmaier</i> | |
| Frequency Bin-Wise Single Channel Speech Presence Probability Estimation Using Multiple DNNS..... | 7151 |
| <i>Shuai Tao, Himavanth Reddy, Jesper Rindom Jensen, Mads Græsbøll Christensen</i> | |
| SYNTACC : Synthesizing Multi-Accent Speech by Weight Factorization..... | 7156 |
| <i>Tuan-Nam Nguyen, Ngoc-Quan Pham, Alexander Waibel</i> | |
| Mutually Guided Few-Shot Learning for Relational Triple Extraction..... | 7161 |
| <i>Chengmei Yang, Shuai Jiang, Bowei He, Chen Ma, Lianghua He</i> | |
| Graph Contrastive Learning with Learnable Graph Augmentation..... | 7166 |
| <i>Xinyan Pu, Ke Zhang, Huazhong Shu, Jean Louis Coatrieux, Youyong Kong</i> | |
| Graph Signal Processing for Neuroimaging to Reveal Dynamics of Brain Structure-Function Coupling..... | 7171 |
| <i>Maria Giulia Preti, Thomas A. W. Bolton, Alessandra Griffa, Dimitri Van De Ville</i> | |
| Modelling Low-Resource Accents Without Accent-Specific TTS Frontend..... | 7176 |
| <i>Georgi Tinchev, Marta Czarnowska, Kamil Deja, Kayoko Yanagisawa, Marius Cotescu</i> | |
| DO-FAM: Disentangled Non-Linear Latent Navigation for Facial Attribute Manipulation..... | 7181 |
| <i>Yifan Yuan, Siteng Ma, Hongming Shan, Junping Zhang</i> | |
| Newton-Based Trainable Learning Rate..... | 7186 |
| <i>George Retsinas, Giorgos Sfikas, Panagiotis Paraskevas Filntisis, Petros Maragos</i> | |
| Stochastic Super-Resolution for Gaussian Textures..... | 7191 |
| <i>Émile Pierret, Bruno Galerne</i> | |
| Semantically-Informed Deep Neural Networks for Sound Recognition..... | 7196 |
| <i>Michele Esposito, Giancarlo Valente, Yenisel Plasencia-Calaña, Michel Dumontier, Bruno L. Giordano, Elia Formisano</i> | |
| F-PABEE: Flexible-Patience-Based Early Exiting for Single-Label and Multi-Label Text Classification Tasks..... | 7201 |
| <i>Xiangxiang Gao, Wei Zhu, Jiasheng Gao, Congrui Yin</i> | |
| Hierarchical Hypergraph Recurrent Attention Network for Temporal Knowledge Graph Reasoning..... | 7206 |
| <i>Jiayan Guo, Meiqi Chen, Yan Zhang, Jianqiang Huang, Zhiwei Liu</i> | |
| Topgformer: Topological-Based Graph Transformer for Mapping Brain Structural Connectivity to Functional Connectivity..... | 7211 |
| <i>Dalu Guo, Ke Zhang, Jiaying Li, Youyong Kong</i> | |
| MoLE : Mixture of Language Experts for Multi-Lingual Automatic Speech Recognition..... | 7216 |
| <i>Yoohwan Kwon, Soo-Whan Chung</i> | |
| Source-Filter HiFi-GAN: Fast and Pitch Controllable High-Fidelity Neural Vocoder..... | 7221 |
| <i>Reo Yoneyama, Yi-Chiao Wu, Tomoki Toda</i> | |
| Analysis of Noisy-Target Training for Dnn-Based Speech Enhancement..... | 7226 |
| <i>Takuya Fujimura, Tomoki Toda</i> | |

| | |
|---|------|
| Constrained Non-Negative PARAFAC2 for Electromyogram Separation | 7231 |
| <i>Abilé Magbonde, Franck Quaine, Bertrand Rivet</i> | |
| Rethinking the Reasonability of the Test Set for Simultaneous Machine Translation | 7236 |
| <i>Mengge Liu, Wen Zhang, Xiang Li, Jian Luan, Bin Wang, Yuhang Guo, Shuoying Chen</i> | |
| Streaming Joint Speech Recognition and Disfluency Detection | 7241 |
| <i>Hayato Futami, Emiru Tsunoo, Kentaro Shibata, Yosuke Kashiwagi, Takao Okuda, Siddhant Arora, Shinji Watanabe</i> | |
| Efficient Practices for Profile-To-Frontal Face Synthesis and Recognition | 7246 |
| <i>Huijiao Wang, Xulei Yang</i> | |
| Distance-Based Online Label Inference Attacks Against Split Learning | 7251 |
| <i>Junlin Liu, Xinchun Lyu</i> | |
| Dereverberation in Acoustic Sensor Networks Using Weighted Prediction Error with Microphone-Dependent Prediction Delays | 7256 |
| <i>Anselm Lohmann, Toon Van Waterschoot, Joerg Bitzer, Simon Doclo</i> | |
| Towards Adversarially Robust Continual Learning..... | 7261 |
| <i>Tao Bai, Chen Chen, Lingjuan Lyu, Jun Zhao, Bihan Wen</i> | |
| Learning Causal Representations for Generalizable Face Anti Spoofing | 7266 |
| <i>Guanghao Zheng, Yuchen Liu, Wenrui Dai, Chenglin Li, Junni Zou, Hongkai Xiong</i> | |
| Speech Emotion Recognition Based on Low-Level Auto-Extracted Time-Frequency Features | 7271 |
| <i>Ke Liu, Jingzhao Hu, Jun Feng</i> | |
| A Principled Approach to Model Validation in Domain Generalization..... | 7276 |
| <i>Boyang Lyu, Thuan Nguyen, Matthias Scheutz, Prakash Ishwar, Shuchin Aeron</i> | |
| X-SEPFORMER: End-To-End Speaker Extraction Network with Explicit Optimization on Speaker Confusion | 7281 |
| <i>Kai Liu, Ziqing Du, Xucheng Wan, Huan Zhou</i> | |
| SpeakerAugment: Data Augmentation for Generalizable Source Separation Via Speaker Parameter Manipulation | 7286 |
| <i>Kai Wang, Yuhang Yang, Hao Huang, Ying Hu, Sheng Li</i> | |
| SW-WAVENET: Learning Representation from Spectrogram and Wavegram Using Wavenet for Anomalous Sound Detection | 7291 |
| <i>Haihui Chen, Likai Ran, Xixia Sun, Chao Cai</i> | |
| Scale-Adaptive Tiny Object Detection Enhanced by Across-Scale and Shape-Preserved Semantic Location..... | 7296 |
| <i>Yuting He, Renjie Huang, Yangguang Shi, Guoqiang Xiao, Bin Yang, Yuqi Li</i> | |
| Enlightening the Student in Knowledge Distillation | 7301 |
| <i>Yujie Zheng, Chong Wang, Yi Chen, Jiangbo Qian, Jun Wang, Jiafei Wu</i> | |
| D-CONFORMER: Deformable Sparse Transformer Augmented Convolution for Voxel-Based 3D Object Detection..... | 7306 |
| <i>Xiao Zhao, Liuzhen Su, Xukun Zhang, Dingkang Yang, Mingyang Sun, Shunli Wang, Peng Zhai, Lihua Zhang</i> | |

| | |
|---|------|
| Adaptive Non-Local Generative Adversarial Networks for Low-Dose CT Image Denoising | 7311 |
| <i>Linlin Yang, Hongying Liu, Fanhua Shang, Yuanyuan Liu</i> | |
| Robust Binaural Sound Localisation with Temporal Attention | 7316 |
| <i>Qi Hu, Ning Ma, Guy J. Brown</i> | |
| Near-Field Localization with Dynamic Metasurface Antennas | 7321 |
| <i>Qianyu Yang, Anna Guerra, Francesco Guidi, Nir Shlezinger, Haiyang Zhang, Davide Dardari, Baoyun Wang, Yonina C. Eldar</i> | |
| NRTSI: Non-Recurrent Time Series Imputation | 7326 |
| <i>Siyuan Shan, Yang Li, Junier B. Oliva</i> | |
| Hierarchical Interactive Reconstruction Network for Video Compressive Sensing | 7331 |
| <i>Tong Zhang, Wenxue Cui, Chen Hui, Feng Jiang</i> | |
| Time-Frequency Awareness Network for Human Mesh Recovery from Videos | 7336 |
| <i>Boyang Zhang, Suping Wu, Meining Jia</i> | |
| Cardiac Disease Diagnosis on Imbalanced Electrocardiography Data Through Optimal Transport Augmentation | 7341 |
| <i>Jielin Qiu, Jiacheng Zhu, Mengdi Xu, Peide Huang, Michael Rosenberg, Douglas Weber, Emerson Liu, Ding Zhao</i> | |
| Estimating Uncertainty on Video Quality Metrics..... | 7346 |
| <i>Pierre David, Patrick Le Callet, Suiyi Ling, Haixiong Wang, Ioannis Katsavounidis, Zafar Shahid, Cosmin Stejerean</i> | |
| Hiding Speaker's Sex in Speech Using Zero-Evidence Speaker Representation in an Analysis/Synthesis Pipeline..... | 7351 |
| <i>Paul-Gauthier No , Xiaoxiao Miao, Xin Wang, Junichi Yamagishi, Jean-Fran ois Bonastre, Driss Matrouf</i> | |
| Predictive Skim: Contrastive Predictive Coding for Low-Latency Online Speech Separation | 7356 |
| <i>Chenda Li, Yifei Wu, Yanmin Qian</i> | |
| Topological Slepians: Maximally Localized Representations of Signals Over Simplicial Complexes | 7361 |
| <i>Claudio Battiloro, Paolo Di Lorenzo, Sergio Barbarossa</i> | |
| Quaternion Orthogonal Transformer for Facial Expression Recognition in the Wild | 7366 |
| <i>Yu Zhou, Liyuan Guo, Lianghai Jin</i> | |
| Integrating the Sensing and Radio Communications Channel Modelling from Radar Mutual Interference..... | 7371 |
| <i>Narcis Cardona, J. Samuel Romero, Wenfei Yang, Jian Li</i> | |
| HyperSteg: Hyperbolic Learning for Deep Steganography..... | 7376 |
| <i>Shivam Agarwal, Ritesh Soun, Rahul Shivani, Vishnu Varanasi, Navroop Gill, Ramit Sawhney</i> | |
| Gaze Pre-Train for Improving Disparity Estimation Networks..... | 7381 |
| <i>Ron M. Hecht, Ohad Rahamim, Shaul Oron, Andrea Forgacs, Gershon Celniker, Dan Levi, Omer Tsimhoni</i> | |
| Color Guided Depth Map Super-Resolution with Nonlocal Autoregressive Modeling..... | 7386 |
| <i>Wei Xu, Na Qi, Qing Zhu, Jingzhong Qi, Longlu Huang, Kun Cao, Yuxin Bao, Qianwen Wang</i> | |

| | |
|---|------|
| On Parametric Misspecified Bayesian Cramér-Rao Bound: An Application to Linear/Gaussian Systems..... | 7391 |
| <i>Shuo Tang, Gerald Lamountain, Tales Imbiriba, Pau Closas</i> | |
| A Novel Efficient Multi-View Traffic-Related Object Detection Framework..... | 7396 |
| <i>Kun Yang, Jing Liu, Dingkang Yang, Hanqi Wang, Peng Sun, Yanni Zhang, Yan Liu, Liang Song</i> | |
| Pseudo-Inverted Bottleneck Convolution for Darts Search Space | 7401 |
| <i>Arash Ahmadian, Louis S. P. Liu, Yue Fei, Konstantinos N. Plataniotis, Mahdi S. Hosseini</i> | |
| Multi-Head Feature Pyramid Networks for Breast Mass Detection | 7406 |
| <i>Hexiang Zhang, Zhenghua Xu, Dan Yao, Shuo Zhang, Junyang Chen, Thomas Lukasiewicz</i> | |
| On Neural Architectures for Deep Learning-Based Source Separation of Co-Channel OFDM Signals | 7411 |
| <i>Gary C. F. Lee, Amir Weiss, Alejandro Lancho, Yury Polyanskiy, Gregory W. Wornell</i> | |
| A Sidecar Separator Can Convert a Single-Talker Speech Recognition System to a Multi-Talker One | 7416 |
| <i>Lingwei Meng, Jiawen Kang, Mingyu Cui, Yuejiao Wang, Xixin Wu, Helen Meng</i> | |
| Learning from Single-Expert Annotated Labels for Automatic Sleep Staging | 7421 |
| <i>Zhiheng Luan, Yanzhen Ren, Li Peng, Xiong Chen, Xiuping Yang, Weiping Tu, Yuhong Yang</i> | |
| Label-Guided Contrastive Learning for Out-of-Domain Detection | 7426 |
| <i>Shun Zhang, Tongliang Li, Jiaqi Bai, Zhoujun Li</i> | |
| Ontology-Aware Network for Zero-Shot Sketch-Based Image Retrieval | 7431 |
| <i>Haoxiang Zhang, He Jiang, Ziqiang Wang, Deqiang Cheng</i> | |
| Fast Multiscale 3D Reconstruction Using Single-Photon Lidar Data | 7436 |
| <i>Sandor Plosz, Istvan Gyongy, Jonathan Leach, Stephen McLaughlin, Gerald S. Buller, Abderrahim Halimi</i> | |
| Quantile Online Learning for Semiconductor Failure Analysis..... | 7441 |
| <i>Bangjian Zhou, Pan Jieming, Maheswari Sivan, Aaron Voon-Yew Thean, J. Senthilnath</i> | |
| Distributed Signal Processing for Out-Of-System Interference Suppression in Cell-Free Massive MIMO..... | 7446 |
| <i>Zakir Hussain Shaik, Erik G. Larsson</i> | |
| An Auto-Encoder Based Method for Camera Fingerprint Compression | 7451 |
| <i>Kaixuan Zhang, Zihan Liu, Jiashang Hu, Shilin Wang</i> | |
| Combining Loss Reweighting and Sample Resampling for Long-Tailed Instance Segmentation | 7456 |
| <i>Yaochi Zhao, Sen Chen, Qiong Chen, Zhuhua Hu</i> | |
| Geometry-Aware DOA Estimation Using a Deep Neural Network with Mixed-Data Input Features..... | 7461 |
| <i>Ulrik Kowalk, Simon Doclo, Joerg Bitzer</i> | |
| Target Speaker Extraction with Ultra-Short Reference Speech by VE-VE Framework..... | 7466 |
| <i>Lei Yang, Wei Liu, Lufen Tan, Jaemo Yang, Han-Gil Moon</i> | |
| MarginNCE: Robust Sound Localization with a Negative Margin | 7471 |
| <i>Sooyoung Park, Arda Senocak, Joon Son Chung</i> | |

| | |
|---|------|
| Could the BubbleView Metaphor Be Used to Infer Visual Attention on 3D Graphical Content? | 7476 |
| <i>Alexandre Bruckert, Mona Abid, Matthieu Perreira Da Silva, Patrick Le Callet</i> | |
| An Interpretable Model Using Evidence Information for Multi-Hop Question Answering Over Long Texts | 7481 |
| <i>Yanyi Chen, Ruifang Liu, Xiyan Liu, Yidong Shi, Ge Bai</i> | |
| Self-Convolution for Automatic Speech Recognition | 7486 |
| <i>Tian-Hao Zhang, Qi Liu, Xinyuan Qian, Song-Lu Chen, Feng Chen, Xu-Cheng Yin</i> | |
| Multi-Dimensional Signal Recovery Using Low-Rank Deconvolution | 7491 |
| <i>David Reixach</i> | |
| Room Impulse Response Reconstruction Based on Spatio-Temporal-Spectral Features Learned from a Spherical Microphone Array Measurement | 7496 |
| <i>Amy Bastine, Thushara D. Abhayapala, Jihui Aimee Zhang</i> | |
| Estimating Normalized Graph Laplacians in Financial Markets | 7501 |
| <i>José Vinícius De M. Cardoso, Jiayi Ying, Sandeep Kumar, Daniel P. Palomar</i> | |
| Radio Map Based UAV Target Localization | 7506 |
| <i>Chen He, Weisheng Gong, Yangrui Dong, Xie Xie, Z. Jane Wang</i> | |
| Joint Estimation of Clustered User Activity and Correlated Channels with Unknown Covariance in mMTC | 7511 |
| <i>Hamza Djelouat, Markus Leinonen, Markku Juntti</i> | |
| ICEL: Learning with Inconsistent Explanations | 7516 |
| <i>Biao Liu, Xiaoyu Wu, Bo Yuan</i> | |
| ICStega: Image Captioning-Based Semantically Controllable Linguistic Steganography | 7521 |
| <i>Xilong Wang, Yaofei Wang, Kejiang Chen, Jinyang Ding, Weiming Zhang, Nenghai Yu</i> | |
| VF-Taco2: Towards Fast and Lightweight Synthesis for Autoregressive Models with Variation Autoencoder and Feature Distillation | 7526 |
| <i>Yuhao Liu, Cheng Gong, Longbiao Wang, Xixin Wu, Qiuyu Liu, Jianwu Dang</i> | |
| Evaluating Parameter-Efficient Transfer Learning Approaches on SURE Benchmark for Speech Understanding | 7531 |
| <i>Yingting Li, Ambuj Mehrish, Rishabh Bhardwaj, Navonil Majumder, Bo Cheng, Shuai Zhao, Amir Zadeh, Rada Mihalcea, Soujanya Poria</i> | |
| Multi-User Data Detection in Massive MIMO with 1-Bit ADCS | 7536 |
| <i>Amin Radbord, Italo Atzeni, Antti Tölli</i> | |
| Full-Band General Audio Synthesis with Score-Based Diffusion | 7541 |
| <i>Santiago Pascual, Gautam Bhattacharya, Chunghsin Yeh, Jordi Pons, Joan Serra</i> | |
| Phase-Aware Spoof Speech Detection Based on Res2net with Phase Network | 7546 |
| <i>Juntae Kim, Sung Min Ban</i> | |
| Subband Dependency Modeling for Sound Event Detection | 7551 |
| <i>Yadong Guan, Guibin Zheng, Jiqing Han, Huanliang Wang</i> | |
| Radio Sensing with Large Intelligent Surface for 6G | 7556 |
| <i>Cristian J. Vaca-Rubio, Pablo Ramirez-Espinosa, Kimmo Kansanen, Zheng-Hua Tan, Elisabeth De Carvalho</i> | |

| | |
|--|------|
| Av-Sepformer: Cross-Attention Sepformer for Audio-Visual Target Speaker Extraction | 7561 |
| <i>Jiuxin Lin, Xinyu Cai, Heinrich Dinkel, Jun Chen, Zhiyong Yan, Yongqing Wang, Junbo Zhang, Zhiyong Wu, Yujun Wang, Helen Meng</i> | |
| ZO-DARTS: Differentiable Architecture Search with Zeroth-Order Approximation..... | 7566 |
| <i>Lunchen Xie, Kaiyu Huang, Fan Xu, Qingjiang Shi</i> | |
| Gaussian Prior Reinforcement Learning for Nested Named Entity Recognition | 7571 |
| <i>Yawen Yang, Xuming Hu, Fukun Ma, Shu'Ang Li, Aiwei Liu, Lijie Wen, Philip S. Yu</i> | |
| Single-Shot Fractional Fourier Phase Retrieval..... | 7576 |
| <i>Yixiao Yang, Ran Tao</i> | |
| A Deep Temporal Factor Analysis Method for Large Scale Financial Portfolio Selection..... | 7581 |
| <i>Yao Zhou, Ruidan Su, Shikui Tu, Lei Xu</i> | |
| Role of Lexical Boundary Information in Chunk-Level Segmentation for Speech Emotion Recognition | 7586 |
| <i>Wei-Cheng Lin, Carlos Busso</i> | |
| Towards Zero-Shot Code-Switched Speech Recognition..... | 7591 |
| <i>Brian Yan, Matthew Wiesner, Ondrej Klejch, Preethi Jyothi, Shinji Watanabe</i> | |
| U-Shiftformer: Brain Tumor Segmentation Using a Shifted Attention Mechanism..... | 7596 |
| <i>Chih-Wei Lin, Zhongsheng Chen</i> | |
| Test Your Samples Jointly: Pseudo-Reference for Image Quality Evaluation | 7601 |
| <i>Marcelin Tworski, Stéphane Lathuilière</i> | |
| Exploring Binary Classification Loss for Speaker Verification..... | 7606 |
| <i>Bing Han, Zhengyang Chen, Yanmin Qian</i> | |
| Understanding Shared Speech-Text Representations | 7611 |
| <i>Gary Wang, Kyle Kastner, Ankur Bapna, Zhehuai Chen, Andrew Rosenberg, Bhuvana Ramabhadran, Yu Zhang</i> | |
| Dual-Feature Enhancement for Weakly Supervised Temporal Action Localization..... | 7616 |
| <i>Siyang Liu, Qiankun Liu, Qi Chu, Bin Liu, Nenghai Yu</i> | |
| Semantic Centralized Contrastive Learning for Unsupervised Hashing..... | 7621 |
| <i>Fengming Liang, Changlin Fan, Bo Xiao, Kongming Liang</i> | |
| Suffix Retrieval-Augmented Language Modeling..... | 7626 |
| <i>Zecheng Wang, Yik-Cheung Tam</i> | |
| Information and Sensing Beamforming Optimization for Multi-User Multi-Target MIMO ISAC Systems..... | 7631 |
| <i>Minghe Zhu, Lei Li, Shuqiang Xia, Tsung-Hui Chang</i> | |
| Multiple Signed Graph Learning for Gene Regulatory Network Inference | 7636 |
| <i>Abdullah Karaaslanli, Satabdi Saha, Tapabrata Maiti, Selin Aviyente</i> | |
| Low-Latency Electrolaryngeal Speech Enhancement Based on Fastspeech2-Based Voice Conversion and Self-Supervised Speech Representation | 7641 |
| <i>Kazuhiro Kobayashi, Tomoki Hayashi, Tomoki Toda</i> | |

| | |
|---|------|
| CORSD: Class-Oriented Relational Self Distillation | 7646 |
| <i>Muzhou Yu, Sia Huat Tan, Kailu Wu, Runpei Dong, Linfeng Zhang, Karsheng Ma</i> | |
| Evopose: A Recursive Transformer for 3D Human Pose Estimation with Kinematic Structure Priors | 7651 |
| <i>Yaqi Zhang, Yan Lu, Bin Liu, Zhiwei Zhao, Qi Chu, Nenghai Yu</i> | |
| Improved Wifi-Based Respiration Tracking Via Contrast Enhancement | 7656 |
| <i>Wei-Hsiang Wang, Xiaolu Zeng, Beibei Wang, Yexin Cao, K. J. Ray Liu</i> | |
| Investigating SINDy as a Tool for Causal Discovery in Time Series Signals | 7661 |
| <i>Andrew O'Brien, Rosina Weber, Edward Kim</i> | |
| Align, Write, Re-Order: Explainable End-to-End Speech Translation Via Operation Sequence Generation | 7666 |
| <i>Motoi Omachi, Brian Yan, Siddharth Dalmia, Yuya Fujita, Shinji Watanabe</i> | |
| Faster than Fast: Accelerating the Griffin-Lim Algorithm..... | 7671 |
| <i>Rossen Nenov, Dang-Khoa Nguyen, Peter Balazs</i> | |
| Deep Learning-Based Compressive Sampling Optimization in Massive MIMO Systems | 7676 |
| <i>Saidur R. Pavel, Yimin D. Zhang, Maria S. Greco, Fulvio Gini</i> | |
| MPS-AMS: Masked Patches Selection and Adaptive Masking Strategy Based Self-Supervised Medical Image Segmentation | 7681 |
| <i>Xiangtao Wang, Ruizhi Wang, Biao Tian, Jiaojiao Zhang, Shuo Zhang, Junyang Chen, Thomas Lukasiewicz, Zhenghua Xu</i> | |
| Robust Watermarking Scheme in Encrypted Domain Based on Integer Lifting Wavelet Transform and Compressed Sensing | 7686 |
| <i>Di Xiao, Qin Tang, Aozhu Zhao, Min Li</i> | |
| Multi-Resolution Sequence Aggregation and Model-Agnostic Framework for Time-Series Forecasting | 7691 |
| <i>Juhyun Lyu, Jinseok Yang, Junghee Kim, Woohyung Lim, Wonbin Ahn, Dongwan Kang, Minjae Kim, Nam Soo Kim</i> | |
| Dual-Graph Co-representation Learning for knowledge-Graph Enhanced Recommendation | 7696 |
| <i>Xinbiao Liu, Bin Liang, Junyu Niu, Chaofeng Sha, Dong Wu</i> | |
| Polarized Signal Singular Spectrum Analysis with Complex SSA..... | 7701 |
| <i>Sébastien Journé, Nicolas Le Bihan, Florent Chatelain, Julien Flamant</i> | |
| A Token-Level Contrastive Framework for Sign Language Translation | 7706 |
| <i>Biao Fu, Peigen Ye, Liang Zhang, Pei Yu, Cong Hu, Xiaodong Shi, Yidong Chen</i> | |
| Ultra Real-Time Portrait Matting Via Parallel Semantic Guidance..... | 7711 |
| <i>Xin Huang, Jiake Xie, Bo Xu, Han Huang, Ziwen Li, Cheng Lu, Yandong Guo, Yong Tang</i> | |
| Knowledge-Aware Bayesian Co-Attention for Multimodal Emotion Recognition | 7716 |
| <i>Zihan Zhao, Yu Wang, Yanfeng Wang</i> | |
| Densitytoken: Weakly-Supervised Crowd Counting with Density Classification..... | 7721 |
| <i>Zaiyi Hu, Binglu Wang, Xuelong Li</i> | |
| Contrastive Representation Learning for Acoustic Parameter Estimation..... | 7726 |
| <i>Philipp Götzt, Cagdas Tuna, Andreas Walther, Emanuel A. P. Habets</i> | |

| | |
|---|------|
| Compressive Estimation of Near Field Channels for Ultra Massive-Mimo Wideband THz Systems..... | 7731 |
| <i>Simon Tarboush, Anum Ali, Tareq Y. Al-Naffouri</i> | |
| Improving the out-Of-Distribution Generalization Capability of Language Models: Counterfactually-Augmented Data is Not Enough..... | 7736 |
| <i>Caoyun Fan, Wenqing Chen, Jidong Tian, Yitian Li, Hao He, Yaohui Jin</i> | |
| Towards Realizing the Value of Labeled Target Samples: A Two-Stage Approach for Semi- Supervised Domain Adaptation..... | 7741 |
| <i>Mengqun Jin, Kai Li, Shuyan Li, Chunming He, Xiu Li</i> | |
| Audio-Visual Speech Enhancement with a Deep Kalman Filter Generative Model | 7746 |
| <i>Ali Golmakani, Mostafa Sadeghi, Romain Serizel</i> | |
| WL-MSR: Watch and Listen for Multimodal Subtitle Recognition | 7751 |
| <i>Jiawei Liu, Hao Wang, Weining Wang, Xingjian He, Jing Liu</i> | |
| Deep Unfolding-Enabled Hybrid Beamforming Design for mmWave Massive MIMO Systems..... | 7756 |
| <i>Nhan Nguyen, Mengyuan Ma, Nir Shlezinger, Yonina C. Eldar, A. L. Swindlehurst, Markku Juntti</i> | |
| End-To-End Word-Level Disfluency Detection and Classification in Children’s Reading Assessment..... | 7761 |
| <i>Lavanya Venkatasubramaniam, Vishal Sunder, Eric Fosler-Lussier</i> | |
| Tell Model Where to Attend: Improving Interpretability of Aspect-Based Sentiment Classification Via Small Explanation Annotations..... | 7766 |
| <i>Zhenxiao Cheng, Jie Zhou, Wen Wu, Qin Chen, Liang He</i> | |
| Exploiting Multi-Decision and Deep Refinement for Ultrasound Image Segmentation | 7771 |
| <i>Wenjing Liu, Xuanya Li, Kai Hu, Xieping Gao</i> | |
| Learning from Positive and Unlabeled Data Using Observer-GAN..... | 7776 |
| <i>Omar Zamzam, Haleh Akrami, Richard Leahy</i> | |
| Transformer-Based Multi-Prototype Approach for Diabetic Macular Edema Analysis in OCT Images | 7781 |
| <i>Plácido L. Vidal, Joaquim De Moura, Jorge Novo, Marcos Ortega, Jaime S. Cardoso</i> | |
| On the Design and Training Strategies for Rnn-Based Online Neural Speech Separation Systems..... | 7786 |
| <i>Kai Li, Yi Luo</i> | |
| Learning How to Learn Domain-Invariant Parameters for Domain Generalization..... | 7791 |
| <i>Feng Hou, Yao Zhang, Yang Liu, Jin Yuan, Cheng Zhong, Yang Zhang, Zhongchao Shi, Jianping Fan, Zhiqiang He</i> | |
| Wespeaker: A Research and Production Oriented Speaker Embedding Learning Toolkit..... | 7796 |
| <i>Hongji Wang, Chengdong Liang, Shuai Wang, Zhengyang Chen, Binbin Zhang, Xu Xiang, Yanlei Deng, Yanmin Qian</i> | |
| Multi-Label Temporal Evidential Neural Networks for Early Event Detection | 7801 |
| <i>Xujiang Zhao, Xuchao Zhang, Chen Zhao, Jin-Hee Cho, Lance Kaplan, Dong Hyun Jeong, Audun Jøsang, Haifeng Chen, Feng Chen</i> | |
| Is Quality Enough? Integrating Energy Consumption in a Large-Scale Evaluation of Neural Audio Synthesis Models..... | 7806 |
| <i>Constance Douwes, Giovanni Bindi, Antoine Caillon, Philippe Esling, Jean-Pierre Briot</i> | |

| | |
|--|------|
| On Crowdsourcing-Design with Comparison Category Rating for Evaluating Speech Enhancement Algorithms..... | 7811 |
| <i>Angélica S. Z. Suárez, Clément Laroche, Line H. Clemmensen, Sneha Das</i> | |
| A Prototypical Semantic Decoupling Method Via Joint Contrastive Learning for Few-Shot Named Entity Recognition..... | 7816 |
| <i>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</i> | |
| Revisit Out-Of-Vocabulary Problem for Slot Filling: A Unified Contrastive Framework with Multi-Level Data Augmentations | 7821 |
| <i>Daichi Guo, Guanting Dong, Dayuan Fu, Yuxiang Wu, Chen Zeng, Tingfeng Hui, Liwen Wang, Xuefeng Li, Zechen Wang, Keqing He, Xinyue Cui, Weiran Xu</i> | |
| Trinet: Stabilizing Self-Supervised Learning from Complete Or Slow Collapse | 7826 |
| <i>Lixin Cao, Jun Wang, Ben Yang, Dan Su, Dong Yu</i> | |
| An Adapter Based Multi-Label Pre-Training for Speech Separation and Enhancement | 7831 |
| <i>Tianrui Wang, Xie Chen, Zhuo Chen, Shu Yu, Weibin Zhu</i> | |
| Large-Scale Language Model Rescoring on Long-Form Data..... | 7836 |
| <i>Tongzhou Chen, Cyril Allauzen, Yinghui Huang, Daniel Park, David Rybach, W. Ronny Huang, Rodrigo Cabrera, Kartik Audhkhasi, Bhuvana Ramabhadran, Pedro J. Moreno, Michael Riley</i> | |
| Contrastive Speech Mixup for Low-Resource Keyword Spotting | 7841 |
| <i>Dianwen Ng, Ruixi Zhang, Jia Qi Yip, Chong Zhang, Yukun Ma, Trung Hieu Nguyen, Chongjia Ni, Eng Siong Chng, Bin Ma</i> | |
| Low-Complexity Acoustic Echo Cancellation with Neural Kalman Filtering | 7846 |
| <i>Dong Yang, Fei Jiang, Wei Wu, Xuefei Fang, Muyong Cao</i> | |
| Learning on Entropy Coded Images with CNN..... | 7851 |
| <i>Rémi Piau, Thomas Maugey, Aline Roumy</i> | |
| Asymmetric Polynomial Loss for Multi-Label Classification..... | 7856 |
| <i>Yusheng Huang, Jiexing Qi, Xinbing Wang, Zhouhan Lin</i> | |
| MAID: A Conditional Diffusion Model for Long Music Audio Inpainting..... | 7861 |
| <i>Kaiyang Liu, Wendong Gan, Chenchen Yuan</i> | |
| Rumor Detection Via Assessing the Spreading Propensity of Users | 7866 |
| <i>Peng Zheng, Zhen Huang, Yong Dou, Yeqing Yan</i> | |
| Deep Network Series for Large-Scale High-Dynamic Range Imaging | 7871 |
| <i>Amir Aghabiglou, Matthieu Terris, Adrian Jackson, Yves Wiaux</i> | |
| Towards Improved Sonar Performance Using Environment-Informed Sparse Sub-Array Processing | 7876 |
| <i>Alexandre L'Her, Angélique Drémeau, Florent Le Courtois, Gaultier Real, Xavier Cristol, Yann Stéphane</i> | |
| Learning Properties of Holomorphic Neural Networks of Dual Variables | 7881 |
| <i>Dmitry Kozlov, Mikhail Bakulin, Stanislav Pavlov, Aleksandr Zuev, Mariya Krylova, Igor Kharchikov</i> | |

| | |
|--|------|
| An Effective Anomalous Sound Detection Method Based on Representation Learning with Simulated Anomalies | 7886 |
| <i>Han Chen, Yan Song, Zhu Zhuo, Yu Zhou, Yu-Hong Li, Hui Xue, Ian McLoughlin</i> | |
| Hearing and Seeing Abnormality: Self-Supervised Audio-Visual Mutual Learning for Deepfake Detection | 7891 |
| <i>Chang-Sung Sung, Jun-Cheng Chen, Chu-Song Chen</i> | |
| Sample-Efficient Robust MMV Recovery Algorithm | 7896 |
| <i>Yuvraj Singh, Jahnvi Singh Rohela, Satish Mulleti</i> | |
| Joint Channel and Direction Estimation for Ground-To-UAV Communications Enabled by a Simultaneous Reflecting and Sensing RIS | 7901 |
| <i>Jiguang He, Aymen Fakhreddine, George C. Alexandropoulos</i> | |
| Towards Low-Power Heart Rate Estimation Based on User’s Demographics and Activity Level for Wearables | 7906 |
| <i>Andre G. C. Pacheco, Frank A. C. Cabello, Adriana M. O. Fonoff, Paula G. Rodrigues, Otávio A. B. Penatti, Paula R. Pinto</i> | |
| Covariance Regularization for Probabilistic Linear Discriminant Analysis | 7911 |
| <i>Zhiyuan Peng, Mingjie Shao, Xuanji He, Xu Li, Tan Lee, Ke Ding, Guanglu Wan</i> | |
| I-Tuning: Tuning Frozen Language Models with Image for Lightweight Image Captioning | 7916 |
| <i>Ziyang Luo, Zhipeng Hu, Yadong Xi, Rongsheng Zhang, Jing Ma</i> | |
| Graph-Based Spectro-Temporal Dependency Modeling for Anti-Spoofing..... | 7921 |
| <i>Feng Chen, Shiwen Deng, Tieran Zheng, Yongjun He, Jiqing Han</i> | |
| Mixer: DNN Watermarking Using Image Mixup | 7926 |
| <i>Kassem Kallas, Teddy Furon</i> | |
| Leveraging Positional-Related Local-Global Dependency for Synthetic Speech Detection..... | 7931 |
| <i>Xiaohui Liu, Meng Liu, Longbiao Wang, Kong Aik Lee, Hanyi Zhang, Jianwu Dang</i> | |
| Improving Disfluency Detection with Multi-Scale Self Attention and Contrastive Learning..... | 7936 |
| <i>Peiyang Wang, Chaoqun Duan, Meng Chen, Xiaodong He</i> | |
| Improving Massively Multilingual ASR with Auxiliary CTC Objectives..... | 7941 |
| <i>William Chen, Brian Yan, Jiatong Shi, Yifan Peng, Soumi Maiti, Shinji Watanabe</i> | |
| Two-Branch Multi-Scale Deep Neural Network for Generalized Document Recapture Attack Detection | 7946 |
| <i>Jiaying Li, Chenqi Kong, Shiqi Wang, Haoliang Li</i> | |
| Resource Allocation for UAV-Enabled Integrated Sensing and Communication (ISAC) Via Multi-Objective Optimization | 7951 |
| <i>Omid Rezaei, Mohammad Mahdi Naghsh, Seyed Mohammad Karbasi, Mohammad Mahdi Nayebi</i> | |
| Disentangling the Horowitz Factor: Learning Content and Style from Expressive Piano Performance..... | 7956 |
| <i>Huan Zhang, Simon Dixon</i> | |
| Personalizing Federated Learning with Over-The-Air Computations | 7961 |
| <i>Zihan Chen, Zeshen Li, Howard H. Yang, Tony Q. S. Quek</i> | |

| | |
|---|------|
| Achieving Fair Speech Emotion Recognition Via Perceptual Fairness | 7966 |
| <i>Woan-Shiuan Chien, Chi-Chun Lee</i> | |
| Soft Label Coding for End-To-end Sound Source Localization with Ad-hoc Microphone Arrays | 7971 |
| <i>Linfeng Feng, Yijun Gong, Xiao-Lei Zhang</i> | |
| Towards Reliable Image Outpainting: Learning Structure-Aware Multimodal Fusion with Depth Guidance..... | 7976 |
| <i>Lei Zhang, Chunyu Lin, Kang Liao, Yao Zhao</i> | |
| SafeDeep: A Scalable Robustness Verification Framework for Deep Neural Networks | 7981 |
| <i>Anahita Baninajjar, Kamran Hosseini, Ahmed Rezine, Amir Aminifar</i> | |
| Exploiting Modality-Invariant Feature for Robust Multimodal Emotion Recognition with Missing Modalities..... | 7986 |
| <i>Haolin Zuo, Rui Liu, Jinming Zhao, Guanglai Gao, Haizhou Li</i> | |
| Front-End Adapter: Adapting Front-End Input of Speech Based Self-Supervised Learning for Speech Recognition | 7991 |
| <i>Xie Chen, Ziyang Ma, Changli Tang, Yujin Wang, Zhisheng Zheng</i> | |
| Target-Speaker Voice Activity Detection Via Sequence-to-Sequence Prediction..... | 7996 |
| <i>Ming Cheng, Weiqing Wang, Yucong Zhang, Xiaoyi Qin, Ming Li</i> | |
| Knowledge-Graph Augmented Music Representation for Genre Classification | 8001 |
| <i>Han Ding, Wenjing Song, Cui Zhao, Fei Wang, Ge Wang, Wei Xi, Jizhong Zhao</i> | |
| Pre-Training Strategies Using Contrastive Learning and Playlist Information for Music Classification and Similarity | 8006 |
| <i>Pablo Alonso-Jiménez, Xavier Favory, Hadrien Foughmand, Grigoris Bourdalas, Xavier Serra, Thomas Lidy, Dmitry Bogdanov</i> | |
| Assisted RTF-Vector-Based Binaural Direction of Arrival Estimation Exploiting a Calibrated External Microphone Array | 8011 |
| <i>Daniel Fejgin, Simon Doclo</i> | |
| Synthetic Pseudo Anomalies for Unsupervised Video Anomaly Detection: A Simple Yet Efficient Framework Based on Masked Autoencoder | 8016 |
| <i>Xiangyu Huang, Caidan Zhao, Chenxing Gao, Lvdong Chen, Zhiqiang Wu</i> | |
| Using Modified Adult Speech as Data Augmentation for Child Speech Recognition..... | 8021 |
| <i>Zijian Fan, Xinwei Cao, Giampiero Salvi, Torbjørn Svendsen</i> | |
| Phonetic Anchor-Based Transfer Learning to Facilitate Unsupervised Cross-Lingual Speech Emotion Recognition..... | 8026 |
| <i>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</i> | |
| Improving Retrieval-Based Dialogue System Via Syntax-Informed Attention..... | 8031 |
| <i>Tengtao Song, Nuo Chen, Ji Jiang, Zhihong Zhu, Yuxian Zou</i> | |
| Weakly-Supervised Scene-Specific Crowd Counting Using Real-Synthetic Hybrid Data..... | 8036 |
| <i>Yaowu Fan, Jia Wan, Yuan Yuan, Qi Wang</i> | |
| Phonation Mode Detection in Singing: A Singer Adapted Model | 8041 |
| <i>Yixin Wang, Wei Wei, Ye Wang</i> | |

| | |
|--|------|
| Improving the Stochastic Gradient Descent's Test Accuracy by Manipulating the ℓ_∞ Norm of Its Gradient Approximation | 8046 |
| <i>Paul Rodriguez</i> | |
| Semi-Supervised Domain Generalization with Graph-Based Classifier | 8051 |
| <i>Minxiang Ye, Yifei Zhang, Shiqiang Zhu, Anhuan Xie, Senwei Xiang</i> | |
| Zero-Shot Speech Emotion Recognition Using Generative Learning with Reconstructed Prototypes | 8056 |
| <i>Xinzhou Xu, Jun Deng, Zixing Zhang, Zhen Yang, Björn W. Schuller</i> | |
| Disambiguation of Cognitive Impairment Diagnosis with EEG-Based Dual-Contrastive Learning..... | 8061 |
| <i>Zhenxi Song, Zian Pei, Huixia Ren, Lin Zhu, Yi Guo, Zhiguo Zhang</i> | |
| CN-CVS: A Mandarin Audio-Visual Dataset for Large Vocabulary Continuous Visual to Speech Synthesis..... | 8066 |
| <i>Chen Chen, Dong Wang, Thomas Fang Zheng</i> | |
| Hierarchical Graph Learning for Stock Market Prediction Via a Domain-Aware Graph Pooling Operator..... | 8071 |
| <i>Arie N. Arya, Yao Lei Xu, L. Stankovic, Danilo P. Mandic</i> | |
| Flowgrad: Using Motion for Visual Sound Source Localization..... | 8076 |
| <i>Rajsuryan Singh, Pablo Zinemanas, Xavier Serra, Juan Pablo Bello, Magdalena Fuentes</i> | |
| A Generalized Subspace Distribution Adaptation Framework for Cross-Corpus Speech Emotion Recognition | 8081 |
| <i>Shaokai Li, Peng Song, Liang Ji, Yun Jin, Wenming Zheng</i> | |
| Simplicial Vector Autoregressive Model for Streaming Edge Flows | 8086 |
| <i>Joshin Krishnan, Rohan Money, Baltasar Beferull-Lozano, Elvin Isufi</i> | |
| A Radar-Jammer Zero-Sum Repeated Bayesian Game | 8091 |
| <i>Sofia Suvorova, Ali Pezeshki, Ross Kyprianou, Bill Moran</i> | |
| A Frequency-Weighted Leaky Fxlms Algorithm with Application to Feedback Active Noise Control Systems..... | 8096 |
| <i>Yu Tang, Hongwei Zhang</i> | |
| Speech Modeling with a Hierarchical Transformer Dynamical VAE | 8101 |
| <i>Xiaoyu Lin, Xiaoyu Bie, Simon Leglaive, Laurent Girin, Xavier Alameda-Pineda</i> | |
| Deep Unfolded Tensor Robust PCA with Self-Supervised Learning | 8106 |
| <i>Harry Dong, Megna Shah, Sean Donegan, Yuejie Chi</i> | |
| Improved Training of Mixture-Of-Experts Language GANs | 8111 |
| <i>Yekun Chai, Qiyue Yin, Junge Zhang</i> | |
| Dynamic Distributed Convex Optimization "Over-The-Air" in Decentralized Wireless Networks..... | 8116 |
| <i>Navneet Agrawal, Renato L. G. Cavalcante, Slawomir Stanczak</i> | |
| Study and Design of Robust Personal Sound Zones with Vast Using Low Rank Rirs | 8121 |
| <i>Sankha Subhra Bhattacharjee, Liming Shi, Guoli Ping, Xiaoxiang Shen, Mads Græsbøll Christensen</i> | |
| Online Binaural Speech Separation of Moving Speakers with a Wavesplit Network | 8126 |
| <i>Cong Han, Nima Mesgarani</i> | |

| | |
|---|------|
| Image Segmentation for Improved Lossless Screen Content Compression | 8131 |
| <i>Shabhrish Reddy Uddehal, Tilo Strutz, Hannah Och, André Kaup</i> | |
| Target Sound Extraction with Variable Cross-Modality Clues | 8136 |
| <i>Chenda Li, Yao Qian, Zhuo Chen, Dongmei Wang, Takuya Yoshioka, Shujie Liu, Yanmin Qian, Michael Zeng</i> | |
| Dynamic Chunk Convolution for Unified Streaming and Non-Streaming Conformer ASR..... | 8141 |
| <i>Xilai Li, Goeric Huybrechts, Srikanth Ronanki, Jeff Farris, Sravan Bodapati</i> | |
| Deep-Unfolded Adaptive Projected Subgradient Method for Mimo Detection..... | 8146 |
| <i>Jochen Fink, Renato L. G. Cavalcante, Zoran Utkovski, Slawomir Stanczak</i> | |
| Predicting Brain Age Using Transferable Covariance Neural Networks..... | 8151 |
| <i>Saurabh Sihag, Gonzalo Mateos, Corey McMillan, Alejandro Ribeiro</i> | |
| Retinal Biomarkers for Detecting Diabetic Retinopathy Using Smartphone-Based Deep Learning Frameworks | 8156 |
| <i>Mahmut Karakaya, Ramazan S. Aygun</i> | |
| Study of Manifold Geometry Using Multiscale Non-Negative Kernel Graphs..... | 8161 |
| <i>Carlos Hurtado, Sarath Shekkizhar, Javier Ruiz-Hidalgo, Antonio Ortega</i> | |
| Understandable Relu Neural Network for Signal Classification | 8166 |
| <i>Marie Guyomard, Susana Barbosa, Lionel Fillatre</i> | |
| A Video Anomaly Detection Framework Based on Appearance-Motion Semantics Representation Consistency | 8171 |
| <i>Xiangyu Huang, Caidan Zhao, Zhiqiang Wu</i> | |
| Syngen: A Syntactic Plug-And-Play Module for Generative Aspect-Based Sentiment Analysis | 8176 |
| <i>Chengze Yu, Taiqiang Wu, Jiayi Li, Xingyu Bai, Yujie Yang</i> | |
| TinyOOD: Effective out-Of-Distribution Detection for TinyML | 8181 |
| <i>Yongchang Li, Juncheng Jia, Yan Zuo, Weipeng Zhu</i> | |
| Radar Clutter Covariance Estimation: A Nonlinear Spectral Shrinkage Approach | 8186 |
| <i>Shashwat Jain, Vikram Krishnamurthy, Muralidhar Rangaswamy, Bosung Kang, Sandeep Gogineni</i> | |
| Gated Contextual Adapters for Selective Contextual Biasing in Neural Transducers | 8191 |
| <i>Anastasios Alexandridis, Kanthashree Mysore Sathyendra, Grant P. Strimel, Feng-Ju Chang, Ariya Rastrow, Nathan Susanj, Athanasios Mouchtaris</i> | |
| LightGrad: Lightweight Diffusion Probabilistic Model for Text-To-Speech | 8196 |
| <i>Jie Chen, Xingchen Song, Zhendong Peng, Binbin Zhang, Fuping Pan, Zhiyong Wu</i> | |
| VAN-ICP: GPU-Accelerated Approximate Nearest Neighbor Search for ICP Registration Via Voxel Dilation..... | 8201 |
| <i>Weimin Wang, Qiong Chang</i> | |
| Distance-Based Weight Transfer for Fine-Tuning from Near-Field to Far-Field Speaker Verification | 8206 |
| <i>Li Zhang, Qing Wang, Hongji Wang, Yue Li, Wei Rao, Yannan Wang, Lei Xie</i> | |
| Utilizing Wav2Vec in Database-Independent Voice Disorder Detection..... | 8211 |
| <i>Saska Tirronen, Farhad Javanmardi, Manila Kodali, Sudarsana Reddy Kadiri, Paavo Alku</i> | |

| | |
|---|------|
| CAT: Causal Audio Transformer for Audio Classification | 8216 |
| <i>Xiaoyu Liu, Hanlin Lu, Jianbo Yuan, Xinyu Li</i> | |
| Water Leak Detection and Localization Using Convolutional Autoencoders..... | 8221 |
| <i>Daniele Ugo Leonzio, Paolo Bestagini, Marco Marcon, Gian Paolo Quarta, Stefano Tubaro</i> | |
| Multimodal Emotion Recognition Based on Deep Temporal Features Using Cross-Modal Transformer and Self-Attention..... | 8226 |
| <i>Bubai Maji, Monorama Swain, Rajlakshmi Guha, Aurobinda Routray</i> | |
| Reliability Estimation for Synthetic Speech Detection | 8231 |
| <i>Davide Salvi, Paolo Bestagini, Stefano Tubaro</i> | |
| VPPT: Visual Pre-Trained Prompt Tuning Framework for Few-Shot Image Classification | 8236 |
| <i>Zhao Song, Ke Yang, Naiyang Guan, Junjie Zhu, Peng Qiao, Qingyong Hu</i> | |
| LINK: Linguistic Steganalysis Framework with External Knowledge | 8241 |
| <i>Jinshuai Yang, Zhongliang Yang, Xinrui Ge, Jiajun Zou, Yue Gao, Yongfeng Huang</i> | |
| Rate Splitting and Precoding Strategies for Multi-User MIMO Broadcast Channels with Common and Private Streams | 8246 |
| <i>Liana Khamidullina, André L. F. De Almeida, Martin Haardt</i> | |
| Sparse Bayesian Learning Assisted Decision Fusion in Millimeter Wave Massive MIMO Sensor Networks | 8251 |
| <i>Apoorva Chawla, Domenico Ciuonzo, Pierluigi Salvo Rossi</i> | |
| Robust Adaptive Beamforming with Proximal Method | 8256 |
| <i>Ruifu Li, Danijela Cabric</i> | |
| Elliptical Wishart Distribution: Maximum Likelihood Estimator from Information Geometry..... | 8261 |
| <i>Imen Ayadi, Florent Bouchard, Frédéric Pascal</i> | |
| Range-ISL Minimization and Spectral Shaping in MIMO Radar Systems Via Waveform Design..... | 8266 |
| <i>Ehsan Raei, Mohammad Alae-Kerahroodi, Bhavani Shankar, Björn Ottersten</i> | |
| Self-Attention for Enhanced OAMP Detection in MIMO Systems..... | 8271 |
| <i>Alexander Fuchs, Christian Knoll, Nima N. Moghadam, Alexey Pak, Jinliang Huang, Erik Leitinger, Franz Pernkopf</i> | |
| Modeling Global Latent Semantic in Multi-Turn Conversations with Random Context Reconstruction..... | 8276 |
| <i>Chengwen Zhang, Danqin Wu</i> | |
| Optimizing Vision Transformers for Medical Image Segmentation..... | 8281 |
| <i>Qianying Liu, Chaitanya Kaul, Jun Wang, Christos Anagnostopoulos, Roderick Murray-Smith, Fani Deligianni</i> | |
| Multimodal Dyadic Impression Recognition Via Listener Adaptive Cross-Domain Fusion..... | 8286 |
| <i>Yuanchao Li, Peter Bell, Catherine Lai</i> | |
| Universal Speaker Recognition Encoders for Different Speech Segments Duration | 8291 |
| <i>Sergey Novoselov, Vladimir Volokhov, Galina Lavrentyeva</i> | |
| Unsupervised Speaker Verification Using Pre-Trained Model and Label Correction | 8296 |
| <i>Zhicong Chen, Jie Wang, Wenxuan Hu, Lin Li, Qingyang Hong</i> | |

| | |
|--|------|
| High-Resolution Embedding Extractor for Speaker Diarisation | 8301 |
| <i>Hee-Soo Heo, Youngki Kwon, Bong-Jin Lee, You Jin Kim, Jee-Weon Jung</i> | |
| Targeted Adversarial Attacks Against Neural Machine Translation | 8306 |
| <i>Sahar Sadrizadeh, Amirhossein Dabiri Aghdam, Ljiljana Dolamic, Pascal Frossard</i> | |
| Deep Learning Sparse Array Design Using Binary Switching Configurations | 8311 |
| <i>Syed A. Hamza, Kyle Juretus, Moeness G. Amin, Fauzia Ahmad</i> | |
| Sparsity-Driven Joint Blind Deconvolution-Demodulation with Application to Motor Fault Detection | 8316 |
| <i>Varun A. Kelkar, Dehong Liu, Hiroshi Inoue, Makoto Kanemaru</i> | |
| Papez: Resource-Efficient Speech Separation with Auditory Working Memory | 8321 |
| <i>Hyunseok Oh, Juheon Yi, Youngki Lee</i> | |
| Efficient Intelligibility Evaluation Using Keyword Spotting: A Study on Audio-Visual Speech Enhancement | 8326 |
| <i>Cassia Valentini-Botinhao, Andrea Lorena Aldana Blanco, Ondrej Klejch, Peter Bell</i> | |
| A Meta-Gnn Approach to Personalized Seizure Detection and Classification | 8331 |
| <i>Abdellah Rahmani, Arun Venkitaraman, Pascal Frossard</i> | |
| Community Detection Graph Convolutional Network for Overlap-Aware Speaker Diarization | 8336 |
| <i>Jie Wang, Zhicong Chen, Haodong Zhou, Lin Li, Qingyang Hong</i> | |
| Smart Split-Federated Learning Over Noisy Channels for Embryo Image Segmentation | 8341 |
| <i>Zahra Hafezi Kafshgari, Ivan V. Bajic, Parvaneh Saeedi</i> | |
| Variational Inference Aided Estimation of Time Varying Channels | 8346 |
| <i>Benedikt Böck, Michael Baur, Valentina Rizzello, Wolfgang Utschick</i> | |
| Online Learning-Based Waveform Selection for Improved Vehicle Recognition in Automotive Radar | 8351 |
| <i>Charles E. Thornton, William W. Howard, R. Michael Buehrer</i> | |
| Gator: Graph-Aware Transformer with Motion-Disentangled Regression for Human Mesh Recovery from a 2D Pose | 8356 |
| <i>Yingxuan You, Hong Liu, Xia Li, Wenhao Li, Ti Wang, Runwei Ding</i> | |
| Bayesian Network Modeling and Prediction of Transitions Within the Homelessness System | 8361 |
| <i>Khandker Sadia Rahman, Daphney-Stavroula Zois, Charalampos Chelmiss</i> | |
| Distributed Adaptive Norm Estimation for Blind System Identification in Wireless Sensor Networks | 8366 |
| <i>M. Blochberger, F. Elvander, R. Ali, J. Østergaard, J. Jensen, M. Moonen, T. Van Waterschoot</i> | |
| Diffusion-Based Generative Speech Source Separation | 8371 |
| <i>Robin Scheibler, Youna Ji, Soo-Whan Chung, Jaekuk Byun, Soyeon Choe, Min-Seok Choi</i> | |
| Lost in Translation: Generating Adversarial Examples Robust to Round-Trip Translation | 8376 |
| <i>Neel Bhandari, Pin-Yu Chen</i> | |
| Relapse Prediction from Long-Term Wearable Data Using Self-Supervised Learning and Survival Analysis | 8381 |
| <i>E. Fekas, A. Zlatintsi, P. P. Filintisis, C. Garoufis, N. Efthymiou, P. Maragos</i> | |

| | |
|--|------|
| On the Relevance of the Differences Between HRTF Measurement Setups for Machine Learning | 8386 |
| <i>Johan Pauwels, Lorenzo Picinali</i> | |
| IR-ECG: Invertible Reconstruction of ECG..... | 8391 |
| <i>Peng Wang, Xi Huang, Li Cui</i> | |
| Robustdistiller: Compressing Universal Speech Representations for Enhanced Environment Robustness..... | 8396 |
| <i>Heitor R. Guimarães, Arthur Pimentel, Anderson R. Avila, Mehdi Rezagholizadeh, Boxing Chen, Tiago H. Falk</i> | |
| Dynamic Fair Node Representation Learning | 8401 |
| <i>Oyku Deniz Kose, Yanning Shen</i> | |
| Speech Intelligibility Classifiers from 550k Disordered Speech Samples..... | 8406 |
| <i>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</i> | |
| Disentangled Feature Learning for Real-Time Neural Speech Coding | 8411 |
| <i>Xue Jiang, Xiulian Peng, Yuan Zhang, Yan Lu</i> | |
| Neural Maximum-A-Posteriori Beamforming for Ultrasound Imaging..... | 8416 |
| <i>Ben Luijten, Boudewine W. Ossenkuppele, Nico De Jong, Martin D. Verweij, Yonina C. Eldar, Massimo Mischi, Ruud J. G. Van Sloun</i> | |
| Collaborative Audio-Visual Event Localization Based on Sequential Decision and Cross-Modal Consistency | 8421 |
| <i>Yuqian Kuang, Xiaopeng Fan</i> | |
| Accelerated Massive MIMO Detector Based on Annealed Underdamped Langevin Dynamics..... | 8426 |
| <i>Nicolas Zilberstein, Chris Dick, Rahman Doost-Mohammady, Ashutosh Sabharwal, Santiago Segarra</i> | |
| Repetition Counting from Compressed Videos Using Sparse Residual Similarity | 8431 |
| <i>Rishabh Khurana, Jayesh Rajkumar Vachhani, Sourabh Vasant Gothe, Pranay Kashyap</i> | |
| Clean Sample Guided Self-Knowledge Distillation for Image Classification..... | 8436 |
| <i>Jiyue Wang, Yanxiong Li, Qianhua He, Wei Xie</i> | |
| LGVIT: Local-Global Vision Transformer for Breast Cancer Histopathological Image Classification..... | 8441 |
| <i>Lang Wang, Juan Liu, Peng Jiang, Dehua Cao, Baochuan Pang</i> | |
| Robust Angle Estimation for Hybrid mmWave Systems..... | 8446 |
| <i>Yuan-Pei Lin, Ting-Ming Yang</i> | |
| HiSSNet: Sound Event Detection and Speaker Identification Via Hierarchical Prototypical Networks for Low-Resource Headphones..... | 8451 |
| <i>N Shashaank, Berker Banar, Mohammad Rasool Izadi, Jeremy Kemmerer, Shuo Zhang, Chuan-Che Jeff Huang</i> | |
| Bayesian Cramér-Rao Bound Estimation with Score-Based Models..... | 8456 |
| <i>Evan Scope Crafts, Bo Zhao</i> | |
| Enabling Large-Scale Image Search with Co-Attention Mechanism | 8461 |
| <i>Zechao Hu, Adrian G. Bors</i> | |

| | |
|---|------|
| Dialog Act Guided Contextual Adapter for Personalized Speech Recognition | 8466 |
| <i>Feng-Ju Chang, Thejaswi Muniyappa, Kanthashree Mysore Sathyendra, Kai Wei, Grant P. Strimel, Ross McGowan</i> | |
| LMCodec: A Low Bitrate Speech Codec with Causal Transformer Models | 8471 |
| <i>Teerapat Jenrungrot, Michael Chinen, W. Bastiaan Kleijn, Jan Skoglund, Zalán Borsos, Neil Zeghidour, Marco Tagliasacchi</i> | |
| SEPDIFF: Speech Separation Based on Denoising Diffusion Model | 8476 |
| <i>Bo Chen, Chao Wu, Wenbin Zhao</i> | |
| Few but Informative Local Hash Code Matching for Image Retrieval | 8481 |
| <i>Zechao Hu, Adrian G. Bors</i> | |
| Vision, Deduction and Alignment: An Empirical Study on Multi-Modal Knowledge Graph Alignment..... | 8486 |
| <i>Yangning Li, Jiaoyan Chen, Yinghui Li, Yuejia Xiang, Xi Chen, Hai-Tao Zheng</i> | |
| Low-Resource Music Genre Classification with Cross-Modal Neural Model Reprogramming | 8491 |
| <i>Yun-Ning Hung, Chao-Han Huck Yang, Pin-Yu Chen, Alexander Lerch</i> | |
| MAST: Multiscale Audio Spectrogram Transformers | 8496 |
| <i>Sreyan Ghosh, Ashish Seth, S. Umesh, Dinesh Manocha</i> | |
| Dynamic Scalable Self-Attention Ensemble for Task-Free Continual Learning | 8501 |
| <i>Fei Ye, Adrian G. Bors</i> | |
| Stuart: Individualized Classroom Observation of Students with Automatic Behavior Recognition and Tracking | 8506 |
| <i>Huayi Zhou, Fei Jiang, Jiaxin Si, Lili Xiong, Hongtao Lu</i> | |
| Reducing the Communication and Computational Cost of Random Fourier Features Kernel LMS in Diffusion Networks | 8511 |
| <i>Daniel G. Tigea, Renato Candido, Luis A. Azpicueta-Ruiz, Magno T. M. Silva</i> | |
| SLICER: Learning Universal Audio Representations Using Low-Resource Self-Supervised Pre-Training | 8516 |
| <i>Ashish Seth, Sreyan Ghosh, S. Umesh, Dinesh Manocha</i> | |
| Deep Proximal Gradient Method for Learned Convex Regularizers..... | 8521 |
| <i>Aaron Berk, Yanting Ma, Petros Boufounos, Pu Wang, Hassan Mansour</i> | |
| Analysis and Transformation of Voice Level in Singing Voice | 8526 |
| <i>Frederik Bous, Axel Roebel</i> | |
| Data-Driven Graph Convolutional Neural Networks for Power System Contingency Analysis..... | 8531 |
| <i>Valentin Bolz, Johannes Rueß, Andreas Zell</i> | |
| Robust Online Multiband Drift Estimation in Electrophysiology Data..... | 8536 |
| <i>Charlie Windolf, Angelique C. Paulk, Yoav Kfir, Eric Trautmann, Domokos Mészéna, William Muñoz, Irene Caprara, Mohsen Jamali, Julien Boussard, Ziv M. Williams, Sydney S. Cash, Liam Paninski, Erdem Varol</i> | |
| Consistent Estimators of a New Class of Covariance Matrix Distances in the Large Dimensional Regime | 8541 |
| <i>Roberto Pereira, Xavier Mestre, David Gregoratti</i> | |

| | |
|--|------|
| Make More of Your Data: Minimal Effort Data Augmentation for Automatic Speech Recognition and Translation | 8546 |
| <i>Tsz Kin Lam, Shigehiko Schamoni, Stefan Riezler</i> | |
| Compressing Cross-Domain Representation Via Lifelong Knowledge Distillation..... | 8551 |
| <i>Fei Ye, Adrian G. Bors</i> | |
| Phase Unwrapping in Correlated Noise for FMCW Lidar Depth Estimation..... | 8556 |
| <i>A. Ulvog, J. Rapp, T. Koike-Akino, H. Mansour, P. Boufounos, K. Parsons</i> | |
| A Bayesian Perspective on Noise2Noise: Theory and Extensions | 8561 |
| <i>Sarah Miller, Christina Karam, Achour Idoughi, Kodai Kikuchi, Keigo Hirakawa</i> | |
| Self-Similarity is All You Need for Fast and Light-Weight Generic Event Boundary Detection | 8566 |
| <i>Sourabh Vasant Gothe, Jayesh Rajkumar Vachhani, Rishabh Khurana, Pranay Kashyap</i> | |
| On Batching Variable Size Inputs for Training End-To-End Speech Enhancement Systems | 8571 |
| <i>Philippe Gonzalez, Tommy Sonne Alstrøm, Tobias May</i> | |
| Learning to Reconnect Interrupted Trajectories for Weakly Supervised Multi-Object Tracking | 8576 |
| <i>Yu-Lei Li, Yang Lu, Jie Li, Hanzi Wang</i> | |
| Textless Direct Speech-To-Speech Translation with Discrete Speech Representation | 8581 |
| <i>Xinjian Li, Ye Jia, Chung-Cheng Chiu</i> | |
| Improving Self-Supervised Learning for Audio Representations by Feature Diversity and Decorrelation | 8586 |
| <i>Bac Nguyen, Stefan Uhlich, Fabien Cardinaux</i> | |
| Asymptotically Optimal Nonparametric Classification Rules for Spike Train Data | 8591 |
| <i>Mirosław Pawlak, Mateusz Pabian, Dominik Rzepka</i> | |
| Fully Complex-Valued Deep Learning Model for Visual Perception | 8596 |
| <i>Aniruddh Sikdar, Sumanth Udupa, Suresh Sundaram</i> | |
| Beamformer-Guided Target Speaker Extraction..... | 8601 |
| <i>Mohamed Elminshawi, Srikanth Raj Chetupalli, Emanuel A. P. Habets</i> | |
| Context-Aware Coherent Speaking Style Prediction with Hierarchical Transformers for Audiobook Speech Synthesis | 8606 |
| <i>Shun Lei, Yixuan Zhou, Liyang Chen, Zhiyong Wu, Shiyin Kang, Helen Meng</i> | |
| Less is More: A Unified Architecture for Device-Directed Speech Detection with Multiple Invocation Types | 8611 |
| <i>Oggi Rudovic, Wonil Chang, Vineet Garg, Pranay Dighe, Pramod Simha, Jack Berkowitz, Ahmed H. Abdelaziz, Sachin Kajarekar, Erik Marchi, Saurabh Adya</i> | |
| Lasso-Based Fast Residual Recovery for Modulo Sampling | 8616 |
| <i>Shaik Basheeruddin Shah, Satish Mulleti, Yonina C. Eldar</i> | |
| Ensemble Prosody Prediction for Expressive Speech Synthesis | 8621 |
| <i>Tian Huey Teh, Vivian Hu, Devang S Ram Mohan, Zack Hodari, Christopher G. R. Wallis, Tomás Gómez Ibarrondo, Alexandra Torresquintero, James Leoni, Mark Gales, Simon King</i> | |
| Massively Multilingual Shallow Fusion with Large Language Models | 8626 |
| <i>Ke Hu, Tara N. Sainath, Bo Li, Nan Du, Yanping Huang, Andrew M. Dai, Yu Zhang, Rodrigo Cabrera, Zhifeng Chen, Trevor Strohman</i> | |

| | |
|--|------|
| Identifying Coordination in a Cognitive Radar Network - A Multi-Objective Inverse Reinforcement Learning Approach | 8631 |
| <i>Luke Snow, Vikram Krishnamurthy, Brian M. Sadler</i> | |
| Dual-Based Online Learning of Dynamic Network Topologies..... | 8636 |
| <i>Seyed Saman Saboksayr, Gonzalo Mateos</i> | |
| Downlink Covariance Estimation in URA FDD Massive MIMO Systems..... | 8641 |
| <i>Salime Bameri, Khalid Almahrog, Ramy H. Gohary, Amr El-Keyi, Yahia Ahmed</i> | |
| Pyramid Dynamic Inference: Encouraging Faster Inference Via Early Exit Boosting | 8646 |
| <i>Ershad Banijamali, Pegah Kharazmi, Sepehr Eghbali, Jixuan Wang, Clement Chung, Samridhi Choudhary</i> | |
| Improving Automatic Sleep Staging Via Temporal Smoothness Regularization..... | 8651 |
| <i>Huy Phan, Elisabeth Heremans, Oliver Y. Chén, Philipp Koch, Alfred Mertins, Maarten De Vos</i> | |
| Spatially Selective Deep Non-Linear Filters for Speaker Extraction | 8656 |
| <i>Kristina Tesch, Timo Gerkmann</i> | |
| Egocentric Audio-Visual Noise Suppression..... | 8661 |
| <i>Roshan Sharma, Weipeng He, Ju Lin, Egor Lakomkin, Yang Liu, Kaustubh Kalgaonkar</i> | |
| HRTF Field: Unifying Measured HRTF Magnitude Representation with Neural Fields | 8666 |
| <i>You Zhang, Yuxiang Wang, Zhiyao Duan</i> | |
| Toward Privacy-Enhancing Ambulatory-Based Well-Being Monitoring: Investigating User Re-Identification Risk in Multimodal Data..... | 8671 |
| <i>Ravi Pranjali, Ranjana Seshadri, Rakesh Kumar Sanath Kumar Kadaba, Tiantian Feng, Shrikanth S. Narayanan, Theodora Chaspari</i> | |
| A Knowledge-Driven Vowel-Based Approach of Depression Classification from Speech Using Data Augmentation..... | 8676 |
| <i>Kexin Feng, Theodora Chaspari</i> | |
| InfoShape: Task-Based Neural Data Shaping Via Mutual Information..... | 8681 |
| <i>Homa Esfahanizadeh, William Wu, Manya Ghobadi, Regina Barzilay, Muriel Médard</i> | |
| Conversational Text-To-SQL: An Odyssey into State-of-The-Art and Challenges Ahead..... | 8686 |
| <i>Sree Hari Krishnan Parthasarathi, Lu Zeng, Dilek Hakkani-Tür</i> | |
| Interpolation of Spatial Room Impulse Responses Using Partial Optimal Transport..... | 8691 |
| <i>Aaron Geldert, Nils Meyer-Kahlen, Sebastian J. Schlecht</i> | |
| Sign Language Recognition Via Deformable 3D Convolutions and Modulated Graph Convolutional Networks | 8696 |
| <i>Katerina Papadimitriou, Gerasimos Potamianos</i> | |
| Efficient Feature Extraction for Non-Maximum Suppression in Visual Person Detection..... | 8701 |
| <i>Charalampos Symeonidis, Ioannis Mademlis, Ioannis Pitas, Nikos Nikolaidis</i> | |
| Deformable Temporal Convolutional Networks for Monaural Noisy Reverberant Speech Separation | 8706 |
| <i>William Ravenscroft, Stefan Goetze, Thomas Hain</i> | |
| Speech Privacy Leakage from Shared Gradients in Distributed Learning | 8711 |
| <i>Zhuohang Li, Jiaxin Zhang, Jian Liu</i> | |

| | |
|---|------|
| VarietySound: Timbre-Controllable Video to Sound Generation Via Unsupervised Information Disentanglement | 8716 |
| <i>Chenye Cui, Zhou Zhao, Yi Ren, Jinglin Liu, Rongjie Huang, Feiyang Chen, Zhefeng Wang, Baoxing Huai, Fei Wu</i> | |
| Neural Networks with Quantization Constraints | 8721 |
| <i>Ignacio Hounie, Juan Elenter, Alejandro Ribeiro</i> | |
| Wireless Power Transfer Using Chirp Waveforms | 8726 |
| <i>Arijit Roy, Constantinos Psomas, Ioannis Krikidis</i> | |
| Classifying Non-Individual Head-Related Transfer Functions with a Computational Auditory Model: Calibration and Metrics | 8731 |
| <i>Rapolas Daugintis, Roberto Barumerli, Lorenzo Picinali, Michele Geronazzo</i> | |
| Cov Loss: Covariance-Based Loss for Deep Face Recognition | 8736 |
| <i>Ibrahim Alkanhal, Abdullah Almansour, Lamia Alsalloom, Raied Aljadaany, Marios Savvides</i> | |
| Improved Small Sample Hypothesis Testing Using the Uncertain Likelihood Ratio | 8741 |
| <i>James Zachary Hare, Lance M. Kaplan</i> | |
| Machine Learning-Aided Piece-Wise Modeling Technique of Power Amplifier for Digital Predistortion | 8746 |
| <i>S. S. Krishna Chaitanya Bulusu, Nuutti Tervo, Praneeth Susarla, Mikko J. Sillanpää, Olli Silvén, Markku Juntti, Aarno Pärssinen</i> | |
| End-To-End Spoken Language Understanding with Tree-Constrained Pointer Generator | 8751 |
| <i>Guangzhi Sun, Chao Zhang, Philip C. Woodland</i> | |
| Recursive/Iterative Unique Projection-Aggregation Decoding of Reed-Muller Codes | 8756 |
| <i>Marzieh Hashemipour-Nazari, Renate Debets, Kees Goossens, Alexios Balatsoukas-Stimming</i> | |
| WIFI-Based Robust Child Presence Detection for Smart Cars | 8761 |
| <i>Sakila S. Jayaweera, Beibei Wang, Xiaolu Zeng, Wei-Hsiang Wang, K. J. Ray Liu</i> | |
| Exploring Sequence-To-Sequence Transformer-Transducer Models for Keyword Spotting | 8766 |
| <i>Beltrán Labrador, Guanlong Zhao, Ignacio López Moreno, Angelo Scorza Scarpati, Liam Fowl, Quan Wang</i> | |
| Performance Comparison of TTS Models for Brazilian Portuguese to Establish a Baseline | 8771 |
| <i>Wilmer Lobato, Felipe Farias, William Cruz, Marcellus Amadeus</i> | |
| Low-Rank Tensor Decompositions for Quaternion Multiway Arrays | 8776 |
| <i>Osimone Imhogiemhe, Julien Flamant, Xavier Luciani, Yassine Zniyed, Sebastian Miron</i> | |
| Received Power Maximization with Practical Phase-Dependent Amplitude Response in RIS-Aided OFDM Wireless Communications | 8781 |
| <i>Dimitris Kompostiotis, Dimitris Vordonis, Vassilis Paliouras</i> | |
| The 2nd Clarity Enhancement Challenge for Hearing Aid Speech Intelligibility Enhancement: Overview and Outcomes | 8786 |
| <i>Michael A. Akeroyd, Will Bailey, Jon Barker, Trevor J. Cox, John F. Culling, Simone Graetzer, Graham Naylor, Zuzanna Podwinska, Zehai Tu</i> | |
| Internal Language Model Estimation Based Adaptive Language Model Fusion for Domain Adaptation | 8791 |
| <i>Rao Ma, Xiaobo Wu, Jin Qiu, Yanan Qin, Haihua Xu, Peihao Wu, Zejun Ma</i> | |

| | |
|---|------|
| Cochlear Decomposition: A Novel Bio-Inspired Multiscale Analysis Framework | 8796 |
| <i>Hessa Alfalahi, Ahsan Khandoker, Ghada Alhussein, Leontios Hadjileontiadis</i> | |
| Motion Matters: A Novel Motion Modeling for Cross-View Gait Feature Learning | 8801 |
| <i>Jingqi Li, Jiaqi Gao, Yuzhen Zhang, Hongming Shan, Junping Zhang</i> | |
| Last: Scalable Lattice-Based Speech Modelling in Jax | 8806 |
| <i>Ke Wu, Ehsan Variansi, Tom Bagby, Michael Riley</i> | |
| Sparsity Constraint Implementation for the Joint Eigenvalue Decomposition of Matrices..... | 8811 |
| <i>Rémi André, Xavier Luciani</i> | |
| Client Selection for Generalization in Accelerated Federated Learning: A Bandit Approach | 8816 |
| <i>Dan Ben Ami, Kobi Cohen, Qing Zhao</i> | |
| Multilingual End-To-End Spoken Language Understanding for Ultra-Low Footprint Applications | 8821 |
| <i>Markus Müller, Anastasios Alexandridis, Zach Trozenski, Joel Whiteman, Grant Strimel, Nathan Susanj, Athanasios Mouchtaris, Siegfried Kunzmann</i> | |
| Robust Monocular Localization of Drones by Adapting Domain Maps to Depth Prediction Inaccuracies | 8826 |
| <i>Priyesh Shukla, S. Sureshkumar, Alex C. Stutts, Sathya Ravi, Theja Tulabandhula, Amit R. Trivedi</i> | |
| Acoustic Source Localization in the Spherical Harmonics Domain Exploiting Low-Rank Approximations | 8831 |
| <i>Maximo Cobos, Mirco Pezzoli, Fabio Antonacci, Augusto Sarti</i> | |
| Investigating Content-Aware Neural Text-to-Speech MOS Prediction Using Prosodic and Linguistic Features | 8836 |
| <i>Alexandra Vioni, Georgia Maniati, Nikolaos Ellinas, June Sig Sung, Inchul Hwang, Aimilios Chalamandaris, Pirros Tsiakoulis</i> | |
| Joint Angle and Respiration Estimation for Passive and Device-Free Respiration Monitoring | 8841 |
| <i>Gerrit Maus, Dieter Brückmann</i> | |
| The Role of Initial Entanglement in Adaptive Gibbs State Preparation on Quantum Computers | 8846 |
| <i>Sophia E. Economou, Ada Warren, Edwin Barnes</i> | |
| Learning Dependencies of Discrete Speech Representations with Neural Hidden Markov Models | 8851 |
| <i>Sung-Lin Yeh, Hao Tang</i> | |
| Supervised Contrastive Learning as Multi-Objective Optimization for Fine-Tuning Large Pre-Trained Language Models | 8856 |
| <i>Youness Moukafih, Mounir Ghogho, Kamel Smaili</i> | |
| Unsupervised Domain Adaptation for Preference Learning Based Speech Emotion Recognition..... | 8861 |
| <i>Abinay Reddy Naini, Mary A. Kohler, Carlos Busso</i> | |
| An Augmented Gaussian Sum Filter Through a Mixture Decomposition | 8866 |
| <i>Kostas Tsampourakis, Víctor Elvira</i> | |
| EfficientSpeech: An On-Device Text to Speech Model..... | 8871 |
| <i>Rowel Atienza</i> | |
| Contrastive Self-Supervised Learning for Automated Multi-Modal Dance Performance Assessment | 8876 |
| <i>Yun Zhong, Fan Zhang, Yiannis Demiris</i> | |

| | |
|--|------|
| SG-VAD: Stochastic Gates Based Speech Activity Detection..... | 8881 |
| <i>Jonathan Svirsky, Ofir Lindenbaum</i> | |
| Automatic Error Detection in Integrated Circuits Image Segmentation: A Data-Driven Approach | 8886 |
| <i>Zhikang Zhang, Bruno Machado Trindade, Michael Green, Zifan Yu, Christopher Pawlowicz, Fengbo Ren</i> | |
| End-To-End Spoken Language Understanding Using Joint CTC Loss and Self-Supervised, Pretrained Acoustic Encoders..... | 8891 |
| <i>Jixuan Wang, Martin Radfar, Kai Wei, Clement Chung</i> | |
| Anomalous Signal Detection for Cyber-Physical Systems Using Interpretable Causal Neural Network..... | 8896 |
| <i>Shuo Zhang, Jing Liu</i> | |
| On the Primal and Dual Formulations of the Discrete Mumford-Shah Functional | 8901 |
| <i>Nelly Pustelnik</i> | |
| Abstract Representation for Multi-Intent Spoken Language Understanding..... | 8906 |
| <i>Rim Abrougui, Géraldine Damnati, Johannes Heinecke, Frédéric Béchet</i> | |
| CSM in Motion Vector Steganalysis: The Effect of Coders on Motion Vectors in H.264 Video Encoding..... | 8911 |
| <i>Verena Lachner, Katharina Schaar, Ralf Zimmermann</i> | |
| Distributed Gaussian Process Hyperparameter Optimization for Multi-Agent Systems | 8916 |
| <i>Peiyuan Zhai, Raj Thilak Rajan</i> | |
| Regularized EM Algorithm | 8921 |
| <i>Pierre Houdouin, Esa Ollila, Frédéric Pascal</i> | |
| Waveform Boundary Detection for Partially Spoofed Audio | 8926 |
| <i>Zexin Cai, Weiqing Wang, Ming Li</i> | |
| Simultaneous Acoustic Echo Sorting and 3-D Room Geometry Inference | 8931 |
| <i>Kathleen Macwilliam, Filip Elvander, Toon Van Waterschoot</i> | |
| Simultaneously Learning Robust Audio Embeddings and Balanced Hash Codes for Query-By-Example..... | 8936 |
| <i>Anup Singh, Kris Demuynck, Vipul Arora</i> | |
| Row Conditional-TGAN for Generating Synthetic Relational Databases..... | 8941 |
| <i>Mohamed Gueye, Yazid Attabi, Maxime Dumas</i> | |
| Distill-Quantize-Tune - Leveraging Large Teachers for Low-Footprint Efficient Multilingual NLU on Edge..... | 8946 |
| <i>Pegah Kharazmi, Zhewei Zhao, Clement Chung, Samridhi Choudhary</i> | |
| Comprehensive Complexity Assessment of Emerging Learned Image Compression on CPU and GPU | 8951 |
| <i>Farhad Pakdaman, Moncef Gabbouj</i> | |
| Convolutional Filtering on Sampled Manifolds | 8956 |
| <i>Zhiyang Wang, Luana Ruiz, Alejandro Ribeiro</i> | |
| Detail-Aware Uncalibrated Photometric Stereo..... | 8961 |
| <i>Antonio Agudo</i> | |

| | |
|--|------|
| Optimal Condition Training for Target Source Separation | 8966 |
| <i>Efthymios Tzinis, Gordon Wichern, Paris Smaragdis, Jonathan Le Roux</i> | |
| Torchaudio-Squim: Reference-Less Speech Quality and Intelligibility Measures in Torchaudio | 8971 |
| <i>Anurag Kumar, Ke Tan, Zhaoheng Ni, Pranay Manocha, Xiaohui Zhang, Ethan Henderson, Buye Xu</i> | |
| PRIME: 3D Human Pose and Body Shape Recovery with Perspective Projection..... | 8976 |
| <i>Baobei Xu, Shukai Fang, Zhaoyang Li, Shicai Yang, Di Xie, Shiliang Pu</i> | |
| Spice+: Evaluation of Automatic Audio Captioning Systems with Pre-Trained Language Models..... | 8981 |
| <i>Félix Gontier, Romain Serizel, Christophe Cerisara</i> | |
| Tangent Bundle Filters and Neural Networks: from Manifolds to Cellular Sheaves and Back..... | 8986 |
| <i>C. Battiloro, Z. Wang, H. Riess, P. Di Lorenzo, A. Ribeiro</i> | |
| Neural-AFC: Learning-Based Step-Size Control for Adaptive Feedback Cancellation with Closed-Loop Model Training..... | 8991 |
| <i>Behrad Soleimani, Henning Schepker, Majid Mirbagheri</i> | |
| Continuous Descriptor-Based Control for Deep Audio Synthesis..... | 8996 |
| <i>Ninon Devis, Nils Demerlé, Sarah Nabi, David Genova, Philippe Esling</i> | |
| Native Multi-Band Audio Coding Within Hyper-Autoencoded Reconstruction Propagation Networks | 9001 |
| <i>Darius Petermann, Inseon Jang, Minje Kim</i> | |
| Contrastive Learning-Based Audio to Lyrics Alignment for Multiple Languages | 9006 |
| <i>Simon Durand, Daniel Stoller, Sebastian Ewert</i> | |
| Improved Projection Learning for Lower Dimensional Feature Maps..... | 9011 |
| <i>Ilan Price, Jared Tanner</i> | |
| Counterfactual Explanation for Multivariate Times Series Using a Contrastive Variational Autoencoder | 9016 |
| <i>William Todo, Merwann Selmani, Béatrice Laurent, Jean-Michel Loubes</i> | |
| Gaitmixer: Skeleton-Based Gait Representation Learning Via Wide-Spectrum Multi-Axial Mixer..... | 9021 |
| <i>Ekkasit Pinyoanuntapong, Ayman Ali, Pu Wang, Minwoo Lee, Chen Chen</i> | |
| Applying Independent Vector Analysis on EEG-Based Motor Imagery Classification..... | 9026 |
| <i>Caroline P. A. Moraes, Bruno Aristimunha, Lucas Heck Dos Santos, Walter Hugo Lopez Pinaya, Raphael Yokoingawa De Camargo, Denis G. Fantinato, Aline Neves</i> | |
| Prosody is Not Identity: A Speaker Anonymization Approach Using Prosody Cloning | 9031 |
| <i>Sarina Meyer, Florian Lux, Julia Koch, Pavel Denisov, Pascal Tilli, Ngoc Thang Vu</i> | |
| ViT-Cat: Parallel Vision Transformers with Cross Attention Fusion for Popularity Prediction in MEC Networks..... | 9036 |
| <i>Zohreh Hajiakhondi-Meybodi, Arash Mohammadi, Ming Hou, Jamshid Abouei, Konstantinos N. Plataniotis</i> | |
| BreathIE: Estimating Breathing Inhale Exhale Ratio Using Motion Sensor Data from Consumer Earbuds..... | 9041 |
| <i>Nafiu Rashid, Md Mahbubur Rahman, Tousif Ahmed, Jilong Kuang, Jun Alex Gao</i> | |

| | |
|--|------|
| A Point is a Wave: Point-Wave Network for Place Recognition | 9046 |
| <i>Ge Li, Ruonan Zhang</i> | |
| Exploiting Speaker Embeddings for Improved Microphone Clustering and Speech Separation in Ad-Hoc Microphone Arrays | 9051 |
| <i>Stijn Kindt, Jenthe Thienpondt, Nilesh Madhu</i> | |
| Neural Architecture Search with Multimodal Fusion Methods for Diagnosing Dementia | 9056 |
| <i>Michail Chatzianastasis, Loukas Ilias, Dimitris Askounis, Michalis Vazirgiannis</i> | |
| Explainable Audio Classification of Playing Techniques with Layer-Wise Relevance Propagation..... | 9061 |
| <i>Changhong Wang, Vincent Lostanlen, Mathieu Lagrange</i> | |
| An Edge Alignment-Based Orientation Selection Method for Neutron Tomography | 9066 |
| <i>Diyu Yang, Shimin Tang, Singanallur V. Venkatakrishnan, Mohammad S. N. Chowdhury, Yuxuan Zhang, Hassina Z. Bilheux, Gregory T. Buzzard, Charles A. Bouman</i> | |
| Enhance Transferability of Adversarial Examples with Model Architecture..... | 9071 |
| <i>Mingyuan Fan, Wenzhong Guo, Zuobin Ying, Ximeng Liu</i> | |
| Beyond Neural-On-Neural Approaches to Speaker Gender Protection..... | 9076 |
| <i>Loes Van Bommel, Zhuoran Liu, Nik Vaessen, Martha Larson</i> | |
| Masked Autoencoders Are Articulatory Learners..... | 9081 |
| <i>Ahmed Adel Attia, Carol Y. Espy-Wilson</i> | |
| Super Dilated Nested Arrays with Ideal Critical Weights and Increased Degrees of Freedom | 9086 |
| <i>Ahmed M. A. Shaalan, Jun Du</i> | |
| Multimodal Facial Action Unit Detection with Physiological Signals | 9091 |
| <i>Zhihua Li, Lijun Yin</i> | |
| Estimating and Analyzing Neural Information Flow Using Signal Processing on Graphs..... | 9096 |
| <i>Felix Schwock, Julien Bloch, Les Atlas, Shima Abadi, Azadeh Yazdan-Shahmorad</i> | |
| Matrix Low-Rank Approximation for Policy Gradient Methods..... | 9101 |
| <i>Sergio Rozada, Antonio G. Marques</i> | |
| Soft 2D-To-3D Delivery Using Deep Graph Neural Networks for Holographic-Type Communication | 9106 |
| <i>Takuya Fujihashi, Toshiaki Koike-Akino, Takashi Watanabe</i> | |
| Exploring Subgroup Performance in End-To-End Speech Models | 9111 |
| <i>Alkis Koudounas, Eliana Pastor, Giuseppe Attanasio, Vittorio Mazzia, Manuel Giollo, Thomas Gueudre, Luca Cagliero, Luca De Alfaro, Elena Baralis, Daniele Amberti</i> | |
| Perceptual-Neural-Physical Sound Matching..... | 9116 |
| <i>Han Han, Vincent Lostanlen, Mathieu Lagrange</i> | |
| An Implicit Gradient Method for Constrained Bilevel Problems Using Barrier Approximation | 9121 |
| <i>Ioannis Tsaknakis, Prashant Khanduri, Mingyi Hong</i> | |
| Benchmarking Convolutional Neural Network Inference on Low-Power Edge Devices | 9126 |
| <i>Oscar Ferraz, Helder Araujo, Vitor Silva, Gabriel Falcao</i> | |
| Signal Analysis-Synthesis Using the Quantum Fourier Transform | 9131 |
| <i>Aradhita Sharma, Glen Uehara, Vivek Narayanaswamy, Leslie Miller, Andreas Spanias</i> | |

| | |
|---|------|
| Blind Acoustic Room Parameter Estimation Using Phase Features | 9136 |
| <i>Christopher Ick, Adib Mehrabi, Wenyu Jin</i> | |
| Exploring Wav2vec 2.0 Fine Tuning for Improved Speech Emotion Recognition..... | 9141 |
| <i>Li-Wei Chen, Alexander Rudnicky</i> | |
| BER-Aware Dynamic Resource Management for Edge-Assisted Goal-Oriented Communications..... | 9146 |
| <i>Francesco Binucci, Paolo Banelli</i> | |
| Adaptable End-To-End ASR Models Using Replaceable Internal LMs and Residual Softmax..... | 9151 |
| <i>Keqi Deng, Philip C. Woodland</i> | |
| Co-Operative CNN for Visual Saliency Prediction on WCE Images | 9156 |
| <i>George Dimas, Anastasios Koulaouzidis, Dimitris K. Iakovidis</i> | |
| Outside Knowledge Visual Question Answering Version 2.0..... | 9161 |
| <i>Benjamin Z. Reichman, Anirudh Sundar, Christopher Richardson, Tamara Zubatiy, Prithwjit Chowdhury, Aaryan Shah, Jack Truxal, Micah Grimes, Dristi Shah, Woo Ju Chee, Saif Punjwani, Atishay Jain, Larry Heck</i> | |
| Learning Environmental Structure Using Acoustic Probes with a Deep Neural Network | 9166 |
| <i>Toros Arikan, Amir Weiss, Hari Vishnu, Grant Deane, Andrew Singer, Gregory Wornell</i> | |
| Single-Shot Domain Adaptation Via Target-Aware Generative Augmentations..... | 9171 |
| <i>Rakshith Subramanyam, Kowshik Thopalli, Spring Berman, Pavan Turaga, Jayaraman J. Thiagarajan</i> | |
| A Parallel Attention Mechanism for Image Manipulation Detection and Localization..... | 9176 |
| <i>Qiang Zeng, Hongxia Wang, Yang Zhou, Rui Zhang, Sijiang Meng</i> | |
| Speech MOS Multi-Task Learning and Rater Bias Correction..... | 9181 |
| <i>Haleh Akrami, Hannes Gamper</i> | |
| Improving Audio Captioning Using Semantic Similarity Metrics..... | 9186 |
| <i>Rehana Mahfuz, Yinyi Guo, Erik Visser</i> | |
| Glacier: Glass-Box Transformer for Interpretable Dynamic Neuroimaging | 9191 |
| <i>Usman Mahmood, Zening Fu, Vince Calhoun, Sergey Plis</i> | |
| Efficient Speech Translation with Dynamic Latent Perceivers | 9196 |
| <i>Ioannis Tsiamas, Gerard I. Gállego, José A. R. Fonollosa, Marta R. Costa-Jussà</i> | |
| Early Detection of Cognitive Decline Using Voice Assistant Commands..... | 9201 |
| <i>Eli Kurtz, Youxiang Zhu, Tiffany Driesse, Bang Tran, John A. Batsis, Robert M. Roth, Xiaohui Liang</i> | |
| Boosting the Accuracy of SRAM-Based in-Memory Architectures Via Maximum Likelihood-Based Error Compensation Method | 9206 |
| <i>Hyungyo Kim, Naresh Shanbhag</i> | |
| Using Machine Learning to Understand the Relationships Between Audiometric Data, Speech Perception, Temporal Processing, and Cognition | 9211 |
| <i>Rana M. Khalil, Alexandra Papanicolaou, Renee Ti Chou, Bobby E. Gibbs, Samira Anderson, Sandra Gordon-Salant, Michael P. Cummings, Matthew J. Goupell</i> | |

| | |
|---|------|
| M-SpeechCLIP: Leveraging Large-Scale, Pre-Trained Models for Multilingual Speech to Image Retrieval | 9216 |
| <i>Layne Berry, Yi-Jen Shih, Hsuan-Fu Wang, Heng-Jui Chang, Hung-Yi Lee, David Harwath</i> | |
| Coupled CP Tensor Decomposition with Shared and Distinct Components for Multi-Task Fmri Data Fusion..... | 9221 |
| <i>R. A. Borsoi, I. Lehmann, M. A. B. S. Akhonda, V. D. Calhoun, K. Usevich, D. Brie, T. Adali</i> | |
| ASSD: Synthetic Speech Detection in the AAC Compressed Domain | 9226 |
| <i>Amit Kumar Singh Yadav, Ziyue Xiang, Emily R. Bartusiak, Paolo Bestagini, Stefano Tubaro, Edward J. Delp</i> | |
| Robust Self-Guided Deep Image Prior | 9231 |
| <i>Evan Bell, Shijun Liang, Qing Qu, Saiprasad Ravishankar</i> | |
| Fully Distributed Federated Learning with Efficient Local Cooperations..... | 9236 |
| <i>Evangelos Georgatos, Christos Mavrokefalidis, Kostas Berberidis</i> | |
| HYDRA-HGR: A Hybrid Transformer-Based Architecture for Fusion of Macroscopic and Microscopic Neural Drive Information | 9241 |
| <i>Mansoorah Montazerin, Elahe Rahimian, Farnoosh Naderkhani, S. Farokh Atashzar, Hamid Alinejad-Rokny, Arash Mohammadi</i> | |
| AERO: Audio Super Resolution in the Spectral Domain | 9246 |
| <i>Moshe Mandel, Or Tal, Yossi Adi</i> | |
| On the Reduction of Large-Scale Room Acoustic Models | 9251 |
| <i>Pavlos Stoikos, Olympia Axelou, George Floros, Nestor Evmorfopoulos, George Stamoulis</i> | |
| JPEG Pleno Call for Proposals Responses Quality Assessment | 9256 |
| <i>Joao Prazeres, Zhe Luo, Antonio M. G. Pinheiro, Luis A. Da Silva Cruz, Stuart Perry</i> | |
| Unlimited Sampling Radar: Life Below the Quantization Noise | 9261 |
| <i>Thomas Feuillen, Bhavani Shankar Mrr, Ayush Bhandari</i> | |
| UWB Localization-Of-Things Via Soft Information: Network Experimentation in Indoor Environment | 9266 |
| <i>Carlos A. Gómez-Vega, Moe Z. Win, Andrea Conti</i> | |
| Leveraging Label Correlations in a Multi-Label Setting: A Case Study in Emotion..... | 9271 |
| <i>Georgios Chochlakis, Gireesh Mahajan, Sabyasachee Baruah, Keith Burghardt, Kristina Lerman, Shrikanth Narayanan</i> | |
| Analysing the Masked Predictive Coding Training Criterion for Pre-Training a Speech Representation Model..... | 9276 |
| <i>Hemant Yadav, Sunayana Sitaram, Rajiv Ratn Shah</i> | |
| Using Emotion Embeddings to Transfer Knowledge Between Emotions, Languages, and Annotation Formats | 9281 |
| <i>Georgios Chochlakis, Gireesh Mahajan, Sabyasachee Baruah, Keith Burghardt, Kristina Lerman, Shrikanth Narayanan</i> | |
| Jamming Source Localization Using Augmented Physics-Based Model | 9286 |
| <i>Andrea Nardin, Tales Imbiriba, Pau Closas</i> | |
| Audio-Visual Speaker Diarization in the Framework of Multi-User Human-Robot Interaction..... | 9291 |
| <i>Timothée Dhaussy, Bassam Jabaian, Fabrice Lefèvre, Radu Horaud</i> | |

| | |
|--|------|
| Personalized Task Load Prediction in Speech Communication..... | 9296 |
| <i>Robert P. Spang, Karl El Hajal, Sebastian Möller, Milos Cernak</i> | |
| Text-To-ECG: 12-Lead Electrocardiogram Synthesis Conditioned on Clinical Text Reports..... | 9301 |
| <i>Hyunseung Chung, Jiho Kim, Joon-Myoung Kwon, Ki-Hyun Jeon, Min Sung Lee, Edward Choi</i> | |
| Fixed-Point Quantization Aware Training for on-Device Keyword-Spotting | 9306 |
| <i>Sashank Macha, Om Oza, Alex Escott, Francesco Calivá, Robbie Armitano, Santosh Kumar Cheekatmalla, Sree Hari Krishnan Parthasarathi, Yuzong Liu</i> | |
| Model-Free Learning of Optimal Beamformers for Passive IRS-Assisted Sumrate Maximization | 9311 |
| <i>Hassaan Hashmi, Spyridon Pougkakiotis, Dionysios S. Kalogierias</i> | |
| Self-Supervised Speech Representation Learning for Keyword-Spotting with Light-Weight Transformers..... | 9316 |
| <i>Chenyang Gao, Yue Gu, Francesco Caliva, Yuzong Liu</i> | |
| Towards Dialogue Modeling Beyond Text | 9321 |
| <i>Tongzi Wu, Yuhao Zhou, Wang Ling, Hojin Yang, Joana Veloso, Lin Sun, Ruixin Huang, Norberto Guimaraes, Scott Sanner</i> | |
| Variable Rate Allocation for Vector-Quantized Autoencoders..... | 9326 |
| <i>Federico Baldassarre, Alaaeldin El-Nouby, Hervé Jégou</i> | |
| Simultaneous Estimation of Direction of Arrival and Sound Speed Using a Non-Uniform Sensor Array..... | 9331 |
| <i>Ryouichi Nishimura, Kenichi Takizawa</i> | |
| Identifiable Bounded Component Analysis Via Minimum Volume Enclosing Parallelotope..... | 9336 |
| <i>Jingzhou Hu, Kejun Huang</i> | |
| A Statistical Interpretation of the Maximum Subarray Problem..... | 9341 |
| <i>Dennis Wei, Dmitry M. Malioutov</i> | |
| Learning to Personalize Equalization for High-Fidelity Spatial Audio Reproduction..... | 9346 |
| <i>Arjun Gupta, Pablo F. Hoffmann, Sebastian Prepelita, Philip Robinson, Vamsi K. Ithapu, David L. Alon</i> | |
| Deep Learning for Lagrangian Drift Simulation at the Sea Surface | 9351 |
| <i>Daria Botvynko, Carlos Granero-Belinchon, Simon Van Gennip, Abdesslam Benzinou, Ronan Fablet</i> | |
| Machine Learning Based Early Debris Detection Using Automotive Low Level Radar Data | 9356 |
| <i>Kanishka Tyagi, Shan Zhang, Yihang Zhang, John Kirkwood, Sanling Song, Narbik Manukian</i> | |
| Volume-Regularized Nonnegative Tucker Decomposition with Identifiability Guarantees..... | 9361 |
| <i>Yuchen Sun, Kejun Huang</i> | |
| SAN: A Robust End-To-End ASR Model Architecture | 9366 |
| <i>Zeping Min, Qian Ge, Guanhua Huang</i> | |
| Efficient Domain Adaptation for Speech Foundation Models..... | 9371 |
| <i>Bo Li, Dongseong Hwang, Zhouyuan Huo, Junwen Bai, Guru Prakash, Tara N. Sainath, Khe Chai Sim, Yu Zhang, Wei Han, Trevor Strohman, Francoise Beaufays</i> | |

| | |
|---|------|
| Hearttoheart: The Arts of Infant Versus Adult-Directed Speech Classification..... | 9376 |
| <i>Najla D. Al Futaisi, Alejandrina Cristia, Björn W. Schuller</i> | |
| Deep AHS: A Deep Learning Approach to Acoustic Howling Suppression..... | 9381 |
| <i>Hao Zhang, Meng Yu, Dong Yu</i> | |
| Powerful and Extensible WFST Framework for Rnn-Transducer Losses..... | 9386 |
| <i>Aleksandr Laptev, Vladimir Bataev, Igor Gitman, Boris Ginsburg</i> | |
| Speaker-Independent Acoustic-to-Articulatory Speech Inversion..... | 9391 |
| <i>Peter Wu, Li-Wei Chen, Cheol Jun Cho, Shinji Watanabe, Louis Goldstein, Alan W Black, Gopala K. Anumanchipalli</i> | |
| Hadamard Layer to Improve Semantic Segmentation..... | 9396 |
| <i>Angello Hoyos, Mariano Rivera</i> | |
| SMUG: Towards Robust Mri Reconstruction by Smoothed Unrolling..... | 9401 |
| <i>Hui Li, Jinghan Jia, Shijun Liang, Yuguang Yao, Saiprasad Ravishankar, Sijia Liu</i> | |
| Repackagingaugmt: Overcoming Prediction Error Amplification in Weight-Averaged Speech Recognition Models Subject to Self-Training..... | 9406 |
| <i>Jae-Hong Lee, Dong-Hyun Kim, Joon-Hyuk Chang</i> | |
| Joint Antenna Selection and Beamforming in Integrated Automotive Radar Sensing-Communications with Quantized Double Phase Shifters..... | 9411 |
| <i>Lifan Xu, Shunqiao Sun, Yimin D. Zhang, Athina Petropulu</i> | |
| Exploring Attention Mechanisms for Multimodal Emotion Recognition in an Emergency Call Center Corpus..... | 9416 |
| <i>Theo Deschamps-Berger, Lori Lamel, Laurence Devillers</i> | |
| Evaluation of Categorical Generative Models - Bridging the Gap Between Real and Synthetic Data..... | 9421 |
| <i>Florence Regol, Anja Kroon, Mark Coates</i> | |
| Joint Modelling of Spoken Language Understanding Tasks with Integrated Dialog History..... | 9426 |
| <i>Siddhant Arora, Hayato Futami, Emiru Tsunoo, Brian Yan, Shinji Watanabe</i> | |
| Learn Topological Representation with Flexible Manifold Layer..... | 9431 |
| <i>Ziheng Jiao, Hongyuan Zhang, Xuelong Li</i> | |
| Signal Processing with Optical Quadratic Random Sketches..... | 9436 |
| <i>Rémi Delogne, Vincent Schellekens, Laurent Daudet, Laurent Jacques</i> | |
| SemanticAC: Semantics-Assisted Framework for Audio Classification..... | 9441 |
| <i>Yicheng Xiao, Yue Ma, Shuyan Li, Hantao Zhou, Ran Liao, Xiu Li</i> | |
| DWFormer: Dynamic Window Transformer for Speech Emotion Recognition..... | 9446 |
| <i>Shuaiqi Chen, Xiaofen Xing, Weibin Zhang, Weidong Chen, Xiangmin Xu</i> | |
| Unique Bispectrum Inversion for Signals with Finite Spectral/Temporal Support..... | 9451 |
| <i>Samuel Pinilla, Kumar Vijay Mishra, Brian M. Sadler</i> | |
| Online Model Compression for Federated Learning with Large Models..... | 9456 |
| <i>Tien-Ju Yang, Yonghui Xiao, Giovanni Motta, Françoise Beaufays, Rajiv Mathews, Mingqing Chen</i> | |

| | |
|---|------|
| Deep Born Operator Learning for Reflection Tomographic Imaging..... | 9461 |
| <i>Qingqing Zhao, Yanting Ma, Petros Boufounos, Saleh Nabi, Hassan Mansour</i> | |
| An Isotropy Analysis for Self-Supervised Acoustic Unit Embeddings on the Zero Resource Speech Challenge 2021 Framework..... | 9466 |
| <i>Jianan Chen, Sakriani Sakti</i> | |
| An Adaptive Enhancement Method for Gastrointestinal Low-Light Images of Capsule Endoscope..... | 9471 |
| <i>Peixuan Liu, Yinghui Wang, Jinlong Yang, Wei Li</i> | |
| Multi-Channel Audio Signal Generation..... | 9476 |
| <i>W. Bastiaan Kleijn, Michael Chinen, Felicia S. C. Lim, Jan Skoglund</i> | |
| TF-GRIDNET: Making Time-Frequency Domain Models Great Again for Monaural Speaker Separation..... | 9481 |
| <i>Zhong-Qiu Wang, Samuele Cornell, Shukjae Choi, Younglo Lee, Byeong-Yeol Kim, Shinji Watanabe</i> | |
| Meta Learning for Domain Agnostic Soft Prompt..... | 9486 |
| <i>Ming-Yen Chen, Mahdin Rohmatillah, Ching-Hsien Lee, Jen-Tzung Chien</i> | |
| Selinet: A Lightweight Model for Single Channel Speech Separation..... | 9491 |
| <i>Ha Minh Tan, Duc-Quang Vu, Jia-Ching Wang</i> | |
| Graph Wavelet-Based Point Cloud Geometric Denoising with Surface-Consistent Non-Negative Kernel Regression..... | 9496 |
| <i>Ryosuke Watanabe, Keisuke Nonaka, Eduardo Pavez, Tatsuya Kobayashi, Antonio Ortega</i> | |
| The Secret Source : Incorporating Source Features to Improve Acoustic-To-Articulatory Speech Inversion..... | 9501 |
| <i>Yashish M. Siriwardena, Carol Espy-Wilson</i> | |
| Does a Quieter City Mean Fewer Complaints? the Sounds of New York City During Covid-19 Lockdown..... | 9506 |
| <i>Mark Cartwright, Magdalena Fuentes, Charlie Mydlarz, Fabio Miranda, Juan Pablo Bello</i> | |
| WeSinger 2: Fully Parallel Singing Voice Synthesis Via Multi-Singer Conditional Adversarial Training..... | 9511 |
| <i>Zewang Zhang, Yibin Zheng, Xinhui Li, Li Lu</i> | |
| Robustness-Preserving Lifelong Learning Via Dataset Condensation..... | 9516 |
| <i>Jinghan Jia, Yihua Zhang, Dogyoon Song, Sijia Liu, Alfred Hero</i> | |
| Spatio-Temporal Attention in Multi-Granular Brain Chronnectomes for Detection of Autism Spectrum Disorder..... | 9521 |
| <i>James Orme-Rogers, Ajitesh Srivastava</i> | |
| CM-CS: Cross-Modal Common-Specific Feature Learning for Audio-Visual Video Parsing..... | 9526 |
| <i>Hongbo Chen, Dongchen Zhu, Guanghui Zhang, Wenjun Shi, Xiaolin Zhang, Jiamao Li</i> | |
| Active Selection of Source Patients in Transfer Learning for Epileptic Seizure Detection Using Riemannian Manifold..... | 9531 |
| <i>Toshiki Orihara, Kazi Mahmudul Hassan, Toshihisa Tanaka</i> | |
| A Content Adaptive Learnable "Time-Frequency" Representation for Audio Signal Processing..... | 9536 |
| <i>Prateek Verma, Chris Chafe</i> | |

| | |
|--|------|
| GSWIN: Gated MLP Vision Model with Hierarchical Structure of Shifted Window | 9541 |
| <i>Mocho Go, Hideyuki Tachibana</i> | |
| Two-Phase Prototypical Contrastive Domain Generalization for Cross-Subject EEG-Based Emotion Recognition..... | 9546 |
| <i>Honghua Cai, Jiahui Pan</i> | |
| Factorized Projection-Domain Spatio-Temporal Regularization for Dynamic Tomography | 9551 |
| <i>Berk Iskender, Marc L. Klasky, Brian M. Patterson, Yoram Bresler</i> | |
| Mouth Breathing Detection Using Audio Captured Through Earbuds..... | 9556 |
| <i>Tousif Ahmed, Md. Mahbubur Rahman, Ebrahim Nemati, Jilong Kuang, Alex Gao</i> | |
| Deep Implicit Distribution Alignment Networks for cross-Corpus Speech Emotion Recognition | 9561 |
| <i>Yan Zhao, Jincen Wang, Yuan Zong, Wenming Zheng, Hailun Lian, Li Zhao</i> | |
| Discriminative Vector Learning with Application to Single Channel Speech Separation | 9566 |
| <i>Ha Minh Tan, Kai-Wen Liang, Jia-Ching Wang</i> | |
| Self-Supervised Adversarial Training for Contrastive Sentence Embedding | 9571 |
| <i>Jen-Tzung Chien, Yuan-An Chen</i> | |
| Efficient Personalized Federated Learning on Selective Model Training | 9576 |
| <i>Yeting Guo, Fang Liu, Tongqing Zhou, Zhiping Cai, Nong Xiao</i> | |
| Efficient and Effective Multi-Camera Pose Estimation with Weighted M-Estimate Sample Consensus..... | 9581 |
| <i>Xinyu Lin, Yingjie Zhou, Xun Zhang, Yipeng Liu, Ce Zhu</i> | |
| Euro: Espnet Unsupervised ASR Open-Source Toolkit..... | 9586 |
| <i>Dongji Gao, Jiatong Shi, Shun-Po Chuang, Leibny Paola Garcia, Hung-Yi Lee, Shinji Watanabe, Sanjeev Khudanpur</i> | |
| HeMPPCAT: Mixtures of Probabilistic Principal Component Analysers for Data with Heteroscedastic Noise | 9591 |
| <i>Alec S. Xu, Laura Balzano, Jeffrey A. Fessler</i> | |
| Voice-Preserving Zero-Shot Multiple Accent Conversion | 9596 |
| <i>Mumin Jin, Prashant Serai, Jilong Wu, Andros Tjandra, Vimal Manohar, Qing He</i> | |
| Navigating and Reaching Therapeutic Goals with Dynamical Systems in Conversation-Based Interventions..... | 9601 |
| <i>Victor Ardulov, Shrikanth Narayanan</i> | |
| Dictionary Learning on Graph Data with Weisfieler-Lehman Sub-Tree Kernel and Ksvd | 9606 |
| <i>Kaveen Liyanage, Reese Pearsall, Clemente Izurieta, Bradley M. Whitaker</i> | |
| Networked Policy Gradient Play in Markov Potential Games | 9611 |
| <i>Sarper Aydin, Ceyhun Eksin</i> | |
| Level-Line Guided Edge Drawing for Robust Line Segment Detection..... | 9616 |
| <i>Xinyu Lin, Yingjie Zhou, Yipeng Liu, Ce Zhu</i> | |
| Speech-Text Based Multi-Modal Training with Bidirectional Attention for Improved Speech Recognition | 9621 |
| <i>Yuhang Yang, Haihua Xu, Hao Huang, Eng Siong Chng, Sheng Li</i> | |

| | |
|---|------|
| Estimating Acoustic Direction of Arrival Using a Single Structural Sensor on a Resonant Surface..... | 9626 |
| <i>Tre Dipassio, Michael C. Heilemann, Benjamin Thompson, Mark F. Bocko</i> | |
| Comparative Study of IRS Assisted Opportunistic Communications Over I.i.d. and Los Channels..... | 9631 |
| <i>L. Yashvanth, Chandra R. Murthy</i> | |
| Tensor Low Rank Column-Wise Compressive Sensing for Dynamic Imaging..... | 9636 |
| <i>Silpa Babu, Selin Aviyente, Namrata Vaswani</i> | |
| Graphmad: Graph Mixup for Data Augmentation Using Data-Driven Convex Clustering..... | 9641 |
| <i>Madeline Navarro, Santiago Segarra</i> | |
| Neural Feature Predictor and Discriminative Residual Coding for Low-Bitrate Speech Coding..... | 9646 |
| <i>Haici Yang, Wootae Lim, Minje Kim</i> | |
| Towards Bandwidth Estimation for Graph Signal Reconstruction..... | 9651 |
| <i>Ajinkya Jayawant, Antonio Ortega</i> | |
| Deep Spectrum Cartography Using Quantized Measurements..... | 9656 |
| <i>Subash Timilsina, Sagar Shrestha, Xiao Fu</i> | |
| M22: Rate-Distortion Inspired Gradient Compression..... | 9661 |
| <i>Yangyi Liu, Sadaf Salehkalaibar, Stefano Rini, Jun Chen</i> | |
| Joint Training and Decoding for Multilingual End-To-End Simultaneous Speech Translation..... | 9666 |
| <i>Wuwei Huang, Renren Jin, Wen Zhang, Jian Luan, Bin Wang, Deyi Xiong</i> | |
| Robust Network Topologies for Distributed Learning | 9671 |
| <i>Chutian Wang, Stefan Vlaski</i> | |
| TDMA-Based Multi-User Binary Computation Offloading in the Finite-Block-Length Regime..... | 9676 |
| <i>M. Amin Manouchehrpour, Harvinder Lehal, Mahsa Salmani, Timothy N. Davidson</i> | |
| Robustness and Convergence of Mirror Descent for Blind Deconvolution..... | 9681 |
| <i>Ronak Mehta, Sathya N. Ravi, Vikas Singh</i> | |
| Generalized Two-Stage Particle Filter for High Dimensions | 9686 |
| <i>Marija Iloska, Mónica F. Bugallo</i> | |
| A Learnable Spatial Mapping for Decoding the Directional Focus of Auditory Attention Using EEG | 9691 |
| <i>Yuanming Zhang, Haoxin Ruan, Ziyang Yuan, Haoliang Du, Xia Gao, Jing Lu</i> | |
| Multi-Modal Food Classification in a Diet Tracking System with Spoken and Visual Inputs | 9696 |
| <i>Shivani Gowda, Yifan Hu, Mandy Korpusik</i> | |
| A Role Engineering Approach Based on Spectral Clustering Analysis for Restful Permissions in Cloud..... | 9701 |
| <i>Yutang Xia, Yang Luo, Wu Luo, Qingni Shen, Yahui Yang, Zhonghai Wu</i> | |
| Cross-Modal Matching and Adaptive Graph Attention Network for RGB-D Scene Recognition | 9706 |
| <i>Yuhui Guo, Xun Liang, James T. Kwok, Xiangping Zheng, Bo Wu, Yuefeng Ma</i> | |
| Rapid Audiometric Evaluation for Personalized Headphone Listening | 9711 |
| <i>Matthew J. Goupell, Marjan Davoodian, Sarah Weinstein, David Gadzinski, Dmitry N. Zotkin, Kaushik Sethunath, Ramani Duraiswami</i> | |

| | |
|---|------|
| Adaptive Time-Scale Modification for Improving Speech Intelligibility Based on Phoneme Clustering for Streaming Services..... | 9716 |
| <i>Sohee Jang, Jiye Kim, Yeon-Ju Kim, Joon-Hyuk Chang</i> | |
| Unlimited Sampling in Phase Space..... | 9721 |
| <i>Peiyu Zhang, Ayush Bhandari</i> | |
| Towards Efficient and Optimal Joint Beamforming and Antenna Selection: A Machine Learning Approach..... | 9726 |
| <i>Sagar Shrestha, Xiao Fu, Mingyi Hong</i> | |
| Exploring Progressive Hybrid-Degraded Image Processing for Homography Estimation..... | 9731 |
| <i>Yijun Lin, Xingzhe Su, Fengge Wu, Junsuo Zhao</i> | |
| Training Graph Neural Networks on Growing Stochastic Graphs | 9736 |
| <i>Juan Cerviño, Luana Ruiz, Alejandro Ribeiro</i> | |
| A Fast and Accurate Pitch Estimation Algorithm Based on the Pseudo Wigner-Ville Distribution | 9741 |
| <i>Yisi Liu, Peter Wu, Alan W Black, Gopala K. Anumanchipalli</i> | |
| Federated Self-Learning with Weak Supervision for Speech Recognition..... | 9746 |
| <i>Milind Rao, Gopinath Chennupati, Gautam Tiwari, Anit Kumar Sahu, Anirudh Raju, Ariya Rastrow, Jasha Droppo</i> | |
| Freevc: Towards High-Quality Text-Free One-Shot Voice Conversion..... | 9751 |
| <i>Jingyi Li, Weiping Tu, Li Xiao</i> | |
| Managing Information Updating with Edge Computing: A Distributed and Learning Approach | 9756 |
| <i>Junyi He, Di Zhang, Shumeng Liu, Yuezhi Zhou, Yaoxue Zhang</i> | |
| Recursive Estimation of User Intent from Noninvasive Electroencephalography Using Discriminative Models | 9761 |
| <i>Niklas Smedemark-Margulies, Basak Celik, Tales Imbiriba, Aziz Kocanaogullari, Deniz Erdogmus</i> | |
| Multimodal Microscopy Image Alignment Using Spatial and Shape Information and a Branch-And-Bound Algorithm..... | 9766 |
| <i>Shuonan Chen, Bovey Y Rao, Stephanie Herrlinger, Attila Losonczy, Liam Paninski, Erdem Varol</i> | |
| Learning Cross-Lingual Visual Speech Representations | 9771 |
| <i>Andreas Zinonos, Alexandros Haliassos, Pingchuan Ma, Stavros Petridis, Maja Pantic</i> | |
| Rethinking Learning-Based Method for Lossless Genome Compression | 9776 |
| <i>Han Yang, Fei Gu, Jieping Ye</i> | |
| TEFISTA-NET: GTD Parameter Estimation of Low-Frequency Ultra- Wideband Radar Via Model-Based Deep Learning | 9781 |
| <i>Rui Li, Xueqian Wang, Gang Li, Xiao-Ping Zhang</i> | |
| Multi-User Methods for Vibrational Radar Backscatter Communications..... | 9786 |
| <i>Jessica Centers, Jeffrey Krolik</i> | |
| A New Personalized Efficacy Atlas for Pallidal Deep Brain Stimulation | 9791 |
| <i>Xiongbiao Luo</i> | |

| | |
|---|------|
| Modeling the Wave Equation Using Physics-Informed Neural Networks Enhanced with Attention to Loss Weights | 9796 |
| <i>Shaikhah Alkhadhr, Mohamed Almekkawy</i> | |
| Seri: Sketching-Reasoning-Integrating Progressive Workflow for Empathetic Response Generation..... | 9801 |
| <i>Guanqun Bi, Yanan Cao, Piji Li, Yuqiang Xie, Fang Fang, Zheng Lin</i> | |
| On the Minimum Perimeter Criterion for Bounded Component Analysis | 9806 |
| <i>Sergio Cruces</i> | |
| Weighted Sampling for Masked Language Modeling | 9811 |
| <i>Linhan Zhang, Qian Chen, Wen Wang, Chong Deng, Xin Cao, Kongzhang Hao, Yuxin Jiang, Wei Wang</i> | |
| Cold Diffusion for Speech Enhancement | 9816 |
| <i>Hao Yen, François G. Germain, Gordon Wichern, Jonathan Le Roux</i> | |
| Direction Aware Positional and Structural Encoding for Directed Graph Neural Networks | 9821 |
| <i>Yonas Sium, Georgios Kollias, Tsuyoshi Idé, Payel Das, Naoki Abe, Aurélie Lozano, Qi Li</i> | |
| CryoSWD: Sliced Wasserstein Distance Minimization for 3D Reconstruction in Cryo-Electron Microscopy | 9826 |
| <i>Mona Zehni, Zhizhen Zhao</i> | |
| Towards a Unified Conformer Structure: from ASR to ASV Task | 9831 |
| <i>Dexin Liao, Tao Jiang, Feng Wang, Lin Li, Qingyang Hong</i> | |
| Active Learning for Efficient Few-Shot Classification | 9836 |
| <i>Aymane Abdali, Vincent Gripon, Lucas Drumetz, Bartosz Boguslawski</i> | |
| Robust Spatiotemporal Fusion of Satellite Images Via Convex Optimization | 9841 |
| <i>Ryosuke Isono, Kazuki Naganuma, Shunsuke Ono</i> | |
| CLIP4VideoCap: Rethinking Clip for Video Captioning with Multiscale Temporal Fusion and Commonsense Knowledge | 9846 |
| <i>Tanvir Mahmud, Feng Liang, Yaling Qing, Diana Marculescu</i> | |
| Direct Position Determination with One-Bit Signal for Multiple Targets | 9851 |
| <i>Lihua Ni, Di Zhang, Tianyi Xing, Maoyan Ran, Ning Liu, Qun Wan</i> | |
| Graph-Based Point Cloud Color Denoising with 3-Dimensional Patch-Based Similarity | 9856 |
| <i>Ryosuke Watanabe, Keisuke Nonaka, Eduardo Pavez, Tatsuya Kobayashi, Antonio Ortega</i> | |
| Music Mixing Style Transfer: A Contrastive Learning Approach to Disentangle Audio Effects | 9861 |
| <i>Junghyun Koo, Marco A. Martínez-Ramírez, Wei-Hsiang Liao, Stefan Uhlich, Kyogu Lee, Yuki Mitsufuji</i> | |
| FedRPO: Federated Relaxed Pareto Optimization for Acoustic Event Classification..... | 9866 |
| <i>Meng Feng, Chieh-Chi Kao, Qingming Tang, Amit Solomon, Viktor Rozgic, Chao Wang</i> | |
| Global and Nodal Mutual Information Maximization in Heterogeneous Graphs..... | 9871 |
| <i>Costas Mavromatis, George Karypis</i> | |
| Spherical Sector Harmonics Based Soundfield Radial Extrapolation and Robustness Analysis..... | 9876 |
| <i>Hanwen Bi, Thushara D. Abhayapala, Fei Ma, Prasanga N. Samarasinghe</i> | |

| | |
|---|------|
| Mabnet: Master Assistant Buddy Network with Hybrid Learning for Image Retrieval | 9881 |
| <i>Rohit Agarwal, Gyanendra Das, Saksham Aggarwal, Alexander Horsch, Dilip K. Prasad</i> | |
| Boundary Cue Guidance and Contextual Feature Mining for Glass Segmentation..... | 9886 |
| <i>Qiquan Xiao, Yuan Zhang, Xuanya Li, Kai Hu</i> | |
| Dynamic Local and Global Context Exploration for Small Object Detection..... | 9891 |
| <i>Ziji Zhang, Ping Gong, Haotian Sun, Pingping Wu, Xuanyuan Yang</i> | |
| Distributed Quantum Sensing Network with Geographically Constrained Measurement Strategies..... | 9896 |
| <i>Yingkang Cao, Xiaodi Wu</i> | |
| 3D Point Cloud Completion Based on Multi-Scale Degradation | 9901 |
| <i>Jianing Long, Qingmeng Zhu, Hao He, Zhipeng Yu, Qilin Zhang, Zhihong Zhang</i> | |
| Semi-Supervised Graph Ultra-Sparsifier Using Reweighted ℓ_1 Optimization..... | 9906 |
| <i>Jiayu Li, Tianyun Zhang, Shengmin Jin, Reza Zafarani</i> | |
| Contextually-Rich Human Affect Perception Using Multimodal Scene Information | 9911 |
| <i>Digbalay Bose, Rajat Hebbar, Krishna Somandepalli, Shrikanth Narayanan</i> | |
| Knowledge-Aware Few Shot Learning for Event Detection from Short Texts..... | 9916 |
| <i>Jinjin Guo, Zhichao Huang, Guangning Xu, Bowen Zhang, Chaoqun Duan</i> | |
| ERSAM: Neural Architecture Search for Energy-Efficient and Real-Time Social Ambiance Measurement | 9921 |
| <i>Chaojian Li, Wenwan Chen, Jiayi Yuan, Yingyan Celine Lin, Ashutosh Sabharwal</i> | |
| Conditional Conformer: Improving Speaker Modulation for Single and Multi-User Speech Enhancement | 9926 |
| <i>Tom O'Malley, Shaojin Ding, Arun Narayanan, Quan Wang, Rajeev Rikhye, Qiao Liang, Yanzhang He, Ian McGraw</i> | |
| Loss Function Design for DNN-Based Sound Event Localization and Detection on Low-Resource Realistic Data | 9931 |
| <i>Qing Wang, Jun Du, Zhaoxu Nian, Shutong Niu, Li Chai, Huaxin Wu, Jia Pan, Chin-Hui Lee</i> | |
| EMCLR: Expectation Maximization Contrastive Learning Representations..... | 9936 |
| <i>Meng Liu, Ran Yi, Lizhuang Ma</i> | |
| An Online Algorithm for Contrastive Principal Component Analysis | 9941 |
| <i>Siavash Golkar, David Lipshutz, Tiberiu Tesileanu, Dmitri B. Chklovskii</i> | |
| Multi-Object Localization and Irrelevant-Semantic Separation for Nuclei Segmentation in Histopathology Images..... | 9946 |
| <i>Ya Tang, Xiongjun Ye, Xuanya Li, Zhineng Chen</i> | |
| Motion-Aware Video Paragraph Captioning Via Exploring Object-Centered Internal Knowledge | 9951 |
| <i>Yimin Hu, Guorui Yu, Yuejie Zhang, Rui Feng, Tao Zhang, Xuequan Lu, Shang Gao</i> | |
| Recursive Joint Attention for Audio-Visual Fusion in Regression Based Emotion Recognition | 9956 |
| <i>R Gnana Praveen, Eric Granger, Patrick Cardinal</i> | |
| A Reality Check and a Practical Baseline for Semantic Speech Embedding | 9961 |
| <i>Guangyu Chen, Yuanyuan Cao</i> | |

| | |
|---|-------|
| Graph Representation Learning for Stroke Recurrence Prediction | 9966 |
| <i>Nicholas Glaze, Artun Bayer, Xiaoqian Jiang, Sean Savitz, Santiago Segarra</i> | |
| Toward Auto-Evaluation with Confidence-Based Category Relation-Aware Regression | 9971 |
| <i>Jiexin Wang, Jiahao Chen, Bing Su</i> | |
| Fretnet: Continuous-Valued Pitch Contour Streaming for Polyphonic Guitar Tablature Transcription..... | 9976 |
| <i>Frank Cwitkowitz, Toni Hirvonen, Anssi Klapuri</i> | |
| Minimising Distortion for GAN-Based Facial Attribute Manipulation..... | 9981 |
| <i>Mingyu Shao, Li Lu, Ye Ding, Qing Liao</i> | |
| Space-Time Graph Neural Networks with Stochastic Graph Perturbations | 9986 |
| <i>Samar Hadou, Charilaos I. Kanatsoulis, Alejandro Ribeiro</i> | |
| Interpretation of Neural Networks is Susceptible to Universal Adversarial Perturbations | 9991 |
| <i>Haniyeh Ehsani Oskouie, Farzan Farnia</i> | |
| On Designing a 3D Imaging Summer Project for Ontario’s High School Students During Covid-19 Pandemic | 9996 |
| <i>Fengbo Lan, Gene Cheung, Prabhkirat Arora, Deinabo Richard-Koko, Lisa Cole</i> | |
| MUG: A General Meeting Understanding and Generation Benchmark | 10001 |
| <i>Qinglin Zhang, Chong Deng, Jiaqing Liu, Hai Yu, Qian Chen, Wen Wang, Zhijie Yan, Jinglin Liu, Yi Ren, Zhou Zhao</i> | |
| Active Beam Tracking with Reconfigurable Intelligent Surface | 10006 |
| <i>Han Han, Tao Jiang, Wei Yu</i> | |
| Improving Acoustic Echo Cancellation by Mixing Speech Local and Global Features with Transformer | 10011 |
| <i>Yajie Liu, Xinmeng Xu, Weiping Tu, Yuhong Yang, Li Xiao</i> | |
| CNEG-VC: Contrastive Learning Using Hard Negative Example in Non-Parallel Voice Conversion | 10016 |
| <i>Bima Prihasto, Yi-Xing Lin, Phuong Thi Le, Chien-Lin Huang, Jia-Ching Wang</i> | |
| ITER-SIS: Robust Unlimited Sampling Via Iterative Signal Sieving | 10021 |
| <i>Ruiming Guo, Ayush Bhandari</i> | |
| Modeling Turn-Taking in Human-To-Human Spoken Dialogue Datasets Using Self-Supervised Features | 10026 |
| <i>Edmilson Morais, Matheus Damasceno, Hagai Aronowitz, Aharon Satt, Ron Hoory</i> | |
| Higher-Order Link Prediction Via Learnable Maximum Mean Discrepancy | 10031 |
| <i>Georgios V. Karanikolas, Alba Pagès-Zamora, Georgios B. Giannakis</i> | |
| Predicting Multi-Codebook Vector Quantization Indexes for Knowledge Distillation | 10036 |
| <i>Liyong Guo, Xiaoyu Yang, Quandong Wang, Yuxiang Kong, Zengwei Yao, Fan Cui, Fangjun Kuang, Wei Kang, Long Lin, Mingshuang Luo, Piotr Zelasko, Daniel Povey</i> | |
| Bert is Robust! a Case Against Word Substitution-Based Adversarial Attacks | 10041 |
| <i>Jens Hauser, Zhao Meng, Damian Pascual, Roger Wattenhofer</i> | |
| When is Mimo Massive in Radar? | 10046 |
| <i>Jaimin Shah, Martina Cardone, Alex Dytso, Cynthia Rush</i> | |

| | |
|---|-------|
| Adapting Self-Supervised Models to Multi-Talker Speech Recognition Using Speaker Embeddings..... | 10051 |
| <i>Zili Huang, Desh Raj, Paola García, Sanjeev Khudanpur</i> | |
| Exploring the Role of Fricatives in Classifying Healthy Subjects and Patients with Amyotrophic Lateral Sclerosis and Parkinson’s Disease..... | 10056 |
| <i>Tanuka Bhattacharjee, Yamini Belur, Atchayaram Nalini, Ravi Yadav, Prasanta Kumar Ghosh</i> | |
| Multiple Access Computation Offloading for the K-User Case | 10061 |
| <i>Xiaomeng Liu, Christian Schaible, Timothy N. Davidson</i> | |
| Algebraic Convolutional Filters on Lie Group Algebras..... | 10066 |
| <i>Harshat Kumar, Alejandro Parada-Mayorga, Alejandro Ribeiro</i> | |
| Generalized Invariant Matching Property Via Lasso | 10071 |
| <i>Kang Du, Yu Xiang</i> | |
| Exploring Approaches to Multi-Task Automatic Synthesizer Programming..... | 10076 |
| <i>Daniel Faronbi, Iran Roman, Juan Pablo Bello</i> | |
| Adaptive Eecm for Mitigating Smart Jammers | 10081 |
| <i>Shashwat Jain, Kunal Pattanayak, Vikram Krishnamurthy, Christopher Berry</i> | |
| Cross-Utterance ASR Rescoring with Graph-Based Label Propagation | 10086 |
| <i>Srinath Tankasala, Long Chen, Andreas Stolcke, Anirudh Raju, Qianli Deng, Chander Chandak, Aparna Khare, Roland Maas, Venkatesh Ravichandran</i> | |
| Exploiting Prompt Learning with Pre-Trained Language Models for Alzheimer’s Disease Detection | 10091 |
| <i>Yi Wang, Jiajun Deng, Tianzi Wang, Bo Zheng, Shoukang Hu, Xunying Liu, Helen Meng</i> | |
| Representation Learning of Clinical Multivariate Time Series with Random Filter Banks | 10096 |
| <i>Alireza Keshavarzian, Hojjat Salehinejad, Shahrokh Valaee</i> | |
| Identifying Source Speakers for Voice Conversion Based Spoofing Attacks on Speaker Verification Systems..... | 10101 |
| <i>Danwei Cai, Zexin Cai, Ming Li</i> | |
| Enhancement of Text-Predicting Style Token with Generative Adversarial Network for Expressive Speech Synthesis | 10106 |
| <i>Hiroki Kanagawa, Yusuke Ijima</i> | |
| Static and Dynamic Source and Filter Cues for Classification of Amyotrophic Lateral Sclerosis Patients and Healthy Subjects..... | 10111 |
| <i>Tanuka Bhattacharjee, Chowdam Venkata Thirumala Kumar, Yamini Belur, Atchayaram Nalini, Ravi Yadav, Prasanta Kumar Ghosh</i> | |
| Enhancing Spatio-Spectral Regularization by Structure Tensor Modeling for Hyperspectral Image Denoising | 10116 |
| <i>Shingo Takemoto, Shunsuke Ono</i> | |
| Towards Scale Adaptive Underwater Detection Through Refined Pyramid Grid | 10121 |
| <i>Xiaoheng Deng, Lirong Liao, Ping Jiang, Yurong Qian</i> | |
| Fast Robust Principle Component Analysis Using Gauss-Newton Iterations | 10126 |
| <i>William Chettleburgh, Zhishen Huang, Ming Yan</i> | |
| On Negative Sampling for Contrastive Audio-Text Retrieval | 10131 |
| <i>Huang Xie, Okko Räsänen, Tuomas Virtanen</i> | |

| | |
|---|-------|
| Ideal: Improved Dense Local Contrastive Learning for Semi-Supervised Medical Image Segmentation..... | 10136 |
| <i>Hritam Basak, Soumitri Chattopadhyay, Rohit Kundu, Sayan Nag, Rammohan Mallipeddi</i> | |
| Enhancing Ontology Translation Through Cross-Lingual Agreement | 10141 |
| <i>Mingjie Tian, Fausto Giunchiglia, Rui Song, Xing Chen, Hao Xu</i> | |
| Modular Conformer Training for Flexible End-To-End ASR..... | 10146 |
| <i>Kartik Audhkhasi, Brian Farris, Bhuvana Ramabhadran, Pedro J. Moreno</i> | |
| A Speech Representation Anonymization Framework Via Selective Noise Perturbation | 10151 |
| <i>Minh Tran, Mohammad Soleymani</i> | |
| Quantum Transfer Learning Using the Large-Scale Unsupervised Pre-Trained Model Wavlm-Large for Synthetic Speech Detection | 10156 |
| <i>Ruoyu Wang, Jun Du, Tian Gao</i> | |
| Spammer Detection on Short Video Applications: A New Challenge and Baselines | 10161 |
| <i>Muyang Yi, Dong Liang, Rui Wang, Yue Ding, Hongtao Lu</i> | |
| Low Precision Representations for High Dimensional Models..... | 10166 |
| <i>Rajarshi Saha, Mert Pilanci, Andrea J. Goldsmith</i> | |
| Mingling Or Misalignment? Temporal Shift for Speech Emotion Recognition with Pre-Trained Representations | 10171 |
| <i>Siyuan Shen, Feng Liu, Aimin Zhou</i> | |
| JaCappella Corpus: A Japanese a Cappella Vocal Ensemble Corpus | 10176 |
| <i>Tomohiko Nakamura, Shinnosuke Takamichi, Naoko Tanji, Satoru Fukayama, Hiroshi Saruwatari</i> | |
| Quantum Graph Transformers | 10181 |
| <i>Georgios Kollias, Vassilis Kalantzis, Theodoros Salonidis, Shashanka Ubaru</i> | |
| Articulation GAN: Unsupervised Modeling of Articulatory Learning | 10186 |
| <i>Gasper Beguš, Alan Zhou, Peter Wu, Gopala K. Anumanchipalli</i> | |
| Active IRS-Assisted MIMO Channel Estimation and Prediction..... | 10191 |
| <i>Mirza A. Haider, Saidur R. Pavel, Yimin D. Zhang, Elias Aboutanios</i> | |
| Efficient Stuttering Event Detection Using Siamese Networks..... | 10196 |
| <i>Payal Mohapatra, Bashima Islam, Md Tamzeed Islam, Ruochen Jiao, Qi Zhu</i> | |
| TOPO-MLP : A Simplicial Network Without Message Passing..... | 10201 |
| <i>Karthikeyan Natesan Ramamurthy, Aldo Guzmán-Sáenz, Mustafa Hajj</i> | |
| Improving Transformer-Based End-to-End Speaker Diarization by Assigning Auxiliary Losses to Attention Heads | 10206 |
| <i>Ye-Rin Jeoung, Joon-Young Yang, Jeong-Hwan Choi, Joon-Hyuk Chang</i> | |
| Improving Fairness and Robustness in End-To-End Speech Recognition Through Unsupervised Clustering | 10211 |
| <i>Irina-Elena Veliche, Pascale Fung</i> | |
| Progressive Multi-Stage Neural Audio Codec with Psychoacoustic Loss and Discriminator | 10216 |
| <i>Byeong Hyeon Kim, Hyungseob Lim, Jihyun Lee, Inseon Jang, Hong-Goo Kang</i> | |

| | |
|---|-------|
| Medleyvox: An Evaluation Dataset for Multiple Singing Voices Separation | 10221 |
| <i>Chang-Bin Jeon, Hyeonggi Moon, Keunwoo Choi, Ben Sangbae Chon, Kyogu Lee</i> | |
| On Designing Light-Weight Object Trackers Through Network Pruning: Use CNNs Or Transformers? | 10226 |
| <i>Saksham Aggarwal, Taneesh Gupta, Pawan K. Sahu, Arnav Chavan, Rishabh Tiwari, Dilip K. Prasad, Deepak K. Gupta</i> | |
| Improving Non-Autoregressive Speech Recognition with Autoregressive Pretraining..... | 10231 |
| <i>Yanjia Li, Lahiru Samarakoon, Ivan Fung</i> | |
| Position-Aware Graph-Based Learning of Whole Slide Images..... | 10236 |
| <i>Milan Aryal, Nasim Yahya Soltani</i> | |
| DVQVC: An Unsupervised Zero-Shot Voice Conversion Framework..... | 10241 |
| <i>Dayong Li, Xian Li, Xiaofei Li</i> | |
| Enhancing the Accuracy of Resistive In-Memory Architectures Using Adaptive Signal Processing..... | 10246 |
| <i>Han-Mo Ou, Naresh R. Shanbhag</i> | |
| Learning to Detect Novel and Fine-Grained Acoustic Sequences Using Pretrained Audio Representations | 10251 |
| <i>Vasudha Kowtha, Miquel Espi Marques, Jonathan Huang, Yichi Zhang, Carlos Avendano</i> | |
| MODEFORMER: Modality-Preserving Embedding for Audio-Video Synchronization Using Transformers..... | 10256 |
| <i>Akash Gupta, Rohun Tripathi, Wondong Jang</i> | |
| Compensatory Debiasing for Gender Imbalances in Language Models..... | 10261 |
| <i>Tae-Jin Woo, Woo-Jeoung Nam, Yeong-Joon Ju, Seong-Whan Lee</i> | |
| Federated Semi-Supervised Learning for Object Detection in Autonomous Driving | 10266 |
| <i>Fangyuan Chi, Yixiao Wang, Panos Nasiopoulos, Victor C. M. Leung, Mahsa T. Pourazad</i> | |
| SAMO: Speaker Attractor Multi-Center One-Class Learning for Voice Anti-Spoofing..... | 10271 |
| <i>Siwen Ding, You Zhang, Zhiyao Duan</i> | |
| Choice Fusion as Knowledge for Zero-Shot Dialogue State Tracking..... | 10276 |
| <i>Ruolin Su, Jingfeng Yang, Ting-Wei Wu, Biing-Hwang Juang</i> | |
| Text Classification in the Wild: A Large-Scale Long-Tailed Name Normalization Dataset..... | 10281 |
| <i>Jiexing Qi, Shuhao Li, Zhixin Guo, Yusheng Huang, Chenghu Zhou, Weinan Zhang, Xinbing Wang, Zhouhan Lin</i> | |
| Reverberation as Supervision for Speech Separation | 10286 |
| <i>Rohith Aralikatti, Christoph Boeddeker, Gordon Wichern, Aswin Subramanian, Jonathan Le Roux</i> | |
| A Closer Look at Scoring Functions and Generalization Prediction | 10291 |
| <i>Puja Trivedi, Danai Koutra, Jayaraman J. Thiagarajan</i> | |
| Accelerated Distributed Stochastic Non-Convex Optimization Over Time-Varying Directed Networks | 10296 |
| <i>Yiyue Chen, Abolfazl Hashemi, Haris Vikalo</i> | |

| | |
|---|-------|
| KG-ECO: Knowledge Graph Enhanced Entity Correction for Query Rewriting..... | 10301 |
| <i>Jinglun Cai, Mingda Li, Ziyang Jiang, Eunah Cho, Zheng Chen, Yang Liu, Xing Fan, Chenlei Guo</i> | |
| UML: A Universal Monolingual Output Layer for Multilingual Asr | 10306 |
| <i>Chao Zhang, Bo Li, Tara N. Sainath, Trevor Strohman, Shuo-Yiin Chang</i> | |
| A Phoneme-Informed Neural Network Model for Note-Level Singing Transcription | 10311 |
| <i>Sangeon Yong, Li Su, Juhan Nam</i> | |
| Sub-Band Contrastive Learning-Based Knowledge Distillation for Sound Classification..... | 10316 |
| <i>Achyut Mani Tripathi, Aakansha Mishra</i> | |
| Improving Music Genre Classification from Multi-Modal Properties of Music and Genre Correlations Perspective | 10321 |
| <i>Ganghui Ru, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao</i> | |
| Importance of Different Temporal Modulations of Speech: A Tale of Two Perspectives..... | 10326 |
| <i>Samik Sadhu, Hynek Hermansky</i> | |
| A Progressive Image Dehazing Framework with Inter and Intra Contrastive Learning..... | 10331 |
| <i>Honglei Xu, Shaohui Liu, Yan Shu, Feng Jiang</i> | |
| Waveform Design to Improve the Estimation of Target Parameters Using the Fourier Transform Method in a MIMO OFDM DFRC System..... | 10336 |
| <i>Satwika Bhogavalli, Eric Grivel, K. V. S. Hari, Vincent Corretja</i> | |
| FNeural Speech Enhancement with Very Low Algorithmic Latency and Complexity Via Integrated Full- And Sub-band Modeling..... | 10341 |
| <i>Zhong-Qiu Wang, Samuele Cornell, Shukjae Choi, Younglo Lee, Byeong-Yeol Kim, Shinji Watanabe</i> | |
| A Gaussian Latent Variable Model for Incomplete Mixed Type Data..... | 10346 |
| <i>Marzieh Ajirak, Petar M. Djuric</i> | |
| DMFormer: Closing the Gap Between CNN and Vision Transformers | 10351 |
| <i>Zimian Wei, Hengyue Pan, Lujun Li, Menglong Lu, Xin Niu, Peijie Dong, Dongsheng Li</i> | |
| A Method of Constructing and Automatically Labeling Radio Frequency Signal Training Dataset for UAV | 10356 |
| <i>Chao Liu, Ruipeng Ma, Zheng Si, Mingmin Chi</i> | |
| Improving Fast-Slow Encoder Based Transducer with Streaming Deliberation | 10361 |
| <i>Ke Li, Jay Mahadeokar, Jinxi Guo, Yangyang Shi, Gil Keren, Ozlem Kalinli, Michael L. Seltzer, Duc Le</i> | |
| Dynamic Alignment Mask CTC: Improved Mask CTC with Aligned Cross Entropy..... | 10366 |
| <i>Xulong Zhang, Haobin Tang, Jianzong Wang, Ning Cheng, Jian Luo, Jing Xiao</i> | |
| Online Caching with Fetching Cost for Arbitrary Demand Pattern: A Drift-Plus-Penalty Approach..... | 10371 |
| <i>Shashank P, B. N. Bharath</i> | |
| Improving EEG-Based Emotion Recognition by Fusing Time-Frequency and Spatial Representations | 10376 |
| <i>Kexin Zhu, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao</i> | |

| | |
|--|-------|
| Avoid Overthinking in Self-Supervised Models for Speech Recognition | 10381 |
| <i>Dan Berrebbi, Brian Yan, Shinji Watanabe</i> | |
| Code-Switching Text Generation and Injection in Mandarin-English ASR | 10386 |
| <i>Haibin Yu, Yuxuan Hu, Yao Qian, Ma Jin, Linquan Liu, Shujie Liu, Yu Shi, Yanmin Qian, Edward Lin, Michael Zeng</i> | |
| On the Effectiveness of Monoaural Target Source Extraction for Distant End-To-end Automatic Speech Recognition | 10391 |
| <i>Catalin Zorila, Rama Doddipatla</i> | |
| Channel Estimation with Tightly-Coupled Antenna Arrays | 10396 |
| <i>Bamelak Tadele, Volodymyr Shyianov, Faouzi Bellili, Amine Mezghani</i> | |
| AutoGCF: Personalized Aggregation on Neural Graph Collaborative Filtering | 10401 |
| <i>Xiaoyu You, Chi Li, Jianwei Xu, Mi Zhang</i> | |
| Dual Collaborative Visual-Semantic Mapping for Multi-Label Zero-Shot Image Recognition..... | 10406 |
| <i>Yunqing Hu, Xuan Jin, Xi Chen, Yin Zhang</i> | |
| Joint Estimation of DOA and Distance in Noisy Reverberant Conditions | 10411 |
| <i>Sulian Bu, Tuo Zhao, Yunxin Zhao</i> | |
| Communication-Constrained Exchange of Zeroth-Order Information with Application to Collaborative Target Tracking | 10416 |
| <i>Ege Can Kaya, Mehmet Berk Sahin, Abolfazl Hashemi</i> | |
| On the Robustness of Non-Intrusive Speech Quality Model by Adversarial Examples..... | 10421 |
| <i>Hsin-Yi Lin, Huan-Hsin Tseng, Yu Tsao</i> | |
| More Speaking Or More Speakers?..... | 10426 |
| <i>Dan Berrebbi, Ronan Collobert, Navdeep Jaitly, Tatiana Likhomanenko</i> | |
| Filler Word Detection with Hard Category Mining and Inter-Category Focal Loss..... | 10431 |
| <i>Zhiyuan Zhao, Lijun Wu, Chuanxin Tang, Dacheng Yin, Yucheng Zhao, Chong Luo</i> | |
| Unsupervised Video Anomaly Detection for Stereotypical Behaviours in Autism | 10436 |
| <i>Jiaqi Gao, Xinyang Jiang, Yuqing Yang, Dongsheng Li, Lili Qiu</i> | |
| HPFTN: Hierarchical Progressive Fusion Transformer Network for Video Denoising | 10441 |
| <i>Shuaitao Zhang, Yuan Zhang, Zheng Zhao, Di Xie, Shiliang Pu</i> | |
| Dasformer: Deep Alternating Spectrogram Transformer for Multi/Single-Channel Speech Separation..... | 10446 |
| <i>Shuo Wang, Xiangyu Kong, Xiulian Peng, Hesam Movassagh, Vinod Prakash, Yan Lu</i> | |
| Real-Time Human Reconstruction Based on Human Pose Prior and Epipolar Refinement..... | 10451 |
| <i>Kuncheng Luo, Zhiheng Li</i> | |
| Coded Matrix Computations for D2D-Enabled Linearized Federated Learning..... | 10456 |
| <i>Anindya Bijoy Das, Aditya Ramamoorthy, David J. Love, Christopher G. Brinton</i> | |
| NAS-DYMC: NAS-Based Dynamic Multi-Scale Convolutional Neural Network for Sound Event Detection | 10461 |
| <i>Jun Wang, Peng Yao, Feng Deng, Jianchao Tan, Chengru Song, Xiaorui Wang</i> | |

| | |
|---|-------|
| Signal Processing and Quantum State Tomography on Noisy Devices..... | 10466 |
| <i>Wenbo Shi, Robert Malaney</i> | |
| Class-Aware Contextual Information for Semantic Segmentation | 10471 |
| <i>Huadong Tang, Youpeng Zhao, Yingying Jiang, Zhuoxin Gan, Qiang Wu</i> | |
| Dual-Stage Graph Convolution Network with Graph Learning for Traffic Prediction | 10476 |
| <i>Zilong Li, Qianqian Ren, Long Chen, Jianguo Sun</i> | |
| TFCnet: Time-Frequency Domain Corrector for Speech Separation | 10481 |
| <i>Weinan Tong, Jiaxu Zhu, Jun Chen, Zhiyong Wu, Shiyin Kang, Helen Meng</i> | |
| Conversation-Oriented ASR with Multi-Look-Ahead CBS Architecture..... | 10486 |
| <i>Huaibo Zhao, Shinya Fujie, Tetsuji Ogawa, Jin Sakuma, Yusuke Kida, Tetsunori Kobayashi</i> | |
| Real-Time Audio-Visual End-To-End Speech Enhancement | 10491 |
| <i>Zirun Zhu, Hemin Yang, Min Tang, Ziyi Yang, Sefik Emre Eskimez, Huaming Wang</i> | |
| Unsupervised Anomaly Detection and Localization of Machine Audio: A Gan-Based Approach | 10496 |
| <i>Anbai Jiang, Wei-Qiang Zhang, Yufeng Deng, Pingyi Fan, Jia Liu</i> | |
| Information Extraction from Pill Bottle Images Via Text Stitching | 10501 |
| <i>Rahul Kumar Gupta, Shilka Roy, Sujit Jos, Unni V. S., Lauren Lavoie, Frederic Medous, Walter Smith</i> | |
| VQ-CL: Learning Disentangled Speech Representations with Contrastive Learning and Vector Quantization | 10506 |
| <i>Huaizhen Tang, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao</i> | |
| No Reference Quality Assessment for Screen Content Images Based on Entire and High-Influence Regions..... | 10511 |
| <i>Zhuoran Xu, Yang Yang, Zhixiang Zhang, Weiming Zhang</i> | |
| Enhanced Dcf Tracker Regularized by Reliable Sample Construction | 10516 |
| <i>Kun Hu, Mingyu Cao, Mengzhu Wang, Long Lan, Wenjing Yang, Huibin Tan</i> | |
| Sparse Delay-Doppler Channel Estimation for OTFS Modulation Using 2D-Music..... | 10521 |
| <i>Akshay S. Bondre, Christ D. Richmond, Ahmed Alkhateeb, Nicolò Michelusi</i> | |
| A Contrastive Knowledge Transfer Framework for Model Compression and Transfer Learning..... | 10526 |
| <i>Kaiqi Zhao, Yitao Chen, Ming Zhao</i> | |
| Lightweight Feature Encoder for Wake-Up Word Detection Based on Self-Supervised Speech Representation | 10531 |
| <i>Hyungjun Lim, Younggwon Kim, Kiho Yeom, Eunjoo Seo, Hoodong Lee, Stanley Jungkyu Choi, Honglak Lee</i> | |
| Kernel Ridge Regression for Generalized Graph Signal Processing..... | 10536 |
| <i>Xingchao Jian, Wee Peng Tay</i> | |
| Adversarial Attacks on Genotype Sequences | 10541 |
| <i>Daniel Mas Montserrat, Alexander G. Ioannidis</i> | |
| Learning a Weight Map for Weakly-Supervised Localization..... | 10546 |
| <i>Tal Shaharabany, Lior Wolf</i> | |

| | |
|--|-------|
| Transcription Free Filler Word Detection with Neural Semi-CRFs..... | 10551 |
| <i>Ge Zhu, Yujia Yan, Juan-Pablo Caceres, Zhiyao Duan</i> | |
| Pyramid Spatial Feature Transform and Shared-Offsets Deformable Alignment Based Convolutional Network for HDR Imaging | 10556 |
| <i>Junda Liao, Qin Liu, Takeshi Ikenaga</i> | |
| Balanced Deep CCA for Bird Vocalization Detection | 10561 |
| <i>Sumit Kumar, B. Anshuman, Linus Rüttimann, Richard H. R. Hahnloser, Vipul Arora</i> | |
| Joint Noise Reduction and Listening Enhancement for Full-End Speech Enhancement | 10566 |
| <i>Haoyu Li, Yun Liu, Junichi Yamagishi</i> | |
| Restoration of Time-Varying Graph Signals Using Deep Algorithm Unrolling..... | 10571 |
| <i>Hayate Kojima, Hikari Noguchi, Koki Yamada, Yuichi Tanaka</i> | |
| Mask the Bias: Improving Domain-Adaptive Generalization of CTC-Based ASR with Internal Language Model Estimation..... | 10576 |
| <i>Nilaksh Das, Monica Sunkara, Sravan Bodapati, Jinglun Cai, Devang Kulshreshtha, Jeff Farris, Katrin Kirchhoff</i> | |
| Parallel Sentence-Level Explanation Generation for Real-World Low-Resource Scenarios | 10581 |
| <i>Yan Liu, Xiaokang Chen, Qi Dai</i> | |
| Comparative Layer-Wise Analysis of Self-Supervised Speech Models | 10586 |
| <i>Ankita Pasad, Bowen Shi, Karen Livescu</i> | |
| DepthFormer: Multimodal Positional Encodings and Cross-Input Attention for Transformer-based Segmentation Networks..... | 10591 |
| <i>Francesco Barbato, Giulia Rizzoli, Pietro Zanuttigh</i> | |
| Improved Belief Propagation Decoding of Turbo Codes | 10596 |
| <i>Yifei Shen, Yuqing Ren, Andreas Toftgaard Kristensen, Xiaohu You, Chuan Zhang, Andreas Burg</i> | |
| Learning Speech Representations with Flexible Hidden Feature Dimensions | 10601 |
| <i>Huaizhen Tang, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao</i> | |
| Personalized Lightweight Text-To-Speech: Voice Cloning with Adaptive Structured Pruning..... | 10606 |
| <i>Sung-Feng Huang, Chia-Ping Chen, Zhi-Sheng Chen, Yu-Pao Tsai, Hung-Yi Lee</i> | |
| QI-TTS: Questioning Intonation Control for Emotional Speech Synthesis..... | 10611 |
| <i>Haobin Tang, Xulong Zhang, Jianzong Wang, Ning Cheng, Jing Xiao</i> | |
| Using Adapters to Overcome Catastrophic Forgetting in End-To-End Automatic Speech Recognition | 10616 |
| <i>Steven Vander Eeckt, Hugo Van Hamme</i> | |
| Partially Adaptive Multichannel Joint Reduction of Ego-Noise and Environmental Noise | 10621 |
| <i>Huajian Fang, Niklas Wittmer, Johannes Twiefel, Stefan Wermter, Timo Gerkmann</i> | |
| URM4DMU: An User Representation Model for Darknet Markets Users..... | 10626 |
| <i>Hongmeng Liu, Jiapeng Zhao, Yixuan Huo, Yuyan Wang, Chun Liao, Liyan Shen, Shiyao Cui, Jinqiao Shi</i> | |

| | |
|---|-------|
| A Probabilistic Framework for Pruning Transformers Via a Finite Admixture of Keys..... | 10631 |
| <i>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</i> | |
| Asynchronous Federated Learning for Real-Time Multiple Licence Plate Recognition Through Semantic Communication | 10636 |
| <i>Renyou Xie, Chaojie Li, Xiaojun Zhou, Zhaoyang Dong</i> | |
| Image Reconstruction Without Explicit Priors | 10641 |
| <i>Angela F. Gao, Oscar Leong, He Sun, Katherine L. Bouman</i> | |
| DGN: Descriptor Generation Network for Feature Matching in Monocular Endoscopy 3D Reconstruction..... | 10646 |
| <i>Kaiyun Zhang, Wenkang Fan, Yinran Chen, Xiongbiao Luo</i> | |
| Alternating Phase Langevin Sampling with Implicit Denoiser Priors for Phase Retrieval..... | 10651 |
| <i>Rohun Agrawal, Oscar Leong</i> | |
| Emotion Recognition in Conversation from Variable-Length Context..... | 10656 |
| <i>Mian Zhang, Xiabing Zhou, Wenliang Chen, Min Zhang</i> | |
| Learning to Regularized Resource Allocation with Budget Constraints | 10661 |
| <i>Shaoke Fang, Qingsong Liu, Lei Xu, Wenfei Wu</i> | |
| Solving Audio Inverse Problems with a Diffusion Model..... | 10666 |
| <i>Eloi Moliner, Jaakko Lehtinen, Vesa Välimäki</i> | |
| MRNET: Multi-Refinement Network for Dual-Pixel Images Defocus Deblurring..... | 10671 |
| <i>Dafeng Zhang, Xiaobing Wang, Zhezhu Jin</i> | |
| Articulatory Representation Learning Via Joint Factor Analysis and Neural Matrix Factorization | 10676 |
| <i>Jiachen Lian, Alan W Black, Yijing Lu, Louis Goldstein, Shinji Watanabe, Gopala K. Anumanchipalli</i> | |
| Pretraining Conformer with ASR for Speaker Verification | 10681 |
| <i>Danwei Cai, Weiqing Wang, Ming Li, Rui Xia, Chuanzeng Huang</i> | |
| Deep Triple-Supervision Learning Unannotated Surgical Endoscopic Video Data for Monocular Dense Depth Estimation | 10686 |
| <i>Wenkang Fan, Kaiyun Zhang, Hong Shi, Jianhua Chen, Yinran Chen, Xiongbiao Luo</i> | |
| VE-KWS: Visual Modality Enhanced End-to-End Keyword Spotting | 10691 |
| <i>Ao Zhang, He Wang, Pengcheng Guo, Yihui Fu, Lei Xie, Yingying Gao, Shilei Zhang, Junlan Feng</i> | |
| StreamSpeech: Low-Latency Neural Architecture for High-Quality on-Device Speech Synthesis | 10696 |
| <i>Georgi Shopov, Stefan Gerdjikov, Stoyan Mihov</i> | |
| Local-Global Progressive U-Transformers for Accurate Hepatic and Portal Veins Segmentation in Abdominal MR Images | 10701 |
| <i>Yu Wu, Dongfang Shen, Jiabao Jin, Guanping Xu, Yinran Chen, Xiongbiao Luo</i> | |
| ACF: Aligned Contrastive Finetuning for Language and Vision Tasks..... | 10706 |
| <i>Wei Zhu, Peng Wang, Xiaoling Wang, Yuan Ni, Guotong Xie</i> | |
| DB-UNet: MLP Based Dual Branch UNet for Accurate Vessel Segmentation in OCTA Images | 10711 |
| <i>Chengliang Wang, Haojian Ning, Xinrun Chen, Shiyong Li</i> | |

| | |
|---|-------|
| Structured Pruning of Self-Supervised Pre-Trained Models for Speech Recognition and Understanding | 10716 |
| <i>Yifan Peng, Kwangyoum Kim, Felix Wu, Prashant Sridhar, Shinji Watanabe</i> | |
| Margin-Mixup: A Method for Robust Speaker Verification in Multi-Speaker Audio | 10721 |
| <i>Jenthe Thienpondt, Nilesh Madhu, Kris Demuyne</i> | |
| EI ² SR: Learning an Enhanced Intra-Instance Semantic Relationship for Arbitrary-Shaped Scene Text Detection | 10726 |
| <i>Yan Shu, Shaohui Liu, Yu Zhou, Honglei Xu, Feng Jiang</i> | |
| Numerical Semantic Modeling for Implicit Discourse Relation Recognition | 10731 |
| <i>Chenxu Wang, Ping Jian, Hai Wang</i> | |
| Forensics for Adversarial Machine Learning Through Attack Mapping Identification | 10736 |
| <i>Allen Yan, Jinsub Kim, Raviv Raich</i> | |
| AMPose: Alternately Mixed Global-Local Attention Model for 3D Human Pose Estimation..... | 10741 |
| <i>Hongxin Lin, Yunwei Chiu, Peiyuan Wu</i> | |
| Cross-Modal Mutual Learning for Cued Speech Recognition | 10746 |
| <i>Lei Liu, Li Liu</i> | |
| Multi-Dimensional Frequency Dynamic Convolution with Confident Mean Teacher for Sound Event Detection | 10751 |
| <i>Shengchang Xiao, Xueshuai Zhang, Pengyuan Zhang</i> | |
| Boosting Face Recognition Performance with Synthetic Data and Limited Real Data..... | 10756 |
| <i>Wenqing Wang, Lingqing Zhang, Chi-Man Pun, Jiu-Cheng Xie</i> | |
| The Potential of Neural Speech Synthesis-Based Data Augmentation for Personalized Speech Enhancement | 10761 |
| <i>Anastasia Kuznetsova, Aswin Sivaraman, Minje Kim</i> | |
| Deep Neural Mel-Subband Beamformer for in-Car Speech Separation..... | 10766 |
| <i>Vinay Kothapally, Yong Xu, Meng Yu, Shi-Xiong Zhang, Dong Yu</i> | |
| Multi-Task Bias-Variance Trade-Off Through Functional Constraints..... | 10771 |
| <i>Juan Cerviño, Juan Andrés Bazerque, Miguel Calvo-Fullana, Alejandro Ribeiro</i> | |
| M-CTRL: A Continual Representation Learning Framework with Slowly Improving Past Pre-Trained Model | 10776 |
| <i>Jin-Seong Choi, Jae-Hong Lee, Chae-Won Lee, Joon-Hyuk Chang</i> | |
| Designing and Evaluating Speech Emotion Recognition Systems: A Reality Check Case Study with IEMOCAP | 10781 |
| <i>Nikolaos Antoniou, Athanasios Katsamanis, Theodoros Giannakopoulos, Shrikanth Narayanan</i> | |
| A Deep Fusion Rule for Infrared and Visible Image Fusion: Feature Communication for Importance Assessment | 10786 |
| <i>Xuran Lv, Jinyong Cheng, Guohua Lv, Zhonghe Wei</i> | |
| Optimal Kernel for Real-Time Arbitrary-Shaped Text Detection..... | 10791 |
| <i>Haozhao Ma, Chuang Yang, Yuan Yuan, Qi Wang</i> | |

| | |
|---|-------|
| Particle Flow Gaussian Sum Particle Filter | 10796 |
| <i>Karthik Comandur, Yunpeng Li, Santosh Nannuru</i> | |
| Self-Supervised Accent Learning for Under-Resourced Accents Using Native Language Data | 10801 |
| <i>Mehul Kumar, Jiyeon Kim, Dhananjaya Gowda, Abhinav Garg, Chanwoo Kim</i> | |
| Learning to Locate the Text Forgery in Smartphone Screenshots | 10806 |
| <i>Zeqin Yu, Bin Li, Yuzhen Lin, Jinhua Zeng, Jishen Zeng</i> | |
| Noncoherent Multiuser Grassmannian Constellations for the Mimo Multiple Access Channel | 10811 |
| <i>Javier Álvarez-Vizoso, Diego Cuevas, Carlos Beltrán, Ignacio Santamaria, Vít Tucek, Gunnar Peters</i> | |
| Attention-Guided Deep Learning Framework for Movement Quality Assessment..... | 10816 |
| <i>Aditya Kanade, Mansi Sharma, Manivannan Muniyandi</i> | |
| Learning from Yourself: A Self-Distillation Method for Fake Speech Detection..... | 10821 |
| <i>Jun Xue, Cunhang Fan, Jiangyan Yi, Chenglong Wang, Zhengqi Wen, Dan Zhang, Zhao Lv</i> | |
| Rate-Distortion Optimization with Alternative References for UGC Video Compression..... | 10826 |
| <i>Xin Xiong, Eduardo Pavez, Antonio Ortega, Balu Adsumilli</i> | |
| MDR-MFI:Multi-Branch Decoupled Regression and Multi-Scale Feature Interaction for Partial-to-Partial Cloud Registration | 10831 |
| <i>Weidong Dai, Xuejun Yan, Jingjing Wang, Di Xie, Shiliang Pu</i> | |
| Static-Scene Constrained Optimization for Matrix/Tensor-Decomposition-free Foreground-Background Separation | 10836 |
| <i>Kazuki Naganuma, Shunsuke Ono</i> | |
| Exploring Universal Singing Speech Language Identification Using Self-Supervised Learning Based Front-End Features | 10841 |
| <i>Xingming Wang, Hao Wu, Chen Ding, Chuanzeng Huang, Ming Li</i> | |
| A Hybrid Deep Neural Network for Nonlinear Causality Analysis in Complex Industrial Control System | 10846 |
| <i>Tian Feng, Qiming Chen, Yao Shi, Xun Lang, Lei Xie, Hongye Su</i> | |
| Binauralization Robust to Camera Rotation Using 360° Videos | 10851 |
| <i>Masaki Yoshida, Ren Togo, Takahiro Ogawa, Miki Haseyama</i> | |
| KEPS-NET: Robust Parking Slot Detection Based Keypoint Estimation for High Localization Accuracy..... | 10856 |
| <i>Jaewoo Lee, Kapje Sung, Daeul Park, Younghan Jeon</i> | |
| Analysing Diffusion-Based Generative Approaches Versus Discriminative Approaches for Speech Restoration..... | 10861 |
| <i>Jean-Marie Lemercier, Julius Richter, Simon Welker, Timo Gerkmann</i> | |
| EMIX: A Data Augmentation Method for Speech Emotion Recognition..... | 10866 |
| <i>An Dang, Toan H. Vu, Le Dinh Nguyen, Jia-Ching Wang</i> | |
| SDRNet: Shape Decoupled Regression Network for 3D Face Reconstruction..... | 10871 |
| <i>Shikun Zhang, Fengyi Song, Ge Song, Ming Yang</i> | |

| | |
|--|-------|
| Low in Resolution, High in Precision: UAV Detection with Super-Resolution and Motion Information Extraction | 10876 |
| <i>Hanzhuo Wang, Xingjian Wang, Chengwei Zhou, Wenchao Meng, Zhiguo Shi</i> | |
| Self-Supervised Hierarchical Metrical Structure Modeling | 10881 |
| <i>Junyan Jiang, Gus Xia</i> | |
| Cosmopolite Sound Monitoring (CoSMo): A Study of Urban Sound Event Detection Systems Generalizing to Multiple Cities | 10886 |
| <i>Florian Angulo, Slim Essid, Geoffroy Peeters, Christophe Mietlicki</i> | |
| Improving Heart Rate and Heart Rate Variability Estimation from Video Through a HR-RR-Tuned Filter | 10891 |
| <i>Michael Chan, Li Zhu, Korosh Vatanparvar, Hewon Jung, Jilong Kuang, Alex Gao</i> | |
| Dense Adversarial Transfer Learning Based on Class-Invariance | 10896 |
| <i>Bach-Tung Pham, Ting-Yu Wang, Phuong Le Thi, Khai-Thinh Nguyen, Yuan-Shan Lee, Tzu-Chiang Tai, Jia-Ching Wang</i> | |
| Uncertainty Estimation in Deep Speech Enhancement Using Complex Gaussian Mixture Models | 10901 |
| <i>Huajian Fang, Timo Gerkmann</i> | |
| Multi-View Learning for Speech Emotion Recognition with Categorical Emotion, Categorical Sentiment, and Dimensional Scores | 10906 |
| <i>Daniel Tompkins, Dimitra Emmanouilidou, Soham Deshmukh, Benjamin Elizalde</i> | |
| Performance Above All? Energy Consumption Vs. Performance, a Study on Sound Event Detection with Heterogeneous Data | 10911 |
| <i>Romain Serizel, Samuele Cornell, Nicolas Turpault</i> | |
| Unsupervised Action Segmentation of Untrimmed Egocentric Videos | 10916 |
| <i>Sam Perochon, Laurent Oudre</i> | |
| Self-Remixing: Unsupervised Speech Separation VIA Separation and Remixing | 10921 |
| <i>Kohei Saijo, Tetsuji Ogawa</i> | |
| Stargan-Vc Based Cross-Domain Data Augmentation for Speaker Verification | 10926 |
| <i>Hang-Rui Hu, Yan Song, Jian-Tao Zhang, Li-Rong Dai, Ian McLoughlin, Zhu Zhuo, Yu Zhou, Yu-Hong Li, Hui Xue</i> | |
| Investigation into Phone-Based Subword Units for Multilingual End-to-End Speech Recognition | 10931 |
| <i>Saierdaer Yusuyin, Hao Huang, Junhua Liu, Cong Liu</i> | |
| Bat: Bi-Alignment Based on Transformation in Multi-Target Domain Adaptation for Semantic Segmentation | 10936 |
| <i>Xian Zhong, Wei Li, Liang Liao, Jing Xiao, Wenxuan Liu, Wenxin Huang, Zheng Wang</i> | |
| Building Change Detection Using Cross-Temporal Feature Interaction Network | 10941 |
| <i>Yuchao Feng, Jiawei Jiang, Honghui Xu, Jianwei Zheng</i> | |
| Flowpose: Conditional Normalizing Flows for 3D Human Pose and Shape Estimation from Monocular Videos | 10946 |
| <i>Yaoyao Du, Zixiao Zhang, Zhihao Li, Peng Wei, Qingmin Liao, Wenming Yang</i> | |
| Random Projector: Efficient Deep Image Prior | 10951 |
| <i>Taihui Li, Zhong Zhuang, Hengkang Wang, Ju Sun</i> | |

| | |
|---|-------|
| Background-Weakening Consistency Regularization for Semi-Supervised Video Action Detection | 10956 |
| <i>Xian Zhong, Aoyu Yi, Wenxuan Liu, Wenxin Huang, Chengming Zou, Zheng Wang</i> | |
| Convolutional NTF for Ambisonic Source Separation Under Reverberant Conditions | 10961 |
| <i>Mateusz Guzik, Konrad Kowalczyk</i> | |
| Calibrating AI Models for Few-Shot Demodulation VIA Conformal Prediction | 10966 |
| <i>Kfir M. Cohen, Sangwoo Park, Osvaldo Simeone, Shlomo Shamai Shitz</i> | |
| Rethinking Random Walk in Graph Representation Learning | 10971 |
| <i>Dingyi Zeng, Wenyu Chen, Wanlong Liu, Li Zhou, Hong Qu</i> | |
| Beyond Rate Coding: Signal Coding and Reconstruction Using Lean Spike Trains..... | 10976 |
| <i>Anik Chattopadhyay, Arunava Banerjee</i> | |
| Detecting Malicious Migration on Edge to Prevent Running Data Leakage..... | 10981 |
| <i>Yuchen Wong, Qingni Shen, Cong Li, Cunzhan Liu, Tianxiang Ai</i> | |
| Deformable Cross Attention for Learning Optical Flow | 10986 |
| <i>Rokia Abdein, Xuezhi Xiang, Ning Lv, Abdulmotaleb El Saddik</i> | |
| Optimized Quality Feature Learning for Video Quality Assessment..... | 10991 |
| <i>Ngai-Wing Kwong, Yui-Lam Chan, Sik-Ho Tsang, Daniel Pak-Kong Lun</i> | |
| Next-Speaker Prediction Based on Non-Verbal Information in Multi-Party Video Conversation | 10996 |
| <i>Saki Mizuno, Nobukatsu Hojo, Satoshi Kobashikawa, Ryo Masumura</i> | |
| Second-Order Statistic Deviation to Model Anomalies in the Design of Unsupervised Detectors | 11001 |
| <i>Andriy Enttsel, Filippo Martinini, Alex Marchioni, Mauro Mangia, Riccardo Rovatti, Gianluca Setti</i> | |
| Clustering-Based Supervised Contrastive Learning for Identifying Risk Items on Heterogeneous Graph..... | 11006 |
| <i>Ao Li, Yugang Ji, Guanyi Chu, Xiao Wang, Dong Li, Chuan Shi</i> | |
| DiffPhase: Generative Diffusion-Based STFT Phase Retrieval | 11011 |
| <i>Tal Peer, Simon Welker, Timo Gerkmann</i> | |
| LiteG2P: A Fast, Light and High Accuracy Model for Grapheme-To-Phoneme Conversion..... | 11016 |
| <i>Chunfeng Wang, Peisong Huang, Yuxiang Zou, Haoyu Zhang, Shichao Liu, Xiang Yin, Zejun Ma</i> | |
| Expressive-VC: Highly Expressive Voice Conversion with Attention Fusion of Bottleneck and Perturbation Features..... | 11021 |
| <i>Ziqian Ning, Qicong Xie, Pengcheng Zhu, Zhichao Wang, Liumeng Xue, Jixun Yao, Lei Xie, Mengxiao Bi</i> | |
| Recouple Event Field Via Probabilistic Bias for Event Extraction..... | 11026 |
| <i>Xingyu Bai, Taiqiang Wu, Han Guo, Zhe Zhao, Xuefeng Yang, Jiayi Li, Weijie Liu, Qi Ju, Weigang Guo, Yujiu Yang</i> | |
| Self-Supervised Learning for Speech Enhancement Through Synthesis..... | 11031 |
| <i>Bryce Irvin, Marko Stamenovic, Mikolaj Kegler, Li-Chia Yang</i> | |
| Rigid-Body Sound Synthesis with Differentiable Modal Resonators..... | 11036 |
| <i>Rodrigo Diaz, Ben Hayes, Charalampos Saitis, György Fazekas, Mark Sandler</i> | |

| | |
|--|-------|
| Unsupervised Voice Type Discrimination Score Adaptation Using X-Vector Clusters | 11041 |
| <i>Mark Lindsey, Tyler Vuong, Richard M. Stern</i> | |
| Toroidal Probabilistic Spherical Discriminant Analysis | 11046 |
| <i>Anna Silnova, Niko Brümmer, Albert Swart, Lukáš Burget</i> | |
| Large-Scale Nonverbal Vocalization Detection Using Transformers | 11051 |
| <i>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</i> | |
| Multilingual Query-By-Example Keyword Spotting with Metric Learning and Phoneme-to-Embedding Mapping | 11056 |
| <i>Paul M. Reuter, Christian Rollwage, Bernd T. Meyer</i> | |
| Exploiting 3D Human Recovery for Action Recognition with Spatio-Temporal Bifurcation Fusion | 11061 |
| <i>Na Jiang, Wei Quan, Qichuan Geng, Zhiping Shi, Peng Xu</i> | |
| Fast and Efficient Speech Enhancement with Variational Autoencoders | 11066 |
| <i>Mostafa Sadeghi, Romain Serizel</i> | |
| Joint Compression and Demosaicking for Satellite Images | 11071 |
| <i>Pascal Bacchus, Renaud Fraisse, Aline Roumy, Christine Guillemot</i> | |
| Think Before You Speak: Concept-Guided Explicit Persona Reasoning for Personalized Dialogue Generation | 11076 |
| <i>Yunpeng Li, Yue Hu, Wei Peng, Yuqiang Xie</i> | |
| Distributed Bayesian Tracking on the Special Euclidean Group Using Lie Algebra Parametric Approximations | 11081 |
| <i>Claudio J. Bordin, Caio G. De Figueredo, Marcelo G. S. Bruno</i> | |
| Fast and Parallel Decoding for Transducer..... | 11086 |
| <i>Wei Kang, Liyong Guo, Fangjun Kuang, Long Lin, Mingshuang Luo, Zengwei Yao, Xiaoyu Yang, Piotr Zelasko, Daniel Povey</i> | |
| Difficulty-Aware Data Augmentor for Scene Text Recognition | 11091 |
| <i>Guanghao Meng, Tao Dai, Bin Chen, Naiqi Li, Yong Jiang, Shu-Tao Xia</i> | |
| Learning Interpretable Filters in Wav-UNet for Speech Enhancement | 11096 |
| <i>Félix Mathieu, Thomas Courtat, Gaël Richard, Geoffroy Peeters</i> | |
| Background Disturbance Mitigation for Video Captioning Via Entity-Action Relocation | 11101 |
| <i>Zipeng Li, Xian Zhong, Shuqin Chen, Wenxuan Liu, Wenxin Huang, Lin Li</i> | |
| Sinusoidal Frequency Estimation by Gradient Descent | 11106 |
| <i>Ben Hayes, Charalampos Saitis, György Fazekas</i> | |
| Multi-View K-Means with Laplacian Embedding..... | 11111 |
| <i>Zhezheng Hao, Zhoumin Lu, Feiping Nie, Rong Wang, Xuelong Li</i> | |
| Going in Style: Audio Backdoors Through Stylistic Transformations | 11116 |
| <i>Stefanos Koffas, Luca Pajola, Stjepan Picek, Mauro Conti</i> | |
| Model-Based Vs. Data-driven Approaches for Predicting Rain-induced Attenuation in Commercial Microwave Links: A Comparative Empirical Study..... | 11121 |
| <i>Dror Jacoby, Jonatan Ostrometzky, Hagit Messer</i> | |

| | |
|--|-------|
| Joint Multi-Level Feature Network for Lightweight Person Re-Identification | 11126 |
| <i>Yunzuo Zhang, Weili Kang, Yameng Liu, Pengfei Zhu</i> | |
| Audio Signal Enhancement with Learning from Positive and Unlabeled Data..... | 11131 |
| <i>Nobutaka Ito, Masashi Sugiyama</i> | |
| Active Perception System for Enhanced Visual Signal Recovery Using Deep Reinforcement Learning | 11136 |
| <i>Gaurav Chaudhary, Laxmidhar Behera, Tushar Sandhan</i> | |
| Federated Learning for ASR Based on Wav2vec 2.0..... | 11141 |
| <i>Tuan Nguyen, Salima Mdhaffar, Natalia Tomashenko, Jean-François Bonastre, Yannick Estève</i> | |
| Learning on Graphs Under Label Noise..... | 11146 |
| <i>Jingyang Yuan, Xiao Luo, Yifang Qin, Yusheng Zhao, Wei Ju, Ming Zhang</i> | |
| Robust Hyperspectral Anomaly Detection with Simultaneous Mixed Noise Removal Via Constrained Convex Optimization | 11151 |
| <i>Koyo Sato, Shunsuke Ono</i> | |
| Improving Knowledge Distillation for Non-Intrusive Load Monitoring Through Explainability Guided Learning..... | 11156 |
| <i>Djordje Batic, Giulia Tanoni, Lina Stankovic, Vladimir Stankovic, Emanuele Principi</i> | |
| Learning Unbiased Rewards with Mutual Information in Adversarial Imitation Learning | 11161 |
| <i>Lihua Zhang, Quan Liu, Zhigang Huang, Lan Wu</i> | |
| Confidence-Based Event-Centric Online Video Question Answering on a Newly Constructed ATBS Dataset..... | 11166 |
| <i>Weikai Kong, Shuhong Ye, Chenglin Yao, Jianfeng Ren</i> | |
| Auto-AVSR: Audio-Visual Speech Recognition with Automatic Labels | 11171 |
| <i>Pingchuan Ma, Alexandros Haliassos, Adriana Fernandez-Lopez, Honglie Chen, Stavros Petridis, Maja Pantic</i> | |
| ClassA Entropy for the Analysis of Structural Complexity of Physiological Signals | 11176 |
| <i>Hongjian Xiao, Ling Li, Danilo P. Mandic</i> | |
| Classifying Pathological Images Based on Multi-Instance Learning and End-to-End Attention Pooling | 11181 |
| <i>Yuqi Chen, Juan Liu, Zhiqun Zuo, Peng Jiang, Yu Jin, Guangsheng Wu</i> | |
| Imaginary Voice: Face-Styled Diffusion Model for Text-to-Speech | 11186 |
| <i>Jiyoung Lee, Joon Son Chung, Soo-Whan Chung</i> | |
| Learning to Balance the Global Coherence and Informativeness in Knowledge-Grounded Dialogue Generation | 11191 |
| <i>Chenxu Niu, Yue Hu, Wei Peng, Yuqiang Xie</i> | |
| Note and Playing Technique Transcription of Electric Guitar Solos in Real-World Music Performance..... | 11196 |
| <i>Tung-Sheng Huang, Ping-Chung Yu, Li Su</i> | |
| Differential Analysis for Networks Obeying Conservation Laws | 11201 |
| <i>Anirudh Rayas, Rajasekhar Anguluri, Jiajun Cheng, Gautam Dasarathy</i> | |

| | |
|--|-------|
| CPD-GAN: Cascaded Pyramid Deformation GAN for Pose Transfer | 11206 |
| <i>Yuan Huang, Yuting Tang, Xiu Zheng, Jie Tang</i> | |
| Training Robust Spiking Neural Networks with Viewpoint Transform and Spatiotemporal Stretching | 11211 |
| <i>Haibo Shen, Juyu Xiao, Yihao Luo, Xiang Cao, Liangqi Zhang, Tianjiang Wang</i> | |
| BIRD-PCC: Bi-Directional Range Image-Based Deep Lidar Point Cloud Compression | 11216 |
| <i>Chia-Sheng Liu, Jia-Fong Yeh, Hao Hsu, Hung-Ting Su, Ming-Sui Lee, Winston H. Hsu</i> | |
| Sparse Error Correction for Power Network Parameters..... | 11221 |
| <i>Dilan Senaratne, Jinsub Kim</i> | |
| Higher-Order Spatio-Temporal Neural Networks for Covid-19 Forecasting | 11226 |
| <i>Yuzhou Chen, Sotiris Batsakis, H. Vincent Poor</i> | |
| Improving Accented Speech Recognition with Multi-Domain Training | 11231 |
| <i>Lucas Maison, Yannick Esteve</i> | |
| Fully Unsupervised Topic Clustering of Unlabelled Spoken Audio Using Self-Supervised Representation Learning and Topic Model..... | 11236 |
| <i>Takashi Maekaku, Yuya Fujita, Xuankai Chang, Shinji Watanabe</i> | |
| Centralized Cascade Multi-Channel Noise Reduction and Acoustic Feedback Cancellation in a Wireless Acoustic Sensor and Actuator Network | 11241 |
| <i>Santiago Ruiz, Toon Van Waterschoot, Marc Moonen</i> | |
| Hindi as a Second Language: Improving Visually Grounded Speech with Semantically Similar Samples | 11246 |
| <i>Hyeonggon Ryu, Arda Senocak, In So Kweon, Joon Son Chung</i> | |
| Disentangling Speech from Surroundings with Neural Embeddings | 11251 |
| <i>Ahmed Omran, Neil Zeghidour, Zalán Borsos, Félix De Chaumont Quitry, Malcolm Slaney, Marco Tagliasacchi</i> | |
| Progressive Perception Learning for Distribution Modulation in Siamese Tracking | 11256 |
| <i>Kun Hu, Xianchen Zhou, Mingyu Cao, Mengzhu Wang, Guangjie Gao, Wenjing Yang, Huibin Tan</i> | |
| Pooling Strategies for Simplicial Convolutional Networks..... | 11261 |
| <i>Domenico Mattia Cinque, Claudio Battiloro, Paolo Di Lorenzo</i> | |
| Efficient Uncertainty Estimation with Gaussian Process for Reliable Dialog Response Retrieval | 11266 |
| <i>Tong Ye, Zhitao Li, Jianzong Wang, Ning Cheng, Jing Xiao</i> | |
| Immersive Enhancement and Removal of Loudspeaker Sound Using Wireless Assistive Listening Systems and Binaural Hearing Devices..... | 11271 |
| <i>Ryan M. Corey, Andrew C. Singer</i> | |
| Cross-Subject Mental Fatigue Detection Based on Separable Spatio-Temporal Feature Aggregation..... | 11276 |
| <i>Yalan Ye, Yutuo He, Wanjing Huang, Qiaosen Dong, Chong Wang, Guoqing Wang</i> | |
| Preserving Background Sound in Noise-Robust Voice Conversion Via Multi-Task Learning | 11281 |
| <i>Jixun Yao, Yi Lei, Qing Wang, Pengcheng Guo, Ziqian Ning, Lei Xie, Hai Li, Junhui Liu, Danming Xie</i> | |

| | |
|--|-------|
| Any-To-Any Voice Conversion with F0 and Timbre Disentanglement and Novel Timbre Conditioning..... | 11286 |
| <i>Sudheer Kovela, Rafael Valle, Ambrish Dantrey, Bryan Catanzaro</i> | |
| Enhancing Unsupervised Speech Recognition with Diffusion GANS | 11291 |
| <i>Xianchao Wu</i> | |
| Learning Sparse Alignments Via Optimal Transport for Cross-Domain Fake News Detection..... | 11296 |
| <i>Wei Tang, Zuyao Ma, Haifeng Sun, Jingyu Wang</i> | |
| Sampling Order-Limited Signals on the Sphere..... | 11301 |
| <i>Muhammad Salaar Arif Khan, Salman Nadeem, Zubair Khalid</i> | |
| Large Dimensional Analysis of LS-SVM Transfer Learning: Application to Polars Classification..... | 11306 |
| <i>Cyprien Doz, Chengfang Ren, Jean-Philippe Ovarlez, Romain Couillet</i> | |
| Robust Autoencoders for Collective Corruption Removal..... | 11311 |
| <i>Taihui Li, Hengkang Wang, Le Peng, Xian 'E Tang, Ju Sun</i> | |
| Representation of Vocal Tract Length Transformation Based on Group Theory..... | 11316 |
| <i>Atsushi Miyashita, Tomoki Toda</i> | |
| Make Your Enemy Your Friend: Improving Image Rotation Angle Estimation with Harmonics | 11321 |
| <i>Kun Yu, Morteza Darvish Morshedi Hosseini, Anjie Peng, Hui Zeng, Miroslav Goljan</i> | |
| Jeffreys Divergence-Based Regularization of Neural Network Output Distribution Applied to Speaker Recognition..... | 11326 |
| <i>Pierre-Michel Bousquet, Mickael Rouvier</i> | |
| Single-Channel Speech Enhancement with Deep Complex U-Networks and Probabilistic Latent Space Models | 11331 |
| <i>Eike J. Nustede, Jörn Anemüller</i> | |
| Spatial Inference Using Censored Multiple Testing with Fdr Control | 11336 |
| <i>Martin Gözl, Abdelhak M. Zoubir, Visa Koivunen</i> | |
| YOLOX-B: A Better YOLOX Model for Real-Time Driver Behavior Detection | 11341 |
| <i>Xu Guo, Ming Ma, Jiaqiang Zhang, Shaojie Li</i> | |
| Towards a More Stable and General Subgraph Information Bottleneck | 11346 |
| <i>Hongzhi Liu, Kaizhong Zheng, Shujian Yu, Badong Chen</i> | |
| Adaptive Scale and Spatial Aggregation for Real-Time Object Detection..... | 11351 |
| <i>Wei Chen, Yulin He, Zhengfa Liang, Yulan Guo</i> | |
| LE-DTA: Local Extrema Convolution for Drug Target Affinity Prediction..... | 11356 |
| <i>Tanoj Langore, Te-Cheng Hsu, Yi-Hsien Hsieh, Che Lin</i> | |
| An Empirical Study and Improvement for Speech Emotion Recognition..... | 11361 |
| <i>Zhen Wu, Yizhe Lu, Xinyu Dai</i> | |
| Phonetic RNN-Transducer for Mispronunciation Diagnosis..... | 11366 |
| <i>Daniel Yue Zhang, Soumya Saha, Sarah Campbell</i> | |
| High-Acoustic Fidelity Text to Speech Synthesis with Fine-Grained Control of Speech Attributes..... | 11371 |
| <i>Rafael Valle, João Felipe Santos, Kevin J. Shih, Rohan Badlani, Bryan Catanzaro</i> | |

| | |
|--|-------|
| Source Localization for Extremely Large-Scale Antenna Arrays with Spatial Non-Stationarity | 11376 |
| <i>Xiaohuan Wu, Ji Sun, Xiaoyuan Jia, Shuxin Wang</i> | |
| Rain2Avoid: Self-Supervised Single Image Deraining | 11381 |
| <i>Yan-Tsung Peng, Wei-Hua Li</i> | |
| Joint Human Orientation-Activity Recognition Using WIFI Signals for Human-Machine Interaction | 11386 |
| <i>Hojjat Salehinejad, Navid Hasanzadeh, Radomir Djogo, Shahrokh Valaee</i> | |
| AE-Flow: Autoencoder Normalizing Flow | 11391 |
| <i>Jakub Mosinski, Piotr Bilinski, Thomas Merritt, Abdelhamid Ezzer, Daniel Korzekwa</i> | |
| An Empirical Study on Speech Restoration Guided by Self-Supervised Speech Representation | 11396 |
| <i>Jaeuk Byun, Youna Ji, Soo-Whan Chung, Soyeon Choe, Min-Seok Choi</i> | |
| Transferring Quantified Emotion Knowledge for the Detection of Depression in Alzheimer's Disease Using Forestnets..... | 11401 |
| <i>P. A. Pérez-Toro, D. Rodríguez-Salas, T. Arias-Vergara, S. P. Bayerl, P. Klumpp, K. Riedhammer, M. Schuster, E. Nöth, A. Maier, J. R. Orozco-Arroyave</i> | |
| Unsupervised Noise Adaptation Using Data Simulation | 11406 |
| <i>Chen Chen, Yuchen Hu, Heqing Zou, Linhui Sun, Eng Siong Chng</i> | |
| A Privacy-Preserving Trajectory Mining Model | 11411 |
| <i>Ziyang Wang, Sissi Xiaoxiao Wu, Junjie Zhu, Yingying Zhu</i> | |
| Element Selection with Wide Class of Optimization Criteria Using Non-Convex Sparse Optimization..... | 11416 |
| <i>Taiga Kawamura, Natsuki Ueno, Nobutaka Ono</i> | |
| Efficient Implementation of Robust CUSUM Algorithm to Characterize Nanogaps Measurements with Heavy-Tailed Noise | 11421 |
| <i>Javier Kipen, Joakim Jaldén, Shyamprasad N. Raja, Saumey Jain</i> | |
| PMMSD: Development of the Matrix Sentence Intelligibility Dataset for Mandarin with Lombard Effect | 11426 |
| <i>Hanchen Pei, Yuhong Yang, Xufeng Chen, Qingmu Liu, Hongyang Chen, Weiping Tu, Song Lin</i> | |
| Robust Content-Variant Reference Image Quality Assessment Via Similar Patch Matching..... | 11431 |
| <i>Wenbo Shi, Wenming Yang, Qingmin Liao</i> | |
| Neighborhood Information-Based Label Refinement for Person Re-Identification with Label Noise..... | 11436 |
| <i>Xian Zhong, Shuai peng Su, Wenxuan Liu, Xuemei Jia, Wenxin Huang, Mengdie Wang</i> | |
| A Game of Snakes and Gans | 11441 |
| <i>Siddarth Asokan, Fatwir Sheikh Mohammed, Chandra Sekhar Seelamantula</i> | |
| Joint Millimeter-Wave AoD and AoA Estimation Using One OFDM Symbol and Frequency-Dependent Beams..... | 11446 |
| <i>Veljko Boljanovic, Danijela Cabric</i> | |
| Quantized Precoding and RIS-Assisted Modulation for Integrated Sensing and Communications Systems..... | 11451 |
| <i>R. S. Prasobh Sankar, Sundeep Prabhakar Chepuri</i> | |

| | |
|--|-------|
| Deep Architecture for DOA Trajectory Localization..... | 11456 |
| <i>Shreyas Jaiswal, Ruchi Pandey, Santosh Nannuru</i> | |
| Ancient Chinese Word Segmentation and Part-Of-Speech Tagging Using Distant Supervision..... | 11461 |
| <i>Shuo Feng, Piji Li</i> | |
| SIAST: A Slot Imbalance-Aware Self-Training Scheme for Semi-Supervised Slot Filling..... | 11466 |
| <i>Jiachi Liu, Sishi Xiong, Yuehuan He, Tong Zhou, Liwen Wang, Xuefeng Li, Bo Xiao</i> | |
| A Novel Transformer-Based Pipeline for Lung Cytopathological Whole Slide Image Classification | 11471 |
| <i>Gaojie Li, Qing Liu, Haotian Liu, Yixiong Liang</i> | |
| End-To-End Non-Autoregressive Image Captioning..... | 11476 |
| <i>Hong Yu, Yuanqiu Liu, Baokun Qi, Zhaolong Hu, Han Liu</i> | |
| Spatio-Temporal Hybrid Fusion of CAE and SWin Transformers for Lung Cancer Malignancy Prediction | 11481 |
| <i>Sadaf Khademi, Shahin Heidarian, Parnian Afshar, Farnoosh Naderkhani, Anastasia Oikonomou, Konstantinos N. Plataniotis, Arash Mohammadi</i> | |
| Moving Towards Non-Binary Gender Identification Via Analysis of System Errors in Binary Gender Classification | 11486 |
| <i>Sebastian Ellis, Stefan Goetze, Heidi Christensen</i> | |
| Improving Few-Shot Learning for Talking Face System with TTS Data Augmentation..... | 11491 |
| <i>Qi Chen, Ziyang Ma, Tao Liu, Xu Tan, Qu Lu, Kai Yu, Xie Chen</i> | |
| On the Fairness of Multitask Representation Learning | 11496 |
| <i>Yingcong Li, Samet Oymak</i> | |
| Animal Re-Identification Algorithm for Posture Diversity | 11501 |
| <i>Zhimin He, Jiangbo Qian, Diqun Yan, Chong Wang, Yu Xin</i> | |
| Diffusion Probabilistic Modeling for Fine-Grained Urban Traffic Flow Inference with Relaxed Structural Constraint..... | 11506 |
| <i>Xovee Xu, Yutao Wei, Pengyu Wang, Xucheng Luo, Fan Zhou, Goce Trajcevski</i> | |
| PCQA-Graphpoint: Efficient Deep-Based Graph Metric for Point Cloud Quality Assessment..... | 11511 |
| <i>Marouane Tliba, Aladine Chetouani, Giuseppe Valenzise, Frédéric Dufaux</i> | |
| Low-Rank Plus Sparse Trajectory Decomposition for Direct Exoplanet Imaging | 11516 |
| <i>Simon Vary, Hazan Daglayan, Laurent Jacques, P.-A. Absil</i> | |
| Classification of Synthetic Facial Attributes by Means of Hybrid Classification/Localization Patch-Based Analysis | 11521 |
| <i>Jun Wang, Benedetta Tondi, Mauro Barni</i> | |
| Leveraging Pretrained Representations with Task-Related Keywords for Alzheimer’s Disease Detection | 11526 |
| <i>Jinchao Li, Kaitao Song, Junan Li, Bo Zheng, Dongsheng Li, Xixin Wu, Xunying Liu, Helen Meng</i> | |
| Towards Learning Emotion Information from Short Segments of Speech..... | 11531 |
| <i>Tilak Purohit, Sarthak Yadav, Bogdan Vlasenko, S. Pavankumar Dubagunta, Mathew Magimai.-Doss</i> | |

| | |
|--|-------|
| Unsupervised Word Segmentation Based on Word Influence | 11536 |
| <i>Ruohao Yan, Huaping Zhang, Wushour Silamu, Askar Hamdulla</i> | |
| Classification-Based Dynamic Network for Efficient Super-Resolution | 11541 |
| <i>Qi Wang, Weiwei Fang, Meng Wang, Yusong Cheng</i> | |
| Vision Transformer with Progressive Tokenization for CT Metal Artifact Reduction | 11546 |
| <i>Songwei Zheng, Dong Zhang, Chunyan Yu, Danhong Zhu, Longlong Zhu, Hao Liu, Zhongzheng Huang</i> | |
| The DKU Post-Challenge Audio-Visual Wake Word Spotting System for the 2021 MISP Challenge: Deep Analysis..... | 11551 |
| <i>Haoxu Wang, Ming Cheng, Qiang Fu, Ming Li</i> | |
| Absolute Decision Corrupts Absolutely: Conservative Online Speaker Diarisation | 11556 |
| <i>Youngki Kwon, Hee-Soo Heo, Bong-Jin Lee, You Jin Kim, Jee-Weon Jung</i> | |
| Incorporating Uncertainty from Speaker Embedding Estimation to Speaker Verification | 11561 |
| <i>Qionqiong Wang, Kong Aik Lee, Tianchi Liu</i> | |
| DocRED-FE: A Document-Level Fine-Grained Entity and Relation Extraction Dataset | 11566 |
| <i>Hongbo Wang, Weimin Xiong, Yifan Song, Dawei Zhu, Yu Xia, Sujian Li</i> | |
| Heart Rate Extraction from Abdominal Audio Signals..... | 11571 |
| <i>Jake Stuchbury-Wass, Erika Bondareva, Kayla-Jade Butkow, Sanja Šćepanovic, Zoran Radivojevic, Cecilia Mascolo</i> | |
| PUFFIN: Pitch-Synchronous Neural Waveform Generation for Fullband Speech on Modest Devices | 11576 |
| <i>Oliver Watts, Lovisa Wihlborg, Cassia Valentini-Botinhao</i> | |
| Deep Plug-And-Play for Tensor Robust Principal Component Analysis..... | 11581 |
| <i>Hao Tan, Jianjun Wang, Weichao Kong</i> | |
| ST-MVDNet++: Improve Vehicle Detection with Lidar-Radar Geometrical Augmentation Via Self- Training | 11586 |
| <i>Yu-Jhe Li, Matthew O'Toole, Kris Kitani</i> | |
| Two-Stream Joint-Training for Speaker Independent Acoustic-to-Articulatory Inversion..... | 11591 |
| <i>Jianrong Wang, Jinyu Liu, Xuwei Li, Mei Yu, Jie Gao, Qiang Fang, Li Liu</i> | |
| Lattice-Free Sequence Discriminative Training for Phoneme-Based Neural Transducers..... | 11596 |
| <i>Zijian Yang, Wei Zhou, Ralf Schlüter, Hermann Ney</i> | |
| Optimization of Sensor Configurations for Fault Identification in Smart Buildings..... | 11601 |
| <i>Naveed Ahmad, Malcolm Egan, Jean-Marie Gorce, Jilles S. Dibangoye, Frédéric Le Mouël</i> | |
| Prototype-Based Layered Federated Cross-Modal Hashing..... | 11606 |
| <i>Jiale Liu, Yu-Wei Zhan, Xin Luo, Zhen-Duo Chen, Yongxin Wang, Xin-Shun Xu</i> | |
| Hankel Structured Low Rank and Sparse Representation Via L0-Norm Optimization for Compressed Ultrasound Plane Wave Signal Reconstruction..... | 11611 |
| <i>Miaomiao Zhang, Ji Chen, Xiaoyan Fu, Ge Xin, Jingzhi Zhang, Na Jiang, Jan D'Hooge</i> | |
| Variational Bayesian Channel Estimation in Wideband Multi-Scale Multi-Lag Channels | 11616 |
| <i>Niladri Halder, Arunkumar K. P., Chandra R. Murthy</i> | |

| | |
|--|-------|
| Self-Healing Through Error Detection, Attribution, and Retraining | 11621 |
| <i>Ansel Maclaughlin, Anna Rumshisky, Rinat Khaziev, Anil Ramakrishna, Yuval Merhav, Rahul Gupta</i> | |
| Multi-Aspect Interest Neighbor-Augmented Network for Next-Basket Recommendation..... | 11626 |
| <i>Zhiying Deng, Jianjun Li, Zhiqiang Guo, Guohui Li</i> | |
| Wireless Deep Speech Semantic Transmission | 11631 |
| <i>Zixuan Xiao, Shengshi Yao, Jincheng Dai, Sixian Wang, Kai Niu, Ping Zhang</i> | |
| Metric-Oriented Speech Enhancement Using Diffusion Probabilistic Model | 11636 |
| <i>Chen Chen, Yuchen Hu, Weiwei Weng, Eng Siong Chng</i> | |
| Decomposition, Interaction, Reconstruction Meets Global Context Learning in Visual Tracking..... | 11641 |
| <i>Huibin Tan, Kun Hu, Mingyu Cao, Mengzhu Wang, Liyang Xu, Wenjing Yang</i> | |
| Question Answering System with Sparse and Noisy Feedback..... | 11646 |
| <i>Djallel Bouneffouf, Ozgur Alkan, Raphael Feraud, Baihan Lin</i> | |
| Single-Sample Direction-of-Arrival Estimation for Fast and Robust 3D Localization with Real Measurements from a Massive MIMO System..... | 11651 |
| <i>Stepan Mazokha, Sanaz Naderi, Georgios I. Orfanidis, George Sklivanitis, Dimitris A. Pados, Jason O. Hallstrom</i> | |
| Factorized AED: Factorized Attention-Based Encoder-Decoder for Text-Only Domain Adaptive ASR..... | 11656 |
| <i>Xun Gong, Wei Wang, Hang Shao, Xie Chen, Yanmin Qian</i> | |
| Positive-Pair Redundancy Reduction Regularisation for Speech-Based Asthma Diagnosis Prediction | 11661 |
| <i>Georgios Rizos, Rafael A. Calvo, Björn W. Schuller</i> | |
| Incorporating Lip Features into Audio-Visual Multi-Speaker DOA Estimation by Gated Fusion | 11666 |
| <i>Ya Jiang, Hang Chen, Jun Du, Qing Wang, Chin-Hui Lee</i> | |
| A Large-Scale Pretrained Deep Model for Phishing URL Detection | 11671 |
| <i>Yanbin Wang, Weifan Zhu, Haitao Xu, Zhan Qin, Kui Ren, Wenrui Ma</i> | |
| Code-Switching Speech Synthesis Based on Self-Supervised Learning and Domain Adaptive Speaker Encoder..... | 11676 |
| <i>Yi-Xing Lin, Cheng-Hsun Pai, Phuong Thi Le, Bima Prihasto, Chien-Ling Huang, Jia Ching Wang</i> | |
| Geometric Matrix Completion with Collaborative Routing Between Capsules..... | 11681 |
| <i>Xuan Li, Li Zhang</i> | |
| Distinguishable Speaker Anonymization Based on Formant and Fundamental Frequency Scaling | 11686 |
| <i>Jixun Yao, Qing Wang, Yi Lei, Pengcheng Guo, Lei Xie, Namin Wang, Jie Liu</i> | |
| One-Shot Action Detection Via Attention Zooming in..... | 11691 |
| <i>He-Yen Hsieh, Ding-Jie Chen, Cheng-Wei Chang, Tyng-Luh Liu</i> | |
| Toward Universal Text-To-Music Retrieval | 11696 |
| <i>Seungheon Doh, Minz Won, Keunwoo Choi, Juhan Nam</i> | |
| A Hierarchical Regression Chain Framework for Affective Vocal Burst Recognition | 11701 |
| <i>Jinchao Li, Xixin Wu, Kaitao Song, Dongsheng Li, Xunying Liu, Helen Meng</i> | |

| | |
|---|-------|
| Lightweight Annotation and Class Weight Training for Automatic Estimation of Alarm Audibility in Noise | 11706 |
| <i>François Effa, Romain Serizel, Jean-Pierre Arz, Nicolas Grimault</i> | |
| Inv-Senet: Invariant Self Expression Network for Clustering Under Biased Data..... | 11711 |
| <i>Ashutosh Singh, Ashish Singh, Aria Masoomi, Tales Imbiriba, Erik Learned-Miller, Deniz Erdogmus</i> | |
| A Processing Framework to Access Large Quantities of Whispered Speech Found in ASMR..... | 11716 |
| <i>Pablo Pérez Zarazaga, Gustav Eje Henter, Zofia Malisz</i> | |
| Textless Speech-To-Music Retrieval Using Emotion Similarity | 11721 |
| <i>Seungheon Doh, Minz Won, Keunwoo Choi, Juhan Nam</i> | |
| Sequential Invariant Information Bottleneck..... | 11726 |
| <i>Yichen Zhang, Shujian Yu, Badong Chen</i> | |
| Adaptive Semantic Fusion Framework for Unsupervised Monocular Depth Estimation..... | 11731 |
| <i>Ruoqi Li, Huimin Yu, Kaiyang Du, Zhuoling Xiao, Bo Yan, Zhengxi Yuan</i> | |
| FFEDCL: Fair Federated Learning with Contrastive Learning..... | 11736 |
| <i>Xiaorong Shi, Liping Yi, Xiaoguang Liu, Gang Wang</i> | |
| WITT: A Wireless Image Transmission Transformer for Semantic Communications..... | 11741 |
| <i>Ke Yang, Sixian Wang, Jincheng Dai, Kailin Tan, Kai Niu, Ping Zhang</i> | |
| Delay-Penalized Transducer for Low-Latency Streaming ASR | 11746 |
| <i>Wei Kang, Zengwei Yao, Fangjun Kuang, Liyong Guo, Xiaoyu Yang, Long Lin, Piotr Zelasko, Daniel Povey</i> | |
| SIGVIC: Spatial Importance Guided Variable-Rate Image Compression | 11751 |
| <i>Jiaming Liang, Meiqin Liu, Chao Yao, Chunyu Lin, Yao Zhao</i> | |
| A Simple Scheme for Coupled Factorization for Hyperspectral Super-Resolution: Exploiting Sparsity in an Easy Way | 11756 |
| <i>Yuening Li, Wing-Kin Ma, Ruiyuan Wu, Huikang Liu</i> | |
| The Multimodal Information Based Speech Processing (Misp) 2022 Challenge: Audio-Visual Diarization and Recognition..... | 11761 |
| <i>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</i> | |
| A Model-Based Hearing Compensation Method Using a Self-Supervised Framework..... | 11766 |
| <i>Yadong Niu, Nan Li, Xihong Wu, Jing Chen</i> | |
| Multitrack Music Transcription with a Time-Frequency Perceiver..... | 11771 |
| <i>Wei -Tsung Lu, Ju-Chiang Wang, Yun -Ning Hung</i> | |
| Singing Voice Synthesis Based on a Musical Note Position-Aware Attention Mechanism..... | 11776 |
| <i>Yukiya Hono, Kei Hashimoto, Yoshihiko Nankaku, Keiichi Tokuda</i> | |
| Relevance Propagation Through Deep Conditional Random Fields | 11781 |
| <i>Xiangyu Yang, Boris Joukovsky, Nikos Deligiannis</i> | |
| Boosting Transferability of Adversarial Example Via an Enhanced Euler's Method..... | 11786 |
| <i>Anjie Peng, Zhi Lin, Hui Zeng, Wenxin Yu, Xiangui Kang</i> | |

| | |
|--|-------|
| ScaleMix: Intra- And Inter-Layer Multiscale Feature Combination for Change Detection..... | 11791 |
| <i>Rui Huang, Qingyi Zhao, Ruofei Wang, Caihua Liu, Sihua Gao, Yuxiang Zhang, Wei Fan</i> | |
| Synthesizing Speech from ECoG with a Combination of Transformer-Based Encoder and Neural Vocoder..... | 11796 |
| <i>Kai Shigemi, Shuji Komeiji, Takumi Mitsuhashi, Yasushi Iimura, Hiroharu Suzuki, Hidenori Sugano, Koichi Shinoda, Kohei Yatabe, Toshihisa Tanaka</i> | |
| AV-TAD: Audio-Visual Temporal Action Detection with Transformer..... | 11801 |
| <i>Yangcheng Li, Zefang Yu, Suncheng Xiang, Ting Liu, Yuzhuo Fu</i> | |
| Semantic-Preserving Augmentation for Robust Image-Text Retrieval..... | 11806 |
| <i>Sunwoo Kim, Kyuhong Shim, Luong Trung Nguyen, Byonghyo Shim</i> | |
| Enhancing the Vocal Range of Single-Speaker Singing Voice Synthesis with Melody-Unsupervised Pre-Training..... | 11811 |
| <i>Shaohuan Zhou, Xu Li, Zhiyong Wu, Ying Shan, Helen Meng</i> | |
| Embedding a Differentiable Mel-Cepstral Synthesis Filter to a Neural Speech Synthesis System..... | 11816 |
| <i>Takenori Yoshimura, Shinji Takaki, Kazuhiro Nakamura, Keiichiro Oura, Yukiya Hono, Kei Hashimoto, Yoshihiko Nankaku, Keiichi Tokuda</i> | |
| Boosting Semi-Supervised Federated Learning with Model Personalization and Client-Variance-Reduction | 11821 |
| <i>Shuai Wang, Yanqing Xu, Yanli Yuan, Xiuhua Wang, Tony Q. S. Quek</i> | |
| SemGeo: Semantic Keywords for Cross-View Image Geo-Localization | 11826 |
| <i>Royston Rodrigues, Masahiro Tani</i> | |
| Modelling Black-Box Audio Effects with Time-Varying Feature Modulation..... | 11831 |
| <i>Marco Comunità, Christian J. Steinmetz, Huy Phan, Joshua D. Reiss</i> | |
| Symbol-Level Precoding is Related to Parameter Estimation from Quantized Data | 11836 |
| <i>Mingjie Shao, Wing-Kin Ma, Yatao Liu</i> | |
| Runtime Prediction of Machine Learning Algorithms in Automl Systems | 11841 |
| <i>Parijat Dube, Theodoros Salonidis, Parikshit Ram, Ashish Verma</i> | |
| Pondering About Task Spatial Misalignment: Classification-Localization Equilibrated Object Detection | 11846 |
| <i>Yudong Zhang, Wei Lu, Xu Wang, Pengkun Wang, Yang Wang</i> | |
| Efficient Speech Quality Assessment Using Self-Supervised Framewise Embeddings | 11851 |
| <i>Karl El Hajal, Zihan Wu, Neil Scheidwasser-Clow, Gasser Elbanna, Milos Cernak</i> | |
| History, Present and Future: Enhancing Dialogue Generation with Few-Shot History-Future Prompt | 11856 |
| <i>Yihe Wang, Yitong Li, Yasheng Wang, Fei Mi, Pingyi Zhou, Jin Liu, Xin Jiang, Qun Liu</i> | |
| Audio Quality Assessment of Vinyl Music Collections Using Self-Supervised Learning | 11861 |
| <i>Alessandro Ragano, Emmanouil Benetos, Andrew Hines</i> | |
| Massively Multilingual ASR on 70 Languages: Tokenization, Architecture, and Generalization Capabilities..... | 11866 |
| <i>Andros Tjandra, Nayan Singhal, David Zhang, Ozlem Kalinli, Abdelrahman Mohamed, Duc Le, Michael L. Seltzer</i> | |

| | |
|---|-------|
| Free-View Expressive Talking Head Video Editing | 11871 |
| <i>Yuantian Huang, Satoshi Iizuka, Kazuhiro Fukui</i> | |
| Wav2vec-Based Detection and Severity Level Classification of Dysarthria from Speech..... | 11876 |
| <i>Farhad Javanmardi, Saska Tirronen, Manila Kodali, Sudarsana Reddy Kadiri, Paavo Alku</i> | |
| Graph Based Semantic Ensemble of Riemannian Neural Structured Learning for BCI-EEG Signal Classification..... | 11881 |
| <i>Vinay Gupta, Laxmidhar Behera, Tushar Sandhan</i> | |
| Measuring the Transferability of ℓ_∞ Attacks by the ℓ_2 Norm..... | 11886 |
| <i>Sizhe Chen, Qinghua Tao, Zhixing Ye, Xiaolin Huang</i> | |
| An Efficient Relay Selection Scheme for Relay-Assisted HARQ..... | 11891 |
| <i>Weihang Ding, Mohammad Shikh-Bahaei</i> | |
| Towards Privacy and Utility in Tourette TIC Detection Through Pretraining Based on Publicly Available Video Data of Healthy Subjects..... | 11896 |
| <i>Nele Sophie Brügge, Esfandiar Mohammadi, Alexander Münchau, Tobias Bäumer, Christian Frings, Christian Beste, Veit Roessner, Heinz Handels</i> | |
| Enhanced Low-Resolution LiDAR-Camera Calibration Via Depth Interpolation and Supervised Contrastive Learning | 11901 |
| <i>Zhikang Zhang, Zifan Yu, Suyu You, Raghuvver Rao, Sanjeev Agarwal, Fengbo Ren</i> | |
| Nonnegative Block-Term Decomposition with the β -Divergence: Joint Data Fusion and Blind Spectral Unmixing..... | 11906 |
| <i>C. Prévost, V. Leplat</i> | |
| High-Level Feature Fusion Network for Session-Based Social Recommendation | 11911 |
| <i>Liuyin Wang, Mingchao Li, Hai-Tao Zheng</i> | |
| Exploiting CCTV Cameras for Hand Hygiene Recognition in ICU..... | 11916 |
| <i>WeiJun Huang, Jia Huang, Guowei Wang, Hongzhou Lu, Min He, Wenjin Wang</i> | |
| A Database for Multi-Modal Short Video Quality Assessment | 11921 |
| <i>Yukun Zhang, Chuan Wang, Sanyi Zhang, Xiaochun Cao</i> | |
| Sequence-Based Device-Free Gesture Recognition Framework for Multi-Channel Acoustic Signals | 11926 |
| <i>Zhizheng Yang, Xun Wang, Dongyu Xia, Wei Wang, Haipeng Dai</i> | |
| Detecting Out-Of-Distribution Examples Via Class-Conditional Impressions Reappearing | 11931 |
| <i>Jinggang Chen, Xiaoyang Qu, Junjie Li, Jianzong Wang, Jiguang Wan, Jing Xiao</i> | |
| Rethinking Rule-Based Approaches in Session-Based Recommendation..... | 11936 |
| <i>Liuyin Wang, Mingchao Li, Hai-Tao Zheng</i> | |
| MLCGAN: Multi-Lead ECG Synthesis with Multi Label Conditional Generative Adversarial Network..... | 11941 |
| <i>Jian Wu, Liping Wang, Hailin Pan, Binyu Wang</i> | |
| Multilingual Word Error Rate Estimation: E-Wer3 | 11946 |
| <i>Shammur Absar Chowdhury, Ahmed Ali</i> | |
| Scalable and Secure Federated XGBoost | 11951 |
| <i>Quang Minh Nguyen, Nhan Khanh Le, Lam M. Nguyen</i> | |

| | |
|--|-------|
| A DNN Based Normalized Time-Frequency Weighted Criterion for Robust Wideband DoA Estimation..... | 11956 |
| <i>Kuan-Lin Chen, Ching-Hua Lee, Bhaskar D. Rao, Harinath Garudadri</i> | |
| DPP-Based Client Selection for Federated Learning with NON-IID DATA..... | 11961 |
| <i>Yuxuan Zhang, Chao Xu, Howard H. Yang, Xijun Wang, Tony Q. S. Quek</i> | |
| Robust Acoustic and Semantic Contextual Biasing in Neural Transducers for Speech Recognition | 11966 |
| <i>Xuandi Fu, Kanthashree Mysore Sathyendra, Ankur Gandhe, Jing Liu, Grant P. Strimel, Ross McGowan, Athanasios Mouchtaris</i> | |
| Memory-Augmented Contrastive Learning for Talking Head Generation | 11971 |
| <i>Jianrong Wang, Yaxin Zhao, Hongkai Fan, Tianyi Xu, Qi Li, Sen Li, Li Liu</i> | |
| Extreme Audio Time Stretching Using Neural Synthesis..... | 11976 |
| <i>Leonardo Fierro, Alec Wright, Vesa Välimäki, Matti Hämäläinen</i> | |
| Group Personalized Federated Learning | 11981 |
| <i>Zhe Liu, Yue Hui, Fuchun Peng</i> | |
| Guide and Select: A Transformer-Based Multimodal Fusion Method for Points of Interest Description Generation..... | 11986 |
| <i>Hanqing Liu, Wei Wang, Niu Hu, Hai-Tao Zheng, Rui Xie, Wei Wu, Yang Bai</i> | |
| Adaptive CSI Feedback with Hidden Semantic Information Transfer | 11991 |
| <i>Jiaqi Cao, Lixiang Lian, Yijie Mao, Bruno Clerckx</i> | |
| Deep Probabilistic Model for Lossless Scalable Point Cloud Attribute Compression | 11996 |
| <i>Dat Thanh Nguyen, Kamal Gopikrishnan Nambiar, André Kaup</i> | |
| Improved Wordpfcg for Passwords with Maximum Probability Segmentation | 12001 |
| <i>Wenting Li, Jiahong Yang, Haibo Cheng, Ping Wang, Kaitai Liang</i> | |
| The Uniqueness Problem of Physical Law Learning..... | 12006 |
| <i>Philipp Scholl, Aras Bacho, Holger Boche, Gitta Kutyniok</i> | |
| Visual Information Matters for ASR Error Correction | 12011 |
| <i>Vanya Bannihatti Kumar, Shanbo Cheng, Ningxin Peng, Yuchen Zhang</i> | |
| SyncNet: Correlating Objective for Time Delay Estimation in Audio Signals..... | 12016 |
| <i>Akshay Raina, Vipul Arora</i> | |
| Multiple Target Measurements: Bayesian Framework for Moving Object Detection in Mimo Radar..... | 12021 |
| <i>Bastian Eisele, Ali Beryhi, Ralf Müller</i> | |
| FindAdaptNet: Find and Insert Adapters by Learned Layer Importance..... | 12026 |
| <i>Junwei Huang, Karthik Ganesan, Soumi Maiti, Young Min Kim, Xuankai Chang, Paul Liang, Shinji Watanabe</i> | |
| Continuous Action Space-Based Spoken Language Acquisition Agent Using Residual Sentence Embedding and Transformer Decoder..... | 12031 |
| <i>Ryota Komatsu, Yusuke Kimura, Takuma Okamoto, Takahiro Shinozaki</i> | |
| On the Detection of Synthetic Images Generated by Diffusion Models..... | 12036 |
| <i>Riccardo Corvi, Davide Cozzolino, Giada Zingarini, Giovanni Poggi, Koki Nagano, Luisa Verdoliva</i> | |

| | |
|--|-------|
| Automatic Classification of Vocal Intensity Category from Speech | 12041 |
| <i>Manila Kodali, Sudarsana Reddy Kadiri, Laura Laaksonen, Paavo Alku</i> | |
| A Sentiment and Syntactic-Aware Graph Convolutional Network for Aspect-Level Sentiment Classification | 12046 |
| <i>Yuxin Yang, Xia Sun, Qiang Lu, Richard Sutcliffe, Jun Feng</i> | |
| Effectiveness of Text, Acoustic, and Lattice-Based Representations in Spoken Language Understanding Tasks..... | 12051 |
| <i>Esaú Villatoro-Tello, Srikanth Madikeri, Juan Zuluaga-Gomez, Bidisha Sharma, Seyyed Saeed Sarfjoo, Iuliia Nigmatulina, Petr Motlicek, Alexei V. Ivanov, Aravind Ganapathiraju</i> | |
| Retrieval-Based Natural 3D Human Motion Generation..... | N/A |
| <i>Zehan Tan, Weidong Yang, Shuai Wu</i> | |
| Enhancing the Efficiency of WMMSE and FP for Beamforming by Minorization-Maximization | 12061 |
| <i>Zepeng Zhang, Ziping Zhao, Kaiming Shen</i> | |
| Efficient Learning of Balanced Signature Graphs | 12066 |
| <i>Gerald Matz, Claudio Verardo, Thomas Dittrich</i> | |
| Abusive Activity Detection with Multi-Modality Based on Convolutional Neural Network..... | 12071 |
| <i>Jisoo Kim, Hyebin Ahn, Byoungyun Yoo</i> | |
| Lyapunov-Driven Deep Reinforcement Learning for Edge Inference Empowered by Reconfigurable Intelligent Surfaces | 12076 |
| <i>Kyriakos Stylianopoulos, Mattia Merluzzi, Paolo Di Lorenzo, George C. Alexandropoulos</i> | |
| A Proximal Approach to IVA-G with Convergence Guarantees..... | 12081 |
| <i>Clément Cosserat, Ben Gabrielson, Emilie Chouzenoux, Jean-Christophe Pesquet, Tülay Adalı</i> | |
| Low-Dose CT Reconstruction Via Optimization-Inspired GAN..... | 12086 |
| <i>Jiawei Jiang, Yuchao Feng, Honghui Xu, Jianwei Zheng</i> | |
| SPASHT: Semantic and Pragmatic Speech Features for Automatic Assessment of Autism..... | 12091 |
| <i>Ashwini B, Vrinda Narayan, Jainendra Shukla</i> | |
| Effectiveness of Inter- And Intra-Subarray Spatial Features for Acoustic Scene Classification..... | 12096 |
| <i>Takao Kawamura, Yuma Kinoshita, Nobutaka Ono, Robin Scheibler</i> | |
| Estimating Shapley Values of Training Utterances for Automatic Speech Recognition Models..... | 12101 |
| <i>Ali Raza Syed, Michael I. Mandel</i> | |
| Practice of the Conformer Enhanced Audio-Visual Hubert on Mandarin and English | 12106 |
| <i>Xiaoming Ren, Chao Li, Shenjian Wang, Biao Li</i> | |
| Piecewise Position Encoding in Convolutional Neural Network for Cough-Based Covid-19 Detection | 12111 |
| <i>Jiakun Shen, Xueshuai Zhang, Pengyuan Zhang, Yonghong Yan, Shaoxing Zhang, Zhihua Huang, Yanfen Tang, Yu Wang, Fujie Zhang, Aijun Sun</i> | |
| Transplayer: Timbre Style Transfer with Flexible Timbre Control | 12116 |
| <i>Yuxuan Wu, Yifan He, Xinlu Liu, Yi Wang, Roger B. Dannenberg</i> | |

| | |
|---|-------|
| Robust M-Estimation Based Distributed Expectation Maximization Algorithm with Robust Aggregation | 12121 |
| <i>Christian A. Schroth, Stefan Vlaski, Abdelhak M. Zoubir</i> | |
| Improving Transformer-Based Networks with Locality for Automatic Speaker Verification | 12126 |
| <i>Mufan Sang, Yong Zhao, Gang Liu, John H. L. Hansen, Jian Wu</i> | |
| Causal Discovery and Causal Inference Based Counterfactual Fairness in Machine Learning | 12131 |
| <i>Yajing Wang, Zongwei Luo</i> | |
| Learning Silhouettes with Group Sparse Autoencoders | 12136 |
| <i>Emmanouil Theodosis, Demba Ba</i> | |
| Progressive Refinement Learning Based on Feature Cross Perception for Residential Areas Semantic Segmentation | 12141 |
| <i>Xinran Lyu, Libao Zhang</i> | |
| HAG: Hierarchical Attention with Graph Network for Dialogue Act Classification in Conversation | 12146 |
| <i>Changzeng Fu, Zhenghan Chen, Jiaqi Shi, Bowen Wu, Chaoran Liu, Carlos Toshinori Ishi, Hiroshi Ishiguro</i> | |
| Neural Fourier Shift for Binaural Speech Rendering | 12151 |
| <i>Jin Woo Lee, Kyogu Lee</i> | |
| CAN2V: Can-Bus Data-Based Seq2seq Model for Vehicle Velocity Prediction | 12156 |
| <i>Jae-Heung Cho, Joon-Hyuk Chang</i> | |
| Relative Dynamic Time Warping Comparison for Pronunciation Errors | 12161 |
| <i>Caitlin Richter, Jón Guðnason</i> | |
| Unsupervised Out-Of-Distribution Detection Using Few in-Distribution Samples | 12166 |
| <i>Chandan Gautam, Aditya Kane, Savitha Ramasamy, Suresh Sundaram</i> | |
| G2CNN: Geometric Prior Based GCNN for Single-View 3D Reconstruction with Loop Subdivision | 12171 |
| <i>Kun Cao, Na Qi, Wei Xu, Qing Zhu, Shibo Xu, Changxin Pan</i> | |
| Towards Building Text-To-Speech Systems for the Next Billion Users | 12176 |
| <i>Gokul Karthik Kumar, Praveen S V, Pratyush Kumar, Mitesh M. Khapra, Karthik Nandakumar</i> | |
| Data Leakage in Cross-Modal Retrieval Training: A Case Study | 12181 |
| <i>Benno Weck, Xavier Serra</i> | |
| Self Supervised Bert for Legal Text Classification | 12186 |
| <i>Arghya Pal, Sailaja Rajanala, Raphaël C.-W. Phan, Koksheik Wong</i> | |
| Egocentric Action Anticipation for Personal Health | 12191 |
| <i>Ivan Rodin, Antonino Furnari, Dimitrios Mavroeidis, Giovanni Maria Farinella</i> | |
| Global Localisation in Continuous Magnetic Vector Fields Using Gaussian Processes | 12196 |
| <i>William McDonald, Cedric Le Gentil, Teresa Vidal-Calleja</i> | |
| I3D: Transformer Architectures with Input-Dependent Dynamic Depth for Speech Recognition | 12201 |
| <i>Yifan Peng, Jaesong Lee, Shinji Watanabe</i> | |

| | |
|--|-------|
| SSGD: A Smartphone Screen Glass Dataset for Defect Detection..... | 12206 |
| <i>Haonan Han, Rui Yang, Shuyan Li, Runze Hu, Xiu Li</i> | |
| Joint Cryo-ET Alignment and Reconstruction with Neural Deformation Fields..... | 12211 |
| <i>Valentin Debarnot, Sidharth Gupta, Konik Kothari, Ivan Dokmanic</i> | |
| Finer-Grained Decomposition for Parallel Quantum Mimo Processing..... | 12216 |
| <i>Minsung Kim, Kyle Jamieson</i> | |
| In-Sensor & Neuromorphic Computing Are All You Need for Energy Efficient Computer Vision | 12221 |
| <i>Gourav Datta, Zeyu Liu, Md Abdullah-Al Kaiser, Souvik Kundu, Joe Mathai, Zihan Yin, Ajey P. Jacob, Akhilesh R. Jaiswal, Peter A. Beerel</i> | |
| Aiding Speech Harmonic Recovery in DNN-Based Single Channel Noise Reduction Using Cepstral Excitation Manipulation (CEM) Components..... | 12226 |
| <i>Yanjue Song, Nilesh Madhu</i> | |
| Noise-Aware Target Extension with Self-Distillation for Robust Speech Recognition | 12231 |
| <i>Ju-Seok Seong, Jeong-Hwan Choi, Jehyun Kyung, Ye-Rin Jeoung, Joon-Hyuk Chang</i> | |
| A Quantum Approach for Stochastic Constrained Binary Optimization | 12236 |
| <i>Sarthak Gupta, Vassilis Kekatos</i> | |
| SVMV: Spatiotemporal Variance-Supervised Motion Volume for Video Frame Interpolation | 12241 |
| <i>Yao Luo, Jinshan Pan, Jinhui Tang</i> | |
| LED: Label Correlation Enhanced Decoder for Multi-Label Text Classification | 12246 |
| <i>Kefan Ma, Zheng Huang, Xinrui Deng, Jie Guo, Weidong Qiu</i> | |
| Unlimited Sampling of FRI Signals Independent of Sampling Rate | 12251 |
| <i>Ruiming Guo, Ayush Bhandari</i> | |
| New Interpretable Patterns and Discriminative Features from Brain Functional Network Connectivity Using Dictionary Learning..... | 12256 |
| <i>F. Ghayem, H. Yang, F. Kantar, S.-J. Kim, V. D. Calhoun, T. Adali</i> | |
| MRML: Multimodal Rumor Detection by Deep Metric Learning | 12261 |
| <i>Liwen Peng, Songlei Jian, Dongsheng Li, Siqu Shen</i> | |
| DiffVoice: Text-To-Speech with Latent Diffusion..... | 12266 |
| <i>Zhijun Liu, Yiwei Guo, Kai Yu</i> | |
| Improved Appliance Transient Feature Extraction Via Template Matching | 12271 |
| <i>Bo Liu, Fenglei Chang, Wenpeng Luan, Bochao Zhao</i> | |
| Asymptotic Distribution of Stochastic Mirror Descent Iterates in Average Ensemble Models | 12276 |
| <i>Taylan Kargin, Fariborz Salehi, Babak Hassibi</i> | |
| DSPGAN: A Gan-Based Universal Vocoder for High-Fidelity TTS by Time-Frequency Domain Supervision from DSP..... | 12281 |
| <i>Kun Song, Yongmao Zhang, Yi Lei, Jian Cong, Hanzhao Li, Lei Xie, Gang He, Jinfeng Bai</i> | |
| Efficient Feature Fusion for Learning-Based Photometric Stereo | 12286 |
| <i>Yakun Ju, Kin-Man Lam, Jun Xiao, Cong Zhang, Cuixin Yang, Junyu Dong</i> | |
| Good Neighbors Are All You Need for Chinese Grapheme-To-Phoneme Conversion..... | 12291 |
| <i>Jungjun Kim, Changjin Han, Gyuhyeon Nam, Gyeongsu Chae</i> | |

| | |
|--|-------|
| SPECTRANET-SO(3): Learning Satellite Orientation from Optical Spectra by Implicitly Modeling Mutually Exclusive Probability Distributions on the Rotation Manifold | 12296 |
| <i>Matthew Phelps, Thomas Swindle, J. Zachary Gazak, Andrew Vandenberg, Justin Fletcher</i> | |
| A Controllable Lifestyle Simulator for Use in Deep Reinforcement Learning Algorithms..... | 12301 |
| <i>Libio Gonçalves Braz, Allmin Susaiyah</i> | |
| Learning to Generate 3D Representations of Building Roofs Using Single-View Aerial Imagery | 12306 |
| <i>Maxim Khomiakov, Alejandro Valverde Mahou, Alba Reinders Sánchez, Jes Frellsen, Michael Riis Andersen</i> | |
| Building Keyword Search System from End-To-End Asr Systems..... | 12311 |
| <i>Ruizhe Huang, Matthew Wiesner, Leibny Paola Garcia-Perera, Dan Povey, Jan Trmal, Sanjeev Khudanpur</i> | |
| LMBAO: A Landmark Map for Bundle Adjustment Odometry in LiDAR SLAM..... | 12316 |
| <i>Letian Zhang, Jinping Wang, Lu Jie, Nanjie Chen, Xiaojun Tan, Zhifei Duan</i> | |
| Multi-Observation Hidden Semi-Markov Model for Photoplethysmogram Signal Semantic Segmentation | 12321 |
| <i>Navid Hasanzadeh, Shahrokh Valaee, Hojjat Salehinejad</i> | |
| Data2vec-Aqc: Search for the Right Teaching Assistant in the Teacher-Student Training Setup..... | 12326 |
| <i>Vasista Sai Lodagala, Sreyan Ghosh, S. Umesh</i> | |
| Exploiting Interactivity and Heterogeneity for Sleep Stage Classification Via Heterogeneous Graph Neural Network | 12331 |
| <i>Ziyu Jia, Youfang Lin, Yuhan Zhou, Xiyang Cai, Peng Zheng, Qiang Li, Jing Wang</i> | |
| Benchmark of Physiological Model Based and Deep Learning Based Remote Photoplethysmography in Automotive Applications | 12336 |
| <i>Zhiyu Wang, Xuezhi Yang, Hongzhou Lu, Caifeng Shan, Wenjin Wang</i> | |
| Learning Gradients of Convex Functions with Monotone Gradient Networks | 12341 |
| <i>Shreyas Chaudhari, Srinivasa Pranav, José M. F. Moura</i> | |
| Spatial Graph Signal Interpolation with an Application for Merging BCI Datasets with Various Dimensionalities | 12346 |
| <i>Yassine El Ouahidi, Lucas Drumetz, Giulia Lioi, Nicolas Farrugia, Bastien Padeloup, Vincent Gripon</i> | |
| Phoneix: Acoustic Feature Processing Strategy for Enhanced Singing Pronunciation with Phoneme Distribution Predictor | 12351 |
| <i>Yuning Wu, Jiatong Shi, Tao Qian, Dongji Gao, Qin Jin</i> | |
| Codebook-Based User Tracking in IRS-Assisted mmWave Communication Networks..... | 12356 |
| <i>Moritz Garkisch, Vahid Jamali, Robert Schober</i> | |
| Efficient Large-Scale Multi-Unimodular Waveform Design with Good Correlation Properties Via Direct Phase Optimizations | 12361 |
| <i>Xiaohan Zhao, Yongzhe Li, Ran Tao</i> | |
| Diverse and Vivid Sound Generation from Text Descriptions..... | 12366 |
| <i>Guangwei Li, Xuenan Xu, Lingfeng Dai, Mengyue Wu, Kai Yu</i> | |
| Model-Based Spectral Reconstruction of Interferometric Acquisitions | 12371 |
| <i>Mohamad Jouni, Daniele Picone, Mauro Dalla Mura</i> | |

| | |
|---|-------|
| Incorporating Visual Information Reconstruction into Progressive Learning for Optimizing Audio-Visual Speech Enhancement..... | 12376 |
| <i>Chen-Yue Zhang, Hang Chen, Jun Du, Bao-Cai Yin, Jia Pan, Chin-Hui Lee</i> | |
| Improving the Modality Representation with Multi-View Contrastive Learning for Multimodal Sentiment Analysis..... | 12381 |
| <i>Peipei Liu, Xin Zheng, Hong Li, Jie Liu, Yimo Ren, Hongsong Zhu, Limin Sun</i> | |
| Generative Model Based Highly Efficient Semantic Communication Approach for Image Transmission..... | 12386 |
| <i>Tianxiao Han, Jiancheng Tang, Qianqian Yang, Yiping Duan, Zhaoyang Zhang, Zhiguo Shi</i> | |
| Image Source Method Based on the Directional Impulse Responses..... | 12391 |
| <i>Jiarui Wang, Prasanga Samarasinghe, Thushara Abhayapala, Jihui Aimee Zhang</i> | |
| Music Rearrangement Using Hierarchical Segmentation..... | 12396 |
| <i>Christos Plachouras, Marius Miron</i> | |
| Graphit: Iterative Reweighted ℓ_1 Algorithm for Sparse Graph Inference in State-Space Models | 12401 |
| <i>Émilie Chouzenoux, Victor Elvira</i> | |
| Modulo EEG Signal Recovery Using Transformer | 12406 |
| <i>Tianyu Geng, Feng Ji, Pratibha, Wee Peng Tay</i> | |
| Latent Iterative Refinement for Modular Source Separation..... | 12411 |
| <i>Dimitrios Bralios, Efthymios Tzinis, Gordon Wichern, Paris Smaragdis, Jonathan Le Roux</i> | |
| DisCoHead: Audio-And-Video-Driven Talking Head Generation by Disentangled Control of Head Pose and Facial Expressions..... | 12416 |
| <i>Geumbyeol Hwang, Sunwon Hong, Seunghyun Lee, Sungwoo Park, Gyeongsu Chae</i> | |
| MvCo-DoT: Multi-View Contrastive Domain Transfer Network for Medical Report Generation | 12421 |
| <i>Ruizhi Wang, Xiangtao Wang, Zhenghua Xu, Wenting Xu, Junyang Chen, Thomas Lukasiewicz</i> | |
| Motor Activity Recognition Using Eeg Data and Ensemble of Stacked BLSTM-LSTM Network and Transformer Model..... | 12426 |
| <i>Pallavi Kaushik, Ilina Tripathi, Partha Pratim Roy</i> | |
| G2PL: Lexicon Enhanced Chinese Polyphone Disambiguation Using Bert Adapter with a New Dataset..... | 12431 |
| <i>Haifeng Zhao, Hongzhi Wan, Lili Huang, Mingwei Cao</i> | |
| SSVMR: Saliency-Based Self-Training for Video-Music Retrieval..... | 12436 |
| <i>Xuxin Cheng, Zhihong Zhu, Hongxiang Li, Yaowei Li, Yuexian Zou</i> | |
| M3ST: Mix at Three Levels for Speech Translation..... | 12441 |
| <i>Xuxin Cheng, Qianqian Dong, Fengpeng Yue, Tom Ko, Mingxuan Wang, Yuexian Zou</i> | |
| On Weighted Cross-Entropy for Label-Imbalanced Separable Data: An Algorithmic-Stability Study | 12446 |
| <i>Puneesh Deora, Christos Thrampoulidis</i> | |
| Node-Wise Domain Adaptation Based on Transferable Attention for Recognizing Road Rage Via EEG | 12451 |
| <i>Xueqi Gao, Chao Xu, Yihang Song, Jing Hu, Jian Xiao, Zhaopeng Meng</i> | |
| Differentiable Adaptive Short-Time Fourier Transform with Respect to the Window Length..... | 12456 |
| <i>Maxime Leiber, Yosra Marnissi, Axel Barrau, Mohammed El Badaoui</i> | |

| | |
|---|-------|
| Retiformer: Retinex-Based Enhancement in Transformer for Low-Light Image..... | 12461 |
| <i>Junxiang Ruan, Xiangtao Kong, Wenqi Huang, Wenming Yang</i> | |
| Learning Supervised Covariation Projection Through General Covariance..... | 12466 |
| <i>Xiangze Bao, Yun-Hao Yuan, Yun Li, Jipeng Qiang, Yi Zhu</i> | |
| A Contrastive Embedding-Based Domain Adaptation Method for Lung Sound Recognition in Children Community-Acquired Pneumonia..... | 12471 |
| <i>Dongmin Huang, Lingwei Wang, Hongzhou Lu, Wenjin Wang</i> | |
| On the Quantization of Recurrent Neural Networks for Smiles Generation | 12476 |
| <i>Adriano Durao, Joel P. Arrais, Bernardete Ribeiro, Gabriel Falcao</i> | |
| MFCCGAN: A Novel MFCC-Based Speech Synthesizer Using Adversarial Learning..... | 12481 |
| <i>Mohammad Reza Hasanabadi, Majid Behdad, Davood Gharavian</i> | |
| Passive Detection of Rank-One Gaussian Signals for Known Channel Subspaces and Arbitrary Noise..... | 12486 |
| <i>D. Ramirez, I. Santamaria, L. L. Scharf</i> | |
| Exploration into Translation-Equivariant Image Quantization..... | 12491 |
| <i>Woncheol Shin, Gyubok Lee, Jiyoung Lee, Eunyi Lyou, Joonseok Lee, Edward Choi</i> | |
| Towards Controllable Audio Texture Morphing..... | 12496 |
| <i>Chitralkha Gupta, Purnima Kamath, Yize Wei, Zhuoyao Li, Suranga Nanayakkara, Lonce Wysse</i> | |
| Self-Supervised Audio-Visual Speech Representations Learning by Multimodal Self-Distillation..... | 12501 |
| <i>Jing-Xuan Zhang, Genshun Wan, Zhen-Hua Ling, Jia Pan, Jianqing Gao, Cong Liu</i> | |
| A Novel Metric for Evaluating Audio Caption Similarity..... | 12506 |
| <i>Swapnil Bhosale, Rupayan Chakraborty, Sunil Kumar Koppurapu</i> | |
| Evaluating Speech-Phoneme Alignment and Its Impact on Neural Text-To-Speech Synthesis..... | 12511 |
| <i>Frank Zalkow, Prachi Govalkar, Meinard Müller, Emanuel A. P. Habets, Christian Dittmar</i> | |
| Continuous Learning for Blind Image Quality Assessment with Contrastive Transformer | 12516 |
| <i>Jifan Yang, Zhongyuan Wang, Baojin Huang, Lianbing Deng</i> | |
| Utilization of Bessel Beams in Wideband Sub Terahertz Communication Systems to Mitigate Beamsplit Effects in the Near-Field..... | 12521 |
| <i>Arjun Singh, Vitaly Petrov, Josep M. Jornet</i> | |
| NBA-OMP: Near-Field Beam-Split-Aware Orthogonal Matching Pursuit for Wideband THz Channel Estimation | 12526 |
| <i>Ahmet M. Elbir, Kumar Vijay Mishra, Symeon Chatzinotas</i> | |
| Composition of Motion from Video Animation Through Learning Local Transformations..... | 12531 |
| <i>Michalis Vrigkas, Virginia Tagka, Marina E. Plissiti, Christophoros Nikou</i> | |
| Delivering Speaking Style in Low-Resource Voice Conversion with Multi-Factor Constraints..... | 12536 |
| <i>Zhichao Wang, Xinsheng Wang, Lei Xie, Yuanzhe Chen, Qiao Tian, Yuping Wang</i> | |
| Privacy-Enhanced Federated Learning Against Attribute Inference Attack for Speech Emotion Recognition | 12541 |
| <i>Huan Zhao, Haijiao Chen, Yufeng Xiao, Zixing Zhang</i> | |

| | |
|---|-------|
| Stochastic Optimization of Vector Quantization Methods in Application to Speech and Image Processing..... | 12546 |
| <i>Mohammad Hassan Vali, Tom Bäckström</i> | |
| Multi-Speaker End-to-End Multi-Modal Speaker Diarization System for the MISP 2022 Challenge | 12551 |
| <i>Tao Liu, Zhengyang Chen, Yanmin Qian, Kai Yu</i> | |
| Improving Speech-To-Speech Translation Through Unlabeled Text..... | 12553 |
| <i>Xuan-Phi Nguyen, Sravya Popuri, Changhan Wang, Yun Tang, Ilia Kulikov, Hongyu Gong</i> | |
| HappyQuokka System for ICASSP 2023 Auditory EEG Challenge | 12558 |
| <i>Zhenyu Piao, Miseul Kim, Hyungchan Yoon, Hong-Goo Kang</i> | |
| High-Speed Drone Detection Based on Yolo-V8 | 12560 |
| <i>Jun-Hwa Kim, Namho Kim, Chee Sun Won</i> | |
| Tensor Completion for Efficient and Accurate Hyperparameter Optimisation in Large-Scale Statistical Learning..... | 12562 |
| <i>Aaman Rebello, Kriton Konstantinidis, Yao Lei Xu, Danilo P. Mandic</i> | |
| The NPU-ASLP System for Audio-Visual Speech Recognition in MISP 2022 Challenge | 12567 |
| <i>Pengcheng Guo, He Wang, Bingshen Mu, Ao Zhang, Peikun Chen</i> | |
| Personalized Speech Enhancement Combining Band-Split RNN and Speaker Attentive Module..... | 12569 |
| <i>Xiaohuai Le, Li Chen, Chao He, Yiqing Guo, Cheng Chen, Xianjun Xia, Jing Lu</i> | |
| Multi-Task Sub-Band Network for Deep Residual Echo Suppression..... | 12571 |
| <i>Jiayao Sun, Dawei Luo, Zhaoxia Li, Jindong Li, Yukai Ju, Yang Li</i> | |
| Dialogue Context Modelling for Action Item Detection: Solution for ICASSP 2023 Mug Challenge Track 5..... | 12573 |
| <i>Jie Huang, Xiachong Feng, Yangfan Ye, Liang Zhao, Xiaocheng Feng, Bing Qin, Ting Liu</i> | |
| A Multi-Stage Low-Latency Enhancement System for Hearing Aids..... | 12575 |
| <i>Chengwei Ouyang, Kexin Fei, Haoshuai Zhou, Congxi Lu, Linkai Li</i> | |
| Relate Auditory Speech to Eeg by Shallow-Deep Attention-Based Network..... | 12577 |
| <i>Fan Cui, Liyong Guo, Lang He, Jiyao Liu, Ercheng Pei, Yujun Wang, Dongmei Jiang</i> | |
| A Progressive Neural Network for Acoustic Echo Cancellation | 12579 |
| <i>Zhuangqi Chen, Xianjun Xia, Siyu Sun, Ziqian Wang, Cheng Chen, Guoliang Xie, Pingjian Zhang, Yijian Xiao</i> | |
| Inplace Cepstral Speech Enhancement System for the ICASSP 2023 Clarity Challenge | 12581 |
| <i>Jinjiang Liu, Xueliang Zhang</i> | |
| Tayloracnet: A Taylor Style Neural Network for Full-Band Echo Cancellation..... | 12583 |
| <i>Weiming Xu, Zhihao Guo</i> | |
| Multi-Head Attention and GRU for Improved Match-Mismatch Classification of Speech Stimulus and EEG Response | 12585 |
| <i>Marvin Borsdorf, Saurav Pahuja, Gabriel Ivucic, Siqi Cai, Haizhou Li, Tanja Schultz</i> | |
| Multi-Speaker Multi-Lingual VQTTS System for LIMMITS 2023 Challenge..... | 12587 |
| <i>Chenpeng Du, Yiwei Guo, Feiyu Shen, Kai Yu</i> | |

| | |
|--|-------|
| Relating EEG Recordings to Speech Using Envelope Tracking and the Speech-FFR | 12589 |
| <i>Mike Thornton, Danilo Mandic, Tobias Reichenbach</i> | |
| S-Feature Pyramid Network and Attention Model for Drone Detection | 12591 |
| <i>Pengcheng Dong, Chuntao Wang, Zhenyong Lu, Kai Zhang, Wenbo Wan, Jiande Sun</i> | |
| E-Branchformer-Based E2E SLU Toward Stop on-Device Challenge..... | 12593 |
| <i>Yosuke Kashiwagi, Siddhant Arora, Hayato Futami, Jessica Huynh, Shih-Lun Wu, Yifan Peng, Brian Yan, Emiru Tsunoo, Shinji Watanabe</i> | |
| W2KPE: Keyphrase Extraction with Word-Word Relation..... | 12595 |
| <i>Wen Cheng, Shichen Dong, Wei Wang</i> | |
| CONSEN: Complementary and Simultaneous Ensemble for Alzheimer’s Disease Detection and MMSE Score Prediction | 12597 |
| <i>Longbin Jin, Yealim Oh, Hyunseo Kim, Hyuntaek Jung, Hyo Jin Jon, Jung Eun Shin, Eun Yi Kim</i> | |
| The WHU-Alibaba Audio-Visual Speaker Diarization System for the MISP 2022 Challenge | 12599 |
| <i>Ming Cheng, Haoxu Wang, Ziteng Wang, Qiang Fu, Ming Li</i> | |
| A Two-Stage System for Spoken Language Understanding..... | 12601 |
| <i>Gaosheng Zhang, Shilei Miao, Linghui Tang, Peijia Qian</i> | |
| A Low-Latency Deep Hierarchical Fusion Network for Fullband Acoustic Echo Cancellation | 12603 |
| <i>Haoran Zhao, Nan Li, Runqiang Han, Xiguang Zheng, Chen Zhang, Liang Guo, Bing Yu</i> | |
| Transductive Matrix Completion with Calibration for Multi-Task Learning | 12605 |
| <i>Hengfang Wang, Yasi Zhang, Xiaojun Mao, Zhonglei Wang</i> | |
| Deep Learning-Based Path Loss Prediction for Outdoor Wireless Communication Systems | 12610 |
| <i>Kehai Qiu, Stefanos Bakirtzis, Hui Song, Ian Wassell, Jie Zhang</i> | |
| Ensemble and Personalized Transformer Models for Subject Identification and Relapse Detection in E-Prevention Challenge..... | 12612 |
| <i>Salvatore Calcagno, Raffaele Mineo, Daniela Giordano, Concetto Spampinato</i> | |
| Speech Signal Improvement Using Causal Generative Diffusion Models | 12614 |
| <i>Julius Richter, Simon Welker, Jean-Marie Lemercier, Bunlong Lay, Tal Peer, Timo Gerkmann</i> | |
| A Transformer-Based E2E SLU Model for Improved Semantic Parsing | 12616 |
| <i>Othman Istaiteh, Yasmeen Kussad, Yahya Daqour, Maria Habib, Mohammad Habash, Dhananjaya Gowda</i> | |
| Half-Temporal and Half-Frequency Attention U2Net for Speech Signal Improvement..... | 12618 |
| <i>Zehua Zhang, Shiyun Xu, Xuyi Zhuang, Yukun Qian, Lianyu Zhou, Mingjiang Wang</i> | |
| Drone-Vs-Bird: Drone Detection Using YOLOv7 with CSRT Tracker | 12620 |
| <i>Sahaj K. Mistry, Shreyas Chatterjee, Ajeet K. Verma, Vinit Jakhetiya, Badri N. Subudhi, Sunil Jaiswal</i> | |
| Decoding Auditory EEG Responses Using an Adapted Wavenet..... | 12622 |
| <i>Bob Van Dyck, Liuyin Yang, Marc M. Van Hulle</i> | |
| Agile Radio Map Prediction Using Deep Learning..... | 12624 |
| <i>Enes Krijestorac, Hazem Sallouha, Shamik Sarkar, Danijela Cabric</i> | |

| | |
|---|-------|
| A Study on the Integration of Pipeline and E2E SLU Systems for Spoken Semantic Parsing Toward Stop Quality Challenge..... | 12626 |
| <i>Siddhant Arora, Hayato Futami, Shih-Lun Wu, Jessica Huynh, Yifan Peng, Yosuke Kashiwagi, Emiru Tsunoo, Brian Yan, Shinji Watanabe</i> | |
| Multi-Channel Speaker Extraction with Adversarial Training: The Wavlab Submission to the Clarity ICASSP 2023 Grand Challenge..... | 12628 |
| <i>Samuele Cornell, Zhong-Qiu Wang, Yoshiki Masuyama, Shinji Watanabe, Manuel Pariente, Nobutaka Ono, Stefano Squartini</i> | |
| PMNet: Large-Scale Channel Prediction System for ICASSP 2023 First Pathloss Radio Map Prediction Challenge | 12630 |
| <i>Ju-Hyung Lee, Joohan Lee, Seon-Ho Lee, Andreas F. Molisch</i> | |
| The Pipeline System of ASR and NLU with MLM-Based Data Augmentation Toward Stop Low-Resource Challenge..... | 12632 |
| <i>Hayato Futami, Jessica Huynh, Siddhant Arora, Shih-Lun Wu, Yosuke Kashiwagi, Yifan Peng, Brian Yan, Emiru Tsunoo, Shinji Watanabe</i> | |
| Exploring Language-Agnostic Speech Representations Using Domain Knowledge for Detecting Alzheimer’s Dementia..... | 12634 |
| <i>Zehra Shah, Shi-Ang Qi, Fei Wang, Mahtab Farrokh, Mashrura Tasnim, Eleni Stroulia, Russell Greiner, Manos Plitsis, Athanasios Katsamanis</i> | |
| The Ustc System for Adress-M Challenge | 12636 |
| <i>Kangdi Mei, Xinyun Ding, Yinlong Liu, Zhiqiang Guo, Feiyang Xu, Xin Li, Tuya Naren, Jiahong Yuan, Zhenhua Ling</i> | |
| Vani: Very-Lightweight Accent-Controllable TTS for Native and Non-Native Speakers with Identity Preservation..... | 12638 |
| <i>Rohan Badlani, Akshit Arora, Subhankar Ghosh, Rafael Valle, Kevin J. Shih, João Felipe Santos, Boris Ginsburg, Bryan Catanzaro</i> | |
| Cross-Lingual Alzheimer’s Disease Detection Based on Paralinguistic and Pre-Trained Features..... | 12640 |
| <i>Xuchu Chen, Yu Pu, Jinpeng Li, Wei-Qiang Zhang</i> | |
| Two-Step Band-Split Neural Network Approach for Full-Band Residual Echo Suppression | 12642 |
| <i>Zihan Zhang, Shimin Zhang, Mingshuai Liu, Yanhong Leng, Zhe Han, Li Chen, Lei Xie</i> | |
| De’hubert: Disentangling Noise in a Self-Supervised Model for Robust Speech Recognition | 12644 |
| <i>Dianwen Ng, Ruixi Zhang, Jia Qi Yip, Zhao Yang, Jinjie Ni, Chong Zhang, Yukun Ma, Chongjia Ni, Eng Siong Chng, Bin Ma</i> | |
| A Person Identification System for the ICASSP 2023 e-Prevention Challenge | 12649 |
| <i>Jinting Wu, Mei Tu</i> | |
| LeanSpeech: The Microsoft Lightweight Speech Synthesis System for Limmits Challenge 2023 | 12651 |
| <i>Chen Zhang, Shubham Bansal, Aakash Lakhera, Jinzhu Li, Gang Wang, Sandeepkumar Satpal, Sheng Zhao, Lei He</i> | |
| The NPU-Elevoc Personalized Speech Enhancement System for Icassp2023 DNS Challenge..... | 12653 |
| <i>Xiaopeng Yan, Yindi Yang, Zhihao Guo, Liangliang Peng, Lei Xie</i> | |
| The XMU System for Audio-Visual Diarization and Recognition in MISP Challenge 2022..... | 12655 |
| <i>Tao Li, Haodong Zhou, Jie Wang, Qingyang Hong, Lin Li</i> | |

| | |
|--|-------|
| A Low-Latency Hybrid Multi-Channel Speech Enhancement System for Hearing Aids..... | 12657 |
| <i>Tong Lei, Zhongshu Hou, Yuxiang Hu, Wanyu Yang, Tianchi Sun, Xiaobin Rong, Dahan Wang, Kai Chen, Jing Lu</i> | |
| TEA-PSE 3.0: Tencent-Ethereal-Audio-Lab Personalized Speech Enhancement System for ICASSP 2023 Dns-Challenge..... | 12659 |
| <i>Yukai Ju, Jun Chen, Shimin Zhang, Shulin He, Wei Rao, Weixin Zhu, Yannan Wang, Tao Yu, Shidong Shang</i> | |
| Real-Time Speech Interruption Analysis: from Cloud to Client Deployment..... | 12661 |
| <i>Quchen Fu, Szu-Wei Fu, Yaran Fan, Yu Wu, Zhuo Chen, Jayant Gupchup, Ross Cutler</i> | |
| TSpeech-AI System Description to the 5th Deep Noise Suppression (DNS) Challenge..... | 12666 |
| <i>Jianwei Yu, Hangting Chen, Yi Luo, Rongzhi Gu, Weihua Li, Chao Weng</i> | |
| SSI-Net: A Multi-Stage Speech Signal Improvement System for ICASSP 2023 SSI Challenge..... | 12668 |
| <i>Weixin Zhu, Zilin Wang, Jiuxin Lin, Chang Zeng, Tao Yu</i> | |
| Post-Trained Language Model Adaptive to Extractive Summarization of Long Spoken Documents..... | 12670 |
| <i>Hyunjong Ok, Seong-Bae Park</i> | |
| Convolutional Recurrent MetriCGAN with Spectral Dimension Compression for Full-Band Speech Enhancement..... | 12672 |
| <i>Zhongshu Hou, Qinwen Hu, Tianchi Sun, Yuxiang Hu, Changbao Zhu, Kai Chen</i> | |
| The Ajmide Topic Segmentation System for the ICASSP 2023 General Meeting Understanding and Generation Challenge..... | 12674 |
| <i>Beibei Hu, Qiang Li, Xianjun Xia</i> | |
| Signal Processing Grand Challenge 2023 – E-Prevention: Sleep Behavior as an Indicator of Relapses in Psychotic Patients..... | 12676 |
| <i>Kleanthis Avramidis, Kranti Adsul, Digbalay Bose, Shrikanth Narayanan</i> | |
| Dual-Path Dilated Convolutional Recurrent Network with Group Attention for Multi-Channel Speech Enhancement..... | 12678 |
| <i>Jiaming Cheng, Cong Pang, Ruiyu Liang, Jingjie Fan, Li Zhao</i> | |
| Lightweight Prosody-TTS for Multi-Lingual Multi-Speaker Scenario..... | 12680 |
| <i>Giridhar Pamisetty, S Chaitanya Varun, K Sri Rama Murty</i> | |
| Two-Stage Neural Network for ICASSP 2023 Speech Signal Improvement Challenge..... | 12682 |
| <i>Mingshuai Liu, Shubo Lv, Zihan Zhang, Runduo Han, Xiang Hao, Xianjun Xia, Li Chen, Yijian Xiao, Lei Xie</i> | |
| Relapse Detection in Patients with Psychotic Disorders Using Unsupervised Learning on Smartwatch Signals..... | 12684 |
| <i>Salam Hamieh, Vincent Heiries, Hussein Al Osman, Christelle Godin</i> | |
| The NIO System for Audio-Visual Diarization and Recognition in MISP Challenge 2022..... | 12686 |
| <i>Gaopeng Xu, Xianliang Wang, Sang Wang, Junfeng Yuan, Wei Guo, Wei Li, Jie Gao</i> | |
| Stream Attention Based U-Net for L3DAS23 Challenge..... | 12688 |
| <i>Honglong Wang, Yanjie Fu, Junjie Li, Meng Ge, Longbiao Wang, Xinyuan Qian</i> | |
| 3D Audio Signal Processing Systems for Speech Enhancement and Sound Localization and Detection..... | 12690 |
| <i>Jisheng Bai, Siwei Huang, Han Yin, Yafei Jia, Mou Wang, Jianfeng Chen</i> | |

| | |
|--|-------|
| The NERCSLIP-USTC System for the L3DAS23 Challenge Task2: 3D Sound Event Localization and Detection (SELD)..... | 12692 |
| <i>Haoyin Yan, Haitao Xu, Qing Wang, Jie Zhang</i> | |
| Autonomous Soundscape Augmentation with Multimodal Fusion of Visual and Participant-Linked Inputs..... | 12694 |
| <i>Kenneth Ooi, Karn N. Watcharasupat, Bhan Lam, Zhen-Ting Ong, Woon-Seng Gan</i> | |
| Cross-Lingual Transfer Learning for Alzheimer’s Detection from Spontaneous Speech..... | 12699 |
| <i>Bastiaan Tamm, Rik Vandenberghe, Hugo Van Hamme</i> | |
| Person Identification with Wearable Sensing Using Missing Feature Encoding and Multi-Stage Modality Fusion | 12701 |
| <i>Payal Mohapatra, Akash Pandey, Sinan Keten, Wei Chen, Qi Zhu</i> | |
| Encoder-Decoder Graph Convolutional Network for Automatic Timed-Up-and-Go and Sit-to-Stand Segmentation..... | 12703 |
| <i>Bo Wen, Chen Du, Truong Q. Nguyen</i> | |
| PhaseAug: A Differentiable Augmentation for Speech Synthesis to Simulate One-To-Many Mapping | 12708 |
| <i>Junhyeok Lee, Seungu Han, Hyunjae Cho, Wonbin Jung</i> | |
| Lightweight Machine Learning for Seizure Detection on Wearable Devices..... | 12713 |
| <i>Baichuan Huang, Azra Abtahi, Amir Aminifar</i> | |
| Pretrained Transformers for Seizure Detection | 12715 |
| <i>Saarang Panchavati, Samuel Vander Dussen, Hemal Semwal, Ahmed Ali, Justin Chen, Haoran Li, Corey Arnold, William Speier</i> | |
| Towards Interpretable Seizure Detection Using Wearables..... | 12717 |
| <i>Irfan Al-Hussaini, Cassie S. Mitchell</i> | |
| Optimization of the Deep Neural Networks for Seizure Detection | 12719 |
| <i>Alexey Shovkun, Andrey Kiryasov, Ilya Zakharov, Mariam Khayretdinova</i> | |
| Learnt Mutual Feature Compression for Machine Vision | 12721 |
| <i>Tie Liu, Mai Xu, Shengxi Li, Chaoran Chen, Li Yang, Zhuoyi Lv</i> | |
| Boosting Person Re-Identification with Viewpoint Contrastive Learning and Adversarial Training | 12726 |
| <i>Xingyue Shi, Hong Liu, Wei Shi, Zihui Zhou, Yidi Li</i> | |
| Graph-Graph Context Dependency Attention for Graph Edit Distance..... | 12731 |
| <i>Ruiqi Jia, Xianbing Feng, Xiaoqing Lyu, Zhi Tang</i> | |
| Improving Sentence Similarity Estimation for Unsupervised Extractive Summarization..... | 12736 |
| <i>Shichao Sun, Ruifeng Yuan, Wenjie Li, Sujian Li</i> | |
| C2BN: Cross-Modality and Cross-Scale Balance Network for Multi-Modal 3D Object Detection | 12741 |
| <i>Bonan Ding, Jin Xie, Jing Nie</i> | |
| An Attention-Based Approach to Hierarchical Multi-Label Music Instrument Classification | 12746 |
| <i>Zhi Zhong, Masato Hirano, Kazuki Shimada, Kazuya Tateishi, Shusuke Takahashi, Yuki Mitsufuji</i> | |
| K ² NN: Self-Supervised Learning with Hierarchical Nearest Neighbors for Remote Sensing | 12751 |
| <i>Jianlong Yuan, Yuanhong Xu, Zhibin Wang</i> | |

| | |
|--|-------|
| Light Field Compression Via Compact Neural Scene Representation | 12756 |
| <i>Jinglei Shi, Christine Guillemot</i> | |
| Logovit: Local-Global Vision Transformer for Object Re-Identification..... | 12761 |
| <i>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</i> | |
| Semantics-Guided Object Removal for Facial Images: with Broad Applicability and Robust Style Preservation | 12766 |
| <i>Jookyung Song, Yeonjin Chang, Seonguk Park, Nojun Kwak</i> | |
| Adaptive Data Augmentation for Contrastive Learning | 12771 |
| <i>Yuhan Zhang, He Zhu, Shan Yu</i> | |
| Bipartite Graph Convolutional Networks with Adversarial Domain Transfer..... | 12776 |
| <i>Dong Wu, Bin Liang, Xiangjun Liu, Xuan Zang, Mingmin Chi</i> | |
| Context-Aware Face Clustering with Graph Convolutional Networks..... | 12781 |
| <i>Dafeng Zhang, Jiangbo Guo, Zhezhu Jin</i> | |
| A Lightweight Convolutional Neural Network Using Feature Filtering Module | 12786 |
| <i>Nan Jing, Yu Zhang</i> | |
| Meta++ Network for Few-Shot Aerospace Crack Segmentation | 12791 |
| <i>Chengyuan Xu, Kang Liu, Xuelong Li</i> | |
| Time-Domain Speech Enhancement Assisted by Multi-Resolution Frequency Encoder and Decoder | 12796 |
| <i>Hao Shi, Masato Mimura, Longbiao Wang, Jianwu Dang, Tatsuya Kawahara</i> | |
| Contextual Similarity is More Valuable than Character Similarity: An Empirical Study for Chinese Spell Checking | 12801 |
| <i>Ding Zhang, Yinghui Li, Qingyu Zhou, Shirong Ma, Yangning Li, Yunbo Cao, Hai-Tao Zheng</i> | |
| FedVMR: A New Federated Learning Method for Video Moment Retrieval..... | 12806 |
| <i>Yan Wang, Xin Luo, Zhen-Duo Chen, Peng-Fei Zhang, Meng Liu, Xin-Shun Xu</i> | |
| Prompt-Distiller: Few-Shot Knowledge Distillation for Prompt-Based Language Learners with Dual Contrastive Learning | 12811 |
| <i>Boyu Hou, Chengyu Wang, Xiaoqing Chen, Minghui Qiu, Liang Feng, Jun Huang</i> | |
| Hypernetwork-Based Adaptive Image Restoration | 12816 |
| <i>Shai Aharon, Gil Ben-Artzi</i> | |
| Vision2Touch: Imaging Estimation of Surface Tactile Physical Properties..... | 12821 |
| <i>Jie Chen, Shizhe Zhou</i> | |
| Dialogue System with Missing Observation | 12826 |
| <i>Djallel Bouneffouf, Mayank Agarwal, Irina Rish</i> | |
| Adversarially Robust Fairness-Aware Regression..... | 12831 |
| <i>Yulu Jin, Lifeng Lai</i> | |
| I See What You Hear: A Vision-Inspired Method to Localize Words..... | 12836 |
| <i>Mohammad Samragh, Arnav Kundu, Ting-Yao Hu, Aman Chadha, Ashish Srivastava, Minsik Cho, Oncel Tuzel, Devang Naik</i> | |

| | |
|--|-------|
| Stabilising and Accelerating Light Gated Recurrent Units for Automatic Speech Recognition..... | 12841 |
| <i>Adel Moumen, Titouan Parcollet</i> | |
| Constrained Independent Component Analysis Based on Entropy Bound Minimization for Subgroup Identification from Multi-Subject fMRI Data..... | 12846 |
| <i>H. Yang, F. Ghayem, B. Gabrielson, M. A. B. S. Akhonda, V. D. Calhoun, T. Adali</i> | |
| Step Restriction for Improving Adversarial Attacks..... | 12851 |
| <i>Keita Goto, Shinta Otake, Rei Kawakami, Nakamasa Inoue</i> | |
| Semi-Supervised Local Structured Feature Learning with Dynamic Maximum Entropy Graph | 12856 |
| <i>Rui Xu, Xun Liang</i> | |
| Neural Architecture of Speech..... | 12861 |
| <i>Subba Reddy Oota, Khushbu Pahwa, Mounika Marreddy, Manish Gupta, Bapi S. Raju</i> | |
| Cocktail Hubert: Generalized Self-Supervised Pre-Training for Mixture and Single-Source Speech..... | 12866 |
| <i>Maryam Fazel-Zarandi, Wei-Ning Hsu</i> | |
| Image Generation is May All You Need for VQA..... | 12871 |
| <i>Kyungho Kim, Junseo Lee, Jihwa Lee</i> | |
| Top-K Visual Tokens Transformer: Selecting Tokens for Visible-Infrared Person Re-Identification..... | 12876 |
| <i>Bin Yang, Jun Chen, Mang Ye</i> | |
| Weakly- And Semi-Supervised Object Localization | 12881 |
| <i>Zhen-Tang Huang, Yan-He Chen, Mei-Chen Yeh</i> | |
| Kalmanbot: Kalmannet-Aided Bollinger Bands for Pairs Trading..... | 12886 |
| <i>Haoran Deng, Guy Revach, Hai Morgenstern, Nir Shlezinger</i> | |
| Complementary Learning System Based Intrinsic Reward in Reinforcement Learning..... | 12891 |
| <i>Zijian Gao, Kele Xu, Hongda Jia, Tianjiao Wan, Bo Ding, Dawei Feng, Xinjun Mao, Huaimin Wang</i> | |
| Filter Pruning Via Filters Similarity in Consecutive Layers..... | 12896 |
| <i>Xiaorui Wang, Jun Wang, Xin Tang, Peng Gao, Rui Fang, Guotong Xie</i> | |
| Coarse-To-Fine Knowledge Selection for Document Grounded Dialogs..... | 12901 |
| <i>Yeqin Zhang, Haomin Fu, Cheng Fu, Haiyang Yu, Yongbin Li, Cam-Tu Nguyen</i> | |
| Audio-Driven High Definition and Lip-Synchronized Talking Face Generation Based on Face Reenactment | 12906 |
| <i>Xianyu Wang, Yuhan Zhang, Weihua He, Yaoyuan Wang, Minglei Li, Yuchen Wang, Jingyi Zhang, Shunbo Zhou, Ziyang Zhang</i> | |
| Advancing the Dimensionality Reduction of Speaker Embeddings for Speaker Diarisation: Disentangling Noise and Informing Speech Activity | 12911 |
| <i>You Jin Kim, Hee-Soo Heo, Jee-Weon Jung, Youngki Kwon, Bong-Jin Lee, Joon Son Chung</i> | |
| Improving Adversarial Robustness with Hypersphere Embedding and Angular-Based Regularizations..... | 12916 |
| <i>Olukorede Fakorede, Ashutosh Nirala, Modeste Atsague, Jin Tian</i> | |
| Matching-Based Term Semantics Pre-Training for Spoken Patient Query Understanding | 12921 |
| <i>Zefa Hu, Xiuyi Chen, Haoran Wu, Minglun Han, Ziyi Ni, Jing Shi, Shuang Xu, Bo Xu</i> | |

| | |
|--|-------|
| BTS-E: Audio Deepfake Detection Using Breathing-Talking-Silence Encoder..... | 12926 |
| <i>Thien-Phuc Doan, Long Nguyen-Vu, Souhwan Jung, Kihun Hong</i> | |
| Keyword-Specific Acoustic Model Pruning for Open-Vocabulary Keyword Spotting | 12931 |
| <i>Yujie Yang, Kun Zhang, Zhiyong Wu, Helen Meng</i> | |
| Long-Memory Message-Passing for Spatially Coupled Systems..... | 12936 |
| <i>Keigo Takeuchi</i> | |
| Hierarchical Multi-Task Learning for Fabric Component Analysis Based on NIR Spectral Signals | 12941 |
| <i>Joseph Kim, Dong Wu, Mingmin Chi, Gaoqi Xu</i> | |
| Boosting Prompt-Based Few-Shot Learners Through Out-of-Domain Knowledge Distillation | 12946 |
| <i>Xiaoqing Chen, Chengyu Wang, Junwei Dong, Minghui Qiu, Liang Feng, Jun Huang</i> | |
| Transient Dictionary Learning for Compressed Time-Of-Flight Imaging..... | 12951 |
| <i>Miguel Heredia Conde</i> | |
| Biologically-Inspired Continual Learning of Human Motion Sequences..... | 12956 |
| <i>Joachim Ott, Shih-Chii Liu</i> | |
| Towards a Unified Training for Levenshtein Transformer | 12961 |
| <i>Kangjie Zheng, Longyue Wang, Zhihao Wang, Binqi Chen, Ming Zhang, Zhaopeng Tu</i> | |
| Nested Attention Network with Graph Filtering for Visual Question and Answering..... | 12966 |
| <i>Jing Lu, Chunlei Wu, Leiquan Wang, Shaozu Yuan, Jie Wu</i> | |
| Unsupervised Deep Digital Staining for Microscopic Cell Images Via Knowledge Distillation | 12971 |
| <i>Ziwan Xu, Lanqing Guo, Shuyan Zhang, Alex C. Kot, Bihan Wen</i> | |
| A Comparison of Semi-Supervised Learning Techniques for Streaming ASR at Scale | 12976 |
| <i>Cal Peyser, Michael Picheny, Kyunghyun Cho, Rohit Prabhavalkar, W. Ronny Huang, Tara N. Sainath</i> | |
| Extended Expectation Maximization for Under-Fitted Models..... | 12981 |
| <i>Aref Miri Rekavandi, Abd-Krim Seghouane, Farid Boussaid, Mohammed Bennamoun</i> | |
| WAVELET2VEC: A Filter Bank Masked Autoencoder for EEG-Based Seizure Subtype Classification..... | 12986 |
| <i>Ruimin Peng, Changming Zhao, Yifan Xu, Jun Jiang, Guangtao Kuang, Jianbo Shao, Dongrui Wu</i> | |
| Semantics-Aware Gamma Correction for Unsupervised Low-Light Image Enhancement | 12991 |
| <i>Yu-Hsuan Chen, Fu-Cheng Pan, Yu-Chien Liao, Jao-Hong Kao, Yu-Chiang Frank Wang</i> | |
| ST360IQ: No-Reference Omnidirectional Image Quality Assessment with Spherical Vision Transformers..... | 12996 |
| <i>Nafiseh Jabbari Tofighi, Mohamed Hedi Elfkir, Nevrez Imamoglu, Cagri Ozcinar, Erkut Erdem, Aykut Erdem</i> | |
| Quantifying Catastrophic Forgetting in Continual Federated Learning | 13001 |
| <i>Christophe Dupuy, Jimit Majmudar, Jixuan Wang, Tanya G. Roosta, Rahul Gupta, Clement Chung, Jie Ding, Salman Avestimehr</i> | |
| Breaking the Trade-Off in Personalized Speech Enhancement with Cross-Task Knowledge Distillation | 13006 |
| <i>Hassan Taherian, Sefik Emre Eskimez, Takuya Yoshioka</i> | |

| | |
|---|-------|
| Multi-Blank Transducers for Speech Recognition..... | 13011 |
| <i>Hainan Xu, Fei Jia, Somshubra Majumdar, Shinji Watanabe, Boris Ginsburg</i> | |
| E2E Segmentation in a Two-Pass Cascaded Encoder ASR Model..... | 13016 |
| <i>W. Ronny Huang, Shuo-Yiin Chang, Tara N. Sainath, Yan Zhang He, David Rybach, Robert David, Rohit Prabhavalkar, Cyril Allauzen, Cal Peyser, Trevor D. Strohman</i> | |
| CO-NET: Classification-Oriented Point Cloud Sampling Via Informative Feature Learning and Non-Overlapped Local Adjustment..... | 13021 |
| <i>Yanan Lin, Keyu Chen, Shihao Zhou, Yunan Huang, Yunqi Lei</i> | |
| A Study on the Invariance in Security Whatever the Dimension of Images for the Steganalysis by Deep-Learning..... | 13026 |
| <i>Kévin Planolles, Marc Chaumont, Frédéric Comby</i> | |
| Data-Aware Zero-Shot Neural Architecture Search for Image Recognition..... | 13031 |
| <i>Yi Fan, Zhong-Han Niu, Yu-Bin Yang</i> | |
| Structure-Preserving and Redundancy-Free Features Refinement for Generalized Zero-Shot Learning | 13036 |
| <i>Jian Ni, Yong Liao</i> | |
| Variational Message Passing-Based Respiratory Motion Estimation and Detection Using Radar Signals | 13041 |
| <i>Jakob Möderl, Erik Leitinger, Franz Pernkopf, Klaus Witrisal</i> | |
| Depth Estimation for a Single Omnidirectional Image with Reversed-Gradient Warming-up Thresholds Discriminator | 13046 |
| <i>Yihong Wu, Yuwen Heng, Mahesan Niranjan, Hansung Kim</i> | |
| Perceive and Predict: Self-Supervised Speech Representation Based Loss Functions for Speech Enhancement | 13051 |
| <i>George Close, William Ravenscroft, Thomas Hain, Stefan Goetze</i> | |
| A Template Matching Approach for Reference Picture Padding in Video Coding..... | 13056 |
| <i>Nicolas Horst, Priyanka Das, Mathias Wien</i> | |
| Channel Estimation in Massive MIMO with Heavy-Tailed Noise: Gaussian-Mixture Versus Cauchy Models..... | 13061 |
| <i>Ziya Gülgün, Erik G. Larsson</i> | |
| Learning to Locate Visual Answer in Video Corpus Using Question..... | 13065 |
| <i>Bin Li, Yixuan Weng, Bin Sun, Shutao Li</i> | |
| Subspace Hybrid Beamforming for Head-Worn Microphone Arrays..... | 13070 |
| <i>Sina Hafezi, Alastair H. Moore, Pierre Guiraud, Patrick A. Naylor, Jacob Donley, Vladimir Tourbabin, Thomas Lunner</i> | |
| Overlay Cognitive Radio Using Symbol Level Precoding with Quantized CSI | 13075 |
| <i>Lu Liu, A. Lee Swindlehurst</i> | |
| Subject-Specific Adaptation for a Causally-Trained Auditory-Attention Decoding System..... | 13080 |
| <i>Christine Beauchene, Michael Brandstein, Stephanie Haro, Thomas F. Quatieri, Christopher J. Smalt</i> | |
| Synthesizer Preset Interpolation Using Transformer Auto-Encoders | 13085 |
| <i>Gwendal Le Vaillant, Thierry Dutoit</i> | |

| | |
|--|-------|
| Audiodec: An Open-Source Streaming High-Fidelity Neural Audio Codec | 13090 |
| <i>Yi-Chiao Wu, Israel D. Gebru, Dejan Markovic, Alexander Richard</i> | |
| Learning Cross-Modal Audiovisual Representations with Ladder Networks for Emotion Recognition | 13095 |
| <i>Lucas Goncalves, Carlos Busso</i> | |
| Wassertein Gan Synthesis for Time Series with Complex Temporal Dynamics: Frugal Architectures and Arbitrary Sample-Size Generation..... | 13100 |
| <i>Th. Beroud, P. Abry, Y. Malevergne, M. Senneret, G. Perrin, J. Macq</i> | |
| Sine: Similarity-Regularized Intra-Class Exploitation for Cross-Granularity Few-Shot Learning | 13105 |
| <i>Jinhai Yang, Hua Yang</i> | |
| Simulating Realistic Speech Overlaps Improves Multi-Talker ASR | 13110 |
| <i>Muqiao Yang, Naoyuki Kanda, Xiaofei Wang, Jian Wu, Sunit Sivasankaran, Zhuo Chen, Jinyu Li, Takuya Yoshioka</i> | |
| Paaploss: A Phonetic-Aligned Acoustic Parameter Loss for Speech Enhancement | 13115 |
| <i>Muqiao Yang, Joseph Konan, David Bick, Yunyang Zeng, Shuo Han, Anurag Kumar, Shinji Watanabe, Bhiksha Raj</i> | |
| Deep Quantigraphic Image Enhancement Via Comparametric Equations | 13120 |
| <i>Xiaomeng Wu, Yongqing Sun, Akisato Kimura</i> | |
| Transadapt: A Transformative Framework for Online Test Time Adaptive Semantic Segmentation | 13125 |
| <i>Debasmit Das, Shubhankar Borse, Hyojin Park, Kambiz Azarian, Hong Cai, Risheek Garrepalli, Fatih Porikli</i> | |
| Inverse Quadratic Transform for Minimizing a Sum of Ratios | 13130 |
| <i>Yannan Chen, Licheng Zhao, Yaowen Zhang, Kaiming Shen</i> | |
| Boosting Fine-Grained Sketch-Based Image Retrieval with Self-Supervised Learning | 13135 |
| <i>Zhaolong Zhang, Yangdong Chen, Yuejie Zhang, Rui Feng, Tao Zhang</i> | |
| D2Former: A Fully Complex Dual-Path Dual-Decoder Conformer Network Using Joint Complex Masking and Complex Spectral Mapping for Monaural Speech Enhancement | 13140 |
| <i>Shengkui Zhao, Bin Ma</i> | |
| MossFormer: Pushing the Performance Limit of Monaural Speech Separation Using Gated Single- Head Transformer with Convolution-Augmented Joint Self-Attentions | 13145 |
| <i>Shengkui Zhao, Bin Ma</i> | |
| Target Speaker Voice Activity Detection with Transformers and Its Integration with End-To-End Neural Diarization | 13150 |
| <i>Dongmei Wang, Xiong Xiao, Naoyuki Kanda, Takuya Yoshioka, Jian Wu</i> | |
| Space-Time Variable Density Samplings for Sparse Bandlimited Graph Signals Driven by Diffusion Operators | 13155 |
| <i>Qing Yao, Longxiu Huang, Sui Tang</i> | |
| Long-Term Synchronization of Wireless Acoustic Sensor Networks with Nonpersistent Acoustic Activity Using Coherence State | 13160 |
| <i>Aleksej Chinaev, Niklas Knaepper, Gerald Enzner</i> | |
| Mitigating Unintended Memorization in Language Models Via Alternating Teaching..... | 13165 |
| <i>Zhe Liu, Xuedong Zhang, Fuchun Peng</i> | |

| | |
|---|-------|
| Video Captioning Via Relation-Aware Graph Learning | 13170 |
| <i>Yi Zheng, Heming Jing, Qiuji Xie, Yuejie Zhang, Rui Feng, Tao Zhang, Shang Gao</i> | |
| Acoustically-Driven Phoneme Removal that Preserves Vocal Affect Cues..... | 13175 |
| <i>Camille Noufi, Jonathan Berger, Karen J. Parker, Daniel L. Bowling</i> | |
| Multi-Source Templates Learning for Real-Time Aerial Tracking..... | 13180 |
| <i>Yiming Sun, Yang Li, Changbo Wang</i> | |
| Diffroll: Diffusion-Based Generative Music Transcription with Unsupervised Pretraining Capability | 13185 |
| <i>Kin Wai Cheuk, Ryosuke Sawata, Toshimitsu Uesaka, Naoki Murata, Naoya Takahashi, Shusuke Takahashi, Dorien Herremans, Yuki Mitsufuji</i> | |
| Sparsity-Smoothness-Aware Power Spectral Density Estimation with Application to Phased Array Weather Radar | 13190 |
| <i>Hiroki Kuroda, Daichi Kitahara, Eiichi Yoshikawa, Hiroshi Kikuchi, Tomoo Ushio</i> | |
| Optimal Mixed-ADC Arrangement for DOA Estimation Via CRB Using ULA..... | 13195 |
| <i>Xinnan Zhang, Yuanbo Cheng, Xiaolei Shang, Jun Liu</i> | |
| Higher-Order Sparse Convolutions in Graph Neural Networks | 13200 |
| <i>Jhony H. Giraldo, Sajid Javed, Arif Mahmood, Fragkiskos D. Malliaros, Thierry Bouwmans</i> | |
| Improving Weakly Supervised Sound Event Detection with Causal Intervention | 13205 |
| <i>Yifei Xin, Dongchao Yang, Fan Cui, Yujun Wang, Yuexian Zou</i> | |
| Robust Iterative Solution for Linear Array-Based 3-D Localization by Message Passing..... | 13210 |
| <i>Yimao Sun, K. C. Ho, Yanbing Yang, Lei Zhang, Liangyin Chen</i> | |
| Backdoor Defense Via Suppressing Model Shortcuts | 13215 |
| <i>Sheng Yang, Yiming Li, Yong Jiang, Shu-Tao Xia</i> | |
| Sparse Representations with Cone Atoms | 13220 |
| <i>Denis C. Ilie-Ablachim, Andra Baltoiu, Bogdan Dumitrescu</i> | |
| A Spatio-Temporal Decomposition Network for Compressed Video Quality Enhancement | 13225 |
| <i>Kai Wang, Fangdong Chen, Zongmiao Ye, Li Wang, Xiaoyang Wu, Shiliang Pu</i> | |
| A Physically Explainable Framework for Human-Related Anomaly Detection..... | 13230 |
| <i>Yalong Jiang, Huining Li, Changkang Li</i> | |
| High-Resolution Neural Network Processing of LFM Radar Pulses | 13235 |
| <i>Jabran Akhtar</i> | |

Author Index