

International Conference on Learning Representations

(ICLR 2021)

4 May 2021, Vienna, Austria

ISBN: 979-8-3313-0008-1

| Title and Authors | Page# |
|--|-------|
| <i>On the Mapping Between Hopfield Networks and Restricted Boltzmann Machines</i> <i>Matthew Smart, Anton Zilman</i> | 1 |
| <i>Complex Query Answering with Neural Link Predictors</i> <i>Erik Arakelyan, Daniel Daza, Pasquale Minervini, Michael Cochez</i> | 19 |
| <i>Share Or Not? Learning to Schedule Language-Specific Capacity for Multilingual Translation</i> <i>Biao Zhang, Ankur Bapna, Rico Sennrich, Orhan Firat</i> | 33 |
| <i>End-To-End Adversarial Text-To-Speech</i> <i>Jeff Donahue, Sander Dieleman, Mikolaj Binkowski, Erich Elsen, Karen Simonyan</i> | 52 |
| <i>Co-Mixup: Saliency Guided Joint Mixup with Supermodular Diversity</i> <i>Janghyun Kim, Wonho Choo, Hosan Jeong, Hyun Oh Song</i> | 75 |
| <i>Contrastive Explanations for Reinforcement Learning Via Embedded Self Predictions</i> <i>Zhengxian Lin, Kin-Ho Lam, Alan Fern</i> | 96 |
| <i>Why Are Convolutional Nets More Sample-Efficient than Fully-Connected Nets?</i> <i>Zhiyuan Li, Yi Zhang, Sanjeev Arora</i> | 117 |
| <i>Iterated Learning for Emergent Systematicity in VQA</i> <i>Ankit Vani, Max Schwarzer, Yuchen Lu, Eeshan Dhekane, Aaron Courville</i> | 141 |
| <i>When Do Curricula Work?</i> <i>Xiaoxia Wu, Ethan Dyer, Behnam Neyshabur</i> | 162 |
| <i>Deep Symbolic Regression: Recovering Mathematical Expressions from Data Via Risk-Seeking Policy Gradients</i> <i>Brenden K Petersen, Mikel Landajuela Larra, Terrell N. Mundhenk, Claudio Prata Santiago, Soo Kyung Kim, Joanne Taery Kim</i> | 185 |
| <i>Improved Autoregressive Modeling with Distribution Smoothing</i> <i>Chenlin Meng, Jiaming Song, Yang Song, Shengjia Zhao, Stefano Ermon</i> | 211 |
| <i>Score-Based Generative Modeling Through Stochastic Differential Equations</i> <i>Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, Ben Poole</i> | 240 |

| | |
|---|-----|
| <p>Global Convergence of Three-Layer Neural Networks in the Mean Field Regime</p> <p>Huy Tuan Pham, Phan-Minh Nguyen</p> | 276 |
| <p>Rethinking Architecture Selection in Differentiable NAS</p> <p>Ruochen Wang, Minhao Cheng, Xiangning Chen, Xiaocheng Tang, Cho-Jui Hsieh</p> | 311 |
| <p>Evolving Reinforcement Learning Algorithms</p> <p>John D Co-Reyes, Yingjie Miao, Daiyi Peng, Esteban Real, Quoc V Le, Sergey Levine, Honglak Lee, Aleksandra Faust</p> | 329 |
| <p>Image GANs Meet Differentiable Rendering for Inverse Graphics and Interpretable 3D Neural Rendering</p> <p>Yuxuan Zhang, Wenzheng Chen, Huan Ling, Jun Gao, Yinan Zhang, Antonio Torralba, Sanja Fidler</p> | 344 |
| <p>MONGOOSE: A Learnable LSH Framework for Efficient Neural Network Training</p> <p>Beidi Chen, Zichang Liu, Binghui Peng, Zhaozhuo Xu, Jonathan Lingjie Li, Tri Dao, Zhao Song, Anshumali Shrivastava, Christopher Re</p> | 369 |
| <p>Parrot: Data-Driven Behavioral Priors for Reinforcement Learning</p> <p>Avi Singh, Huihan Liu, Gaoyue Zhou, Albert Yu, Nicholas Rhinehart, Sergey Levine</p> | 402 |
| <p>SenSel: Sensitive Set Invariance for Enforcing Individual Fairness</p> <p>Mikhail Yurochkin, Yuekai Sun</p> | 421 |
| <p>Theoretical Analysis of Self-Training with Deep Networks on Unlabeled Data</p> <p>Colin Wei, Kendrick Shen, Yining Chen, Tengyu Ma</p> | 440 |
| <p>Growing Efficient Deep Networks by Structured Continuous Sparsification</p> <p>Xin Yuan, Pedro Henrique Pamplona Savarese, Michael Maire</p> | 470 |
| <p>VCNet and Functional Targeted Regularization for Learning Causal Effects of Continuous Treatments</p> <p>Lizhen Nie, Mao Ye, Qiang Liu, Dan Nicolae</p> | 488 |
| <p>EigenGame: PCA as a Nash Equilibrium</p> <p>Ian Gemp, Brian McWilliams, Claire Vernade, Thore Graepel</p> | 512 |
| <p>Randomized Automatic Differentiation</p> <p>Deniz Oktay, Nick McGreivy, Joshua Aduol, Alex Beatson, Ryan P Adams</p> | 523 |
| <p>A Distributional Approach to Controlled Text Generation</p> <p>Muhammad Khalifa, Hady Elsahar, Marc Dymetman</p> | 542 |
| <p>An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale</p> <p>Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby</p> | 611 |

| | |
|--|-----|
| <p>Getting a CLUE: A Method for Explaining Uncertainty Estimates</p> <p><i>Javier Antoran, Umang Bhatt, Tameem Adel, Adrian Weller, José Miguel Hernández-Lobato</i></p> | 632 |
| <p>Optimal Rates for Averaged Stochastic Gradient Descent Under Neural Tangent Kernel Regime</p> <p><i>Atsushi Nitanda, Taiji Suzuki</i></p> | 666 |
| <p>Learning Invariant Representations for Reinforcement Learning Without Reconstruction</p> <p><i>Amy Zhang, Rowan Thomas McAllister, Roberto Calandra, Yarin Gal, Sergey Levine</i></p> | 701 |
| <p>Rethinking the Role of Gradient-Based Attribution Methods for Model Interpretability</p> <p><i>Suraj Srinivas, Francois Fleuret</i></p> | 718 |
| <p>SMiRL: Surprise Minimizing Reinforcement Learning in Unstable Environments</p> <p><i>Glen Berseth, Daniel Geng, Coline Manon Devin, Nicholas Rhinehart, Chelsea Finn, Dinesh Jayaraman, Sergey Levine</i></p> | 733 |
| <p>Learning to Reach Goals Via Iterated Supervised Learning</p> <p><i>Dibya Ghosh, Abhishek Gupta, Ashwin Reddy, Justin Fu, Coline Manon Devin, Benjamin Eysenbach, Sergey Levine</i></p> | 750 |
| <p>Self-Training for Few-Shot Transfer Across Extreme Task Differences</p> <p><i>Cheng Perng Phoo, Bharath Hariharan</i></p> | 771 |
| <p>Federated Learning Based on Dynamic Regularization</p> <p><i>Durmus Alp Emre Acar, Yue Zhao, Ramon Matas, Matthew Mattina, Paul Whatmough, Venkatesh Saligrama</i></p> | 790 |
| <p>Rao-Blackwellizing the Straight-Through Gumbel-Softmax Gradient Estimator</p> <p><i>Max B Paulus, Chris J. Maddison, Andreas Krause</i></p> | 826 |
| <p>Coupled Oscillatory Recurrent Neural Network (coRNN): An Accurate and (gradient) Stable Architecture for Learning Long Time Dependencies</p> <p><i>T. Konstantin Rusch, Siddhartha Mishra</i></p> | 837 |
| <p>DiffWave: A Versatile Diffusion Model for Audio Synthesis</p> <p><i>Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, Bryan Catanzaro</i></p> | 862 |
| <p>Learning Cross-Domain Correspondence for Control with Dynamics Cycle-Consistency</p> <p><i>Qiang Zhang, Tete Xiao, Alexei A Efros, Lerrel Pinto, Xiaolong Wang</i></p> | 879 |
| <p>Deformable DETR: Deformable Transformers for End-To-End Object Detection</p> <p><i>Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, Jifeng Dai</i></p> | 894 |
| <p>Learning Generalizable Visual Representations Via Interactive Gameplay</p> <p><i>Luca Weihs, Aniruddha Kembhavi, Kiana Ehsani, Sarah M Pratt, Winson Han, Alvaro Herrasti, Eric Kolve, Dustin Schwenk, Roozbeh Mottaghi, Ali Farhadi</i></p> | 910 |
| <p>Gradient Projection Memory for Continual Learning</p> <p><i>Gobinda Saha, Isha Garg, Kaushik Roy</i></p> | 944 |

| | |
|---|------|
| <p>MONGOOSE: A Learnable LSH Framework for Efficient Neural Network Training</p> <p><i>Yuan Yin, Vincent Le Guen, Jérémie Dona, Emmanuel De Bezenac, Ibrahim Ayed, Nicolas Thome, Patrick Gallinari</i></p> | 962 |
| <p>Human-Level Performance in No-Press Diplomacy Via Equilibrium Search</p> <p><i>Jonathan Gray, Adam Lerer, Anton Bakhtin, Noam Brown</i></p> | 984 |
| <p>How Neural Networks Extrapolate: From Feedforward to Graph Neural Networks</p> <p><i>Keyulu Xu, Mozhi Zhang, Jingling Li, Simon Shaolei Du, Ken-Ichi Kawarabayashi, Stefanie Jegelka</i></p> | 1008 |
| <p>Rethinking Attention with Performers</p> <p><i>Krzysztof Marcin Choromanski, Valerii Likhoshesterov, David Dohan, Xingyou Song, Andreea Gane, Tamas Sarlos, Peter Hawkins, Jared Quincy Davis, Afroz Mohiuddin, Lukasz Kaiser, David Benjamin Belanger, Lucy J Colwell, Adrian Weller</i></p> | 1060 |
| <p>What Matters for On-Policy Deep Actor-Critic Methods? a Large-Scale Study</p> <p><i>Marcin Andrychowicz, Anton Raichuk, Piotr Stanczyk, Manu Orsini, Sertan Girgin, Raphaël Marinier, Leonard Hussenot, Matthieu Geist, Olivier Pietquin, Marcin Michalski, Sylvain Gelly, Olivier Bachem</i></p> | 1098 |
| <p>Neural Synthesis of Binaural Speech from Mono Audio</p> <p><i>Alexander Richard, Dejan Markovic, Israel D. Gebru, Steven Krenn, Gladstone Alexander Butler, Fernando Torre, Yaser Sheikh</i></p> | 1108 |
| <p>Dataset Condensation with Gradient Matching</p> <p><i>Bo Zhao, Konda Reddy Mopuri, Hakan Bilen</i></p> | 1131 |
| <p>Scalable Learning and MAP Inference for Nonsymmetric Determinantal Point Processes</p> <p><i>Mike Gartrell, Insu Han, Elvis Dohmatob, Jennifer Gillenwater, Victor-Emmanuel Brunel</i></p> | 1151 |
| <p>Geometry-Aware Instance-Reweight Adversarial Training</p> <p><i>Jingfeng Zhang, Jianing Zhu, Gang Niu, Bo Han, Masashi Sugiyama, Mohan Kankanhalli</i></p> | 1172 |
| <p>Free Lunch for Few-Shot Learning: Distribution Calibration</p> <p><i>Shuo Yang, Lu Liu, Min Xu</i></p> | 1201 |
| <p>Towards Nonlinear Disentanglement in Natural Data with Temporal Sparse Coding</p> <p><i>David A. Klindt, Lukas Schott, Yash Sharma, Ivan Ustyuzhaninov, Wieland Brendel, Matthias Bethge, Dylan Paiton</i></p> | 1214 |
| <p>Do 2D GANs Know 3D Shape? Unsupervised 3D Shape Reconstruction from 2D Image GANs</p> <p><i>Xingang Pan, Bo Dai, Ziwei Liu, Chen Change Loy, Ping Luo</i></p> | 1265 |
| <p>Practical Real Time Recurrent Learning with a Sparse Approximation</p> <p><i>Jacob Menick, Erich Elsen, Utku Evci, Simon Osindero, Karen Simonyan, Alex Graves</i></p> | 1283 |
| <p>Retrieval-Augmented Generation for Code Summarization Via Hybrid GNN</p> <p><i>Shangqing Liu, Yu Chen, Xiaofei Xie, Jing Kai Siow, Yang Liu</i></p> | 1299 |

| | |
|--|------|
| <p>Benefit of Deep Learning with Non-Convex Noisy Gradient Descent: Provable Excess Risk Bound and Superiority to Kernel Methods</p> <p><i>Taiji Suzuki, Shunta Akiyama</i></p> | 1315 |
| <p>Self-Supervised Visual Reinforcement Learning with Object-Centric Representations</p> <p><i>Andrii Zadaianchuk, Maximilian Seitzer, Georg Martius</i></p> | 1336 |
| <p>Discovering a Set of Policies for the Worst Case Reward</p> <p><i>Tom Zahavy, Andre Barreto, Daniel J Mankowitz, Shaobo Hou, Brendan O'Donoghue, Iurii Kemaev, Satinder Singh</i></p> | 1357 |
| <p>Implicit Normalizing Flows</p> <p><i>Cheng Lu, Jianfei Chen, Chongxuan Li, Qiuhan Wang, Jun Zhu</i></p> | 1374 |
| <p>Gauge Equivariant Mesh CNNs: Anisotropic Convolutions on Geometric Graphs</p> <p><i>Pim De Haan, Maurice Weiler, Taco Cohen, Max Welling</i></p> | 1398 |
| <p>Generalization Bounds Via Distillation</p> <p><i>Daniel Hsu, Ziwei Ji, Matus Telgarsky, Lan Wang</i></p> | 1410 |
| <p>Learning Mesh-Based Simulation with Graph Networks</p> <p><i>Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, Peter Battaglia</i></p> | 1435 |
| <p>On Statistical Bias in Active Learning: How and When to Fix it</p> <p><i>Sebastian Farquhar, Yarin Gal, Tom Rainforth</i></p> | 1453 |
| <p>Winning the L2RPN Challenge: Power Grid Management Via Semi-Markov Afterstate Actor-Critic</p> <p><i>Deunsol Yoon, Sunghoon Hong, Byung-Jun Lee, Kee-Eung Kim</i></p> | 1483 |
| <p>Expressive Power of Invariant and Equivariant Graph Neural Networks</p> <p><i>Waiss Azizian, Marc Lelarge</i></p> | 1501 |
| <p>MARS: Markov Molecular Sampling for Multi-Objective Drug Discovery</p> <p><i>Yutong Xie, Chence Shi, Hao Zhou, Yuwei Yang, Weinan Zhang, Yong Yu, Lei Li</i></p> | 1540 |
| <p>On Self-Supervised Image Representations for GAN Evaluation</p> <p><i>Stanislav Morozov, Andrey Voynov, Artem Babenko</i></p> | 1559 |
| <p>Identifying Nonlinear Dynamical Systems with Multiple Time Scales and Long-Range Dependencies</p> <p><i>Dominik Schmidt, Georgia Koppe, Zahra Monfared, Max Beutelspacher, Daniel Durstewitz</i></p> | 1576 |
| <p>Understanding the Role of Importance Weighting for Deep Learning</p> <p><i>Da Xu, Yuting Ye, Chuanwei Ruan</i></p> | 1605 |
| <p>RMSprop Converges with Proper Hyper-Parameter</p> <p><i>Naichen Shi, Dawei Li, Mingyi Hong, Ruoyu Sun</i></p> | 1625 |

| | |
|--|------|
| <p><i>The Intrinsic Dimension of Images and Its Impact on Learning</i></p> <p><i>Phil Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, Tom Goldstein</i></p> | 1635 |
| <p><i>Learning with Feature-Dependent Label Noise: A Progressive Approach</i></p> <p><i>Yikai Zhang, Songzhu Zheng, Pengxiang Wu, Mayank Goswami, Chao Chen</i></p> | 1652 |
| <p><i>Disentangled Recurrent Wasserstein Autoencoder</i></p> <p><i>Jun Han, Martin Renqiang Min, Ligong Han, Li Erran Li, Xuan Zhang</i></p> | 1665 |
| <p><i>Influence Estimation for Generative Adversarial Networks</i></p> <p><i>Naoyuki Terashita, Hiroki Ohashi, Yuichi Nonaka, Takashi Kanemaru</i></p> | 1686 |
| <p><i>Orthogonalizing Convolutional Layers with the Cayley Transform</i></p> <p><i>Asher Trockman, J Zico Kolter</i></p> | 1711 |
| <p><i>Predicting Infectiousness for Proactive Contact Tracing</i></p> <p><i>Yoshua Bengio, Prateek Gupta, Tegan Maharaj, Nasim Rahaman, Martin Weiss, Tristan Deleu, Eilif Benjamin Muller, Meng Qu, Victor Schmidt, Pierre-Luc St-Charles, Hannah Alsdurf, Olexa Bilaniuk, David Buckeridge, Gaetan Caron, Pierre Luc Carrier, Joumana Ghosn, Satya Ortiz Gagne, Christopher Pal, Irina Rish, Bernhard Schölkopf, + 3 More Authors</i></p> | 1732 |
| <p><i>Learning a Latent Simplex in Input Sparsity Time</i></p> <p><i>Ainesh Bakshi, Chiranjib Bhattacharyya, Ravi Kannan, David Woodruff, Samson Zhou</i></p> | 1753 |
| <p><i>CPT: Efficient Deep Neural Network Training Via Cyclic Precision</i></p> <p><i>Yonggan Fu, Han Guo, Meng Li, Xin Yang, Yining Ding, Vikas Chandra, Yingyan Lin</i></p> | 1764 |
| <p><i>Watch-And-Help: A Challenge for Social Perception and Human-AI Collaboration</i></p> <p><i>Xavier Puig, Tianmin Shu, Shuang Li, Zilin Wang, Yuan-Hong Liao, Joshua B. Tenenbaum, Sanja Fidler, Antonio Torralba</i></p> | 1778 |
| <p><i>Gradient Vaccine: Investigating and Improving Multi-Task Optimization in Massively Multilingual Models</i></p> <p><i>Zirui Wang, Yulia Tsvetkov, Orhan Firat, Yuan Cao</i></p> | 1801 |
| <p><i>PlasticineLab: A Soft-Body Manipulation Benchmark with Differentiable Physics</i></p> <p><i>Zhiao Huang, Yuanming Hu, Tao Du, Siyuan Zhou, Hao Su, Joshua B. Tenenbaum, Chuang Gan</i></p> | 1823 |
| <p><i>Distributional Sliced-Wasserstein and Applications to Generative Modeling</i></p> <p><i>Khai Nguyen, Nhat Ho, Tung Pham, Hung Bui</i></p> | 1841 |
| <p><i>Systematic Generalisation with Group Invariant Predictions</i></p> <p><i>Faruk Ahmed, Yoshua Bengio, Harm Van Seijen, Aaron Courville</i></p> | 1853 |
| <p><i>Memory Optimization for Deep Networks</i></p> <p><i>Aashaka Shah, Chao-Yuan Wu, Jayashree Mohan, Vijay Chidambaram, Philipp Kraehenbuehl</i></p> | 1872 |

| | |
|--|------|
| <p>HW-NAS-Bench: Hardware-Aware Neural Architecture Search Benchmark</p> <p>Chaojian Li, Zhongzhi Yu, Yonggan Fu, Yongan Zhang, Yang Zhao, Haoran You, Qixuan Yu, Yue Wang, Cong Hao, Yingyan Lin</p> | 1890 |
| <p>Regularized Inverse Reinforcement Learning</p> <p>Wonseok Jeon, Chen-Yang Su, Paul Barde, Thang Doan, Derek Nowrouzezahrai, Joelle Pineau</p> | 1908 |
| <p>What Are the Statistical Limits of Offline RL with Linear Function Approximation?</p> <p>Ruosong Wang, Dean Foster, Sham M. Kakade</p> | 1934 |
| <p>Behavioral Cloning from Noisy Demonstrations</p> <p>Fumihito Sasaki, Ryota Yamashina</p> | 1954 |
| <p>How Does Mixup Help with Robustness and Generalization?</p> <p>Linjun Zhang, Zhun Deng, Kenji Kawaguchi, Amirata Ghorbani, James Zou</p> | 1968 |
| <p>Emergent Symbols Through Binding in External Memory</p> <p>Taylor Whittington Webb, Ishan Sinha, Jonathan Cohen</p> | 1992 |
| <p>Individually Fair Gradient Boosting</p> <p>Alexander Vargo, Fan Zhang, Mikhail Yurochkin, Yuekai Sun</p> | 2020 |
| <p>Structured Prediction as Translation Between Augmented Natural Languages</p> <p>Giovanni Paolini, Ben Athiwaratkun, Jason Krone, Jie Ma, Alessandro Achille, Rishita Anubhai, Cicero Nogueira Dos Santos, Bing Xiang, Stefano Soatto</p> | 2050 |
| <p>Minimum Width for Universal Approximation</p> <p>Sejun Park, Chulhee Yun, Jaeho Lee, Jinwoo Shin</p> | 2076 |
| <p>Topology-Aware Segmentation Using Discrete Morse Theory</p> <p>Xiaoling Hu, Yusu Wang, Li Fuxin, Dimitris Samaras, Chao Chen</p> | 2101 |
| <p>Interpreting Graph Neural Networks for NLP with Differentiable Edge Masking</p> <p>Michael Sejr Schlichtkrull, Nicola De Cao, Ivan Titov</p> | 2120 |
| <p>Regularization Matters in Policy Optimization - An Empirical Study on Continuous Control</p> <p>Zhuang Liu, Xuanlin Li, Bingyi Kang, Trevor Darrell</p> | 2141 |
| <p>Recurrent Independent Mechanisms</p> <p>Anirudh Goyal, Alex Lamb, Jordan Hoffmann, Shagun Sodhani, Sergey Levine, Yoshua Bengio, Bernhard Schölkopf</p> | 2185 |
| <p>Dataset Inference: Ownership Resolution in Machine Learning</p> <p>Pratyush Maini, Mohammad Yaghini, Nicolas Papernot</p> | 2222 |
| <p>Learning from Protein Structure with Geometric Vector Perceptrons</p> <p>Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, Ron Dror</p> | 2244 |

| | |
|---|------|
| <p>Unsupervised Object Keypoint Learning Using Local Spatial Predictability</p> <p>Anand Gopalakrishnan, Sjoerd Van Steenkiste, Jürgen Schmidhuber</p> | 2262 |
| <p>Fast Geometric Projections for Local Robustness Certification</p> <p>Aymeric Fromherz, Klas Leino, Matt Fredrikson, Bryan Parno, Corina Pasareanu</p> | 2281 |
| <p>Random Feature Attention</p> <p>Hao Peng, Nikolaos Pappas, Dani Yogatama, Roy Schwartz, Noah Smith, Lingpeng Kong</p> | 2296 |
| <p>Sharpness-Aware Minimization for Efficiently Improving Generalization</p> <p>Pierre Foret, Ariel Kleiner, Hossein Mobahi, Behnam Neyshabur</p> | 2315 |
| <p>PMI-Masking: Principled Masking of Correlated Spans</p> <p>Yoav Levine, Barak Lenz, Opher Lieber, Omri Abend, Kevin Leyton-Brown, Moshe Tennenholtz, Yoav Shoham</p> | 2334 |
| <p>Learning Incompressible Fluid Dynamics from Scratch - Towards Fast, Differentiable Fluid Models that Generalize</p> <p>Nils Wandel, Michael Weinmann, Reinhard Klein</p> | 2348 |
| <p>A Gradient Flow Framework for Analyzing Network Pruning</p> <p>Ekdeep Singh Lubana, Robert Dick</p> | 2363 |
| <p>Towards Robustness Against Natural Language Word Substitutions</p> <p>Xinshuai Dong, Anh Tuan Luu, Rongrong Ji, Hong Liu</p> | 2391 |
| <p>Iterative Empirical Game Solving Via Single Policy Best Response</p> <p>Max Smith, Thomas Anthony, Michael Wellman</p> | 2405 |
| <p>Self-Supervised Policy Adaptation During Deployment</p> <p>Nicklas Hansen, Rishabh Jangir, Yu Sun, Guillem Alenyà, Pieter Abbeel, Alexei A Efros, Lerrel Pinto, Xiaolong Wang</p> | 2415 |
| <p>Differentially Private Learning Needs Better Features (or Much More Data)</p> <p>Florian Tramer, Dan Boneh</p> | 2433 |
| <p>Data-Efficient Reinforcement Learning with Self-Predictive Representations</p> <p>Max Schwarzer, Ankesh Anand, Rishab Goel, R Devon Hjelm, Aaron Courville, Philip Bachman</p> | 2458 |
| <p>Meta-GMVAE: Mixture of Gaussian VAE for Unsupervised Meta-Learning</p> <p>Dong Bok Lee, Dongchan Min, Seanie Lee, Sung Ju Hwang</p> | 2476 |
| <p>Implicit Convex Regularizers of CNN Architectures: Convex Optimization of Two- And Three-Layer Networks in Polynomial Time</p> <p>Tolga Ergen, Mert Pilanci</p> | 2492 |
| <p>Image Augmentation is All You Need: Regularizing Deep Reinforcement Learning from Pixels</p> <p>Denis Yarats, Ilya Kostrikov, Rob Fergus</p> | 2517 |

| | |
|--|------|
| <p>Dynamic Tensor Rematerialization</p> <p><i>Marisa Kirisame, Steven Lyubomirsky, Altan Haan, Jennifer Brennan, Mike He, Jared Roesch, Tianqi Chen, Zachary Tatlock</i></p> | 2538 |
| <p>Deep Neural Network Fingerprinting by Conferrable Adversarial Examples</p> <p><i>Nils Lukas, Yuxuan Zhang, Florian Kerschbaum</i></p> | 2569 |
| <p>Model-Based Visual Planning with Self-Supervised Functional Distances</p> <p><i>Stephen Tian, Suraj Nair, Frederik Ebert, Sudeep Dasari, Benjamin Eysenbach, Chelsea Finn, Sergey Levine</i></p> | 2587 |
| <p>Mathematical Reasoning Via Self-Supervised Skip-Tree Training</p> <p><i>Markus Norman Rabe, Dennis Lee, Kshitij Bansal, Christian Szegedy</i></p> | 2606 |
| <p>DeepAveragers: Offline Reinforcement Learning by Solving Derived Non-Parametric MDPs</p> <p><i>Aayam Kumar Shrestha, Stefan Lee, Prasad Tadepalli, Alan Fern</i></p> | 2627 |
| <p>On the Theory of Implicit Deep Learning: Global Convergence with Implicit Layers</p> <p><i>Kenji Kawaguchi</i></p> | 2652 |
| <p>BUSTLE: Bottom-Up Program Synthesis Through Learning-Guided Exploration</p> <p><i>Augustus Odena, Kensen Shi, David Bieber, Rishabh Singh, Charles Sutton, Hanjun Dai</i></p> | 2690 |
| <p>The Traveling Observer Model: Multi-Task Learning Through Spatial Variable Embeddings</p> <p><i>Elliot Meyerson, Risto Miikkulainen</i></p> | 2706 |
| <p>Fidelity-Based Deep Adiabatic Scheduling</p> <p><i>Eli Ovits, Lior Wolf</i></p> | 2723 |
| <p>Deciphering and Optimizing Multi-Task Learning: A Random Matrix Approach</p> <p><i>Malik Tiomoko, Hafiz Tiomoko Ali, Romain Couillet</i></p> | 2741 |
| <p>Learning-Based Support Estimation in Sublinear Time</p> <p><i>Talya Eden, Piotr Indyk, Shyam Narayanan, Ronitt Rubinfeld, Sandeep Silwal, Tal Wagner</i></p> | 2752 |
| <p>Unlearnable Examples: Making Personal Data Unexploitable</p> <p><i>Hanxun Huang, Xingjun Ma, Sarah Monazam Erfani, James Bailey, Yisen Wang</i></p> | 2769 |
| <p>How Benign is Benign Overfitting ?</p> <p><i>Amartya Sanyal, Puneet K. Dokania, Varun Kanade, Philip Torr</i></p> | 2786 |
| <p>Autoregressive Entity Retrieval</p> <p><i>Nicola De Cao, Gautier Izacard, Sebastian Riedel, Fabio Petroni</i></p> | 2812 |
| <p>Neural Approximate Sufficient Statistics for Implicit Models</p> <p><i>Yanzhi Chen, Dinghuai Zhang, Michael U. Gutmann, Aaron Courville, Zhanxing Zhu</i></p> | 2832 |
| <p>Large Scale Image Completion Via Co-Modulated Generative Adversarial Networks</p> <p><i>Shengyu Zhao, Jonathan Cui, Yilun Sheng, Yue Dong, Xiao Liang, Eric I-Chao Chang, Yan Xu</i></p> | 2847 |

| | |
|---|------|
| <p>DDPNOpt: Differential Dynamic Programming Neural Optimizer</p> <p><i>Guan-Hong Liu, Tianrong Chen, Evangelos Theodorou</i></p> | 2872 |
| <p>Geometry-Aware Gradient Algorithms for Neural Architecture Search</p> <p><i>Liam Li, Mikhail Khodak, Nina Balcan, Ameet Talwalkar</i></p> | 2890 |
| <p>Beyond Fully-Connected Layers with Quaternions: Parameterization of Hypercomplex Multiplications with $1/n$ Parameters</p> <p><i>Aston Zhang, Yi Tay, Shuai Zhang, Alvin Chan, Anh Tuan Luu, Siu Hui, Jie Fu</i></p> | 2915 |
| <p>Tent: Fully Test-Time Adaptation by Entropy Minimization</p> <p><i>Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, Trevor Darrell</i></p> | 2928 |
| <p>Neural Topic Model Via Optimal Transport</p> <p><i>He Zhao, Dinh Phung, Viet Huynh, Trung Le, Wray Buntine</i></p> | 2943 |
| <p>A Panda? No, It's a Sloth: Slowdown Attacks on Adaptive Multi-Exit Neural Network Inference</p> <p><i>Sanghyun Hong, Yigitcan Kaya, Ionu?-Vlad Modoranu, Tudor Dumitras</i></p> | 2958 |
| <p>Are Neural Rankers Still Outperformed by Gradient Boosted Decision Trees?</p> <p><i>Zhen Qin, Le Yan, Honglei Zhuang, Yi Tay, Rama Kumar Pasumarthi, Xuanhui Wang, Michael Bendersky, Marc Najork</i></p> | 2975 |
| <p>Contrastive Behavioral Similarity Embeddings for Generalization in Reinforcement Learning</p> <p><i>Rishabh Agarwal, Marlos C. Machado, Pablo Samuel Castro, Marc G Bellemare</i></p> | 2991 |
| <p>Very Deep VAEs Generalize Autoregressive Models and Can Outperform Them on Images</p> <p><i>Rewon Child</i></p> | 3021 |
| <p>Async-RED: A Provably Convergent Asynchronous Block Parallel Stochastic Method Using Deep Denoising Priors</p> <p><i>Yu Sun, Jiaming Liu, Yiran Sun, Brendt Wohlberg, Ulugbek Kamilov</i></p> | 3038 |
| <p>A Good Image Generator is What You Need for High-Resolution Video Synthesis</p> <p><i>Yu Tian, Jian Ren, Menglei Chai, Kyle Olszewski, Xi Peng, Dimitris N. Metaxas, Sergey Tulyakov</i></p> | 3050 |
| <p>Undistillable: Making a Nasty Teacher that CANNOT Teach Students</p> <p><i>Haoyu Ma, Tianlong Chen, Ting-Kuei Hu, Chenyu You, Xiaohui Xie, Zhangyang Wang</i></p> | 3073 |
| <p>Support-Set Bottlenecks for Video-Text Representation Learning</p> <p><i>Mandela Patrick, Po-Yao Huang, Yuki Asano, Florian Metze, Alexander G Hauptmann, Joao F. Henriques, Andrea Vedaldi</i></p> | 3085 |
| <p>Grounded Language Learning Fast and Slow</p> <p><i>Felix Hill, Olivier Tieleman, Tamara Von Glehn, Nathaniel Wong, Hamza Merzic, Stephen Clark</i></p> | 3104 |
| <p>GAN "Steerability" Without Optimization</p> <p><i>Nurit Spingarn, Ron Banner, Tomer Michaeli</i></p> | 3121 |

| | |
|--|------|
| <p>Noise Against Noise: Stochastic Label Noise Helps Combat Inherent Label Noise</p> <p><i>Pengfei Chen, Guangyong Chen, Junjie Ye, Jingwei Zhao, Pheng-Ann Heng</i></p> | 3180 |
| <p>VAEBM: A Symbiosis Between Variational Autoencoders and Energy-Based Models</p> <p><i>Zhisheng Xiao, Karsten Kreis, Jan Kautz, Arash Vahdat</i></p> | 3200 |
| <p>Graph-Based Continual Learning</p> <p><i>Binh Tang, David S. Matteson</i></p> | 3233 |
| <p>Sparse Quantized Spectral Clustering</p> <p><i>Zhenyu Liao, Romain Couillet, Michael W. Mahoney</i></p> | 3251 |
| <p>LambdaNetworks: Modeling Long-Range Interactions Without Attention</p> <p><i>Irwan Bello</i></p> | 3272 |
| <p>Contrastive Divergence Learning is a Time Reversal Adversarial Game</p> <p><i>Omer Yair, Tomer Michaeli</i></p> | 3303 |
| <p>Quantifying Differences in Reward Functions</p> <p><i>Adam Gleave, Michael D Dennis, Shane Legg, Stuart Russell, Jan Leike</i></p> | 3318 |
| <p>Long-Tail Learning Via Logit Adjustment</p> <p><i>Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, Sanjiv Kumar</i></p> | 3360 |
| <p>Locally Free Weight Sharing for Network Width Search</p> <p><i>Xiu Su, Shan You, Tao Huang, Fei Wang, Chen Qian, Changshui Zhang, Chang Xu</i></p> | 3384 |
| <p>Mutual Information State Intrinsic Control</p> <p><i>Rui Zhao, Yang Gao, Pieter Abbeel, Volker Tresp, Wei Xu</i></p> | 3396 |
| <p>Multivariate Probabilistic Time Series Forecasting Via Conditioned Normalizing Flows</p> <p><i>Kashif Rasul, Abdul-Saboor Sheikh, Ingmar Schuster, Urs M Bergmann, Roland Vollgraf</i></p> | 3412 |
| <p>Information Laundering for Model Privacy</p> <p><i>Xinran Wang, Yu Xiang, Jun Gao, Jie Ding</i></p> | 3431 |
| <p>UPDeT: Universal Multi-Agent RL Via Policy Decoupling with Transformers</p> <p><i>Siyi Hu, Fengda Zhu, Xiaojun Chang, Xiaodan Liang</i></p> | 3451 |
| <p>Correcting Experience Replay for Multi-Agent Communication</p> <p><i>Sanjeevan Ahilan, Peter Dayan</i></p> | 3466 |
| <p>Improving Adversarial Robustness Via Channel-Wise Activation Suppressing</p> <p><i>Yang Bai, Yuyuan Zeng, Yong Jiang, Shu-Tao Xia, Xingjun Ma, Yisen Wang</i></p> | 3482 |
| <p>Long-Tailed Recognition by Routing Diverse Distribution-Aware Experts</p> <p><i>Xudong Wang, Long Lian, Zhongqi Miao, Ziwei Liu, Stella Yu</i></p> | 3501 |

| | |
|---|------|
| <p>Generalization in Data-Driven Models of Primary Visual Cortex</p> <p><i>Konstantin-Klemens Lurz, Mohammad Bashiri, Konstantin Willeke, Akshay Jagadish, Eric Wang, Edgar Y. Walker, Santiago A Cadena, Taliah Muhammad, Erick Cobos, Andreas S. Tolias, Alexander S Ecker, Fabian H. Sinz</i></p> | 3516 |
| <p>Sequential Density Ratio Estimation for Simultaneous Optimization of Speed and Accuracy</p> <p><i>Akinori F Ebihara, Taiki Miyagawa, Kazuyuki Sakurai, Hitoshi Imaoka</i></p> | 3528 |
| <p>Uncertainty Sets for Image Classifiers Using Conformal Prediction</p> <p><i>Anastasios Nikolas Angelopoulos, Stephen Bates, Michael Jordan, Jitendra Malik</i></p> | 3593 |
| <p>Graph Convolution with Low-Rank Learnable Local Filters</p> <p><i>Xiuyuan Cheng, Zichen Miao, Qiang Qiu</i></p> | 3610 |
| <p>Mind the Pad -- CNNs Can Develop Blind Spots</p> <p><i>Bilal Alsallakh, Narine Kokhlikyan, Vivek Miglani, Jun Yuan, Orion Reblitz-Richardson</i></p> | 3635 |
| <p>Stabilized Medical Image Attacks</p> <p><i>Gege Qi, Lijun Gong, Yibing Song, Kai Ma, Yefeng Zheng</i></p> | 3671 |
| <p>Net-DNF: Effective Deep Modeling of Tabular Data</p> <p><i>Liran Katzir, Gal Elidan, Ran El-Yaniv</i></p> | 3683 |
| <p>Predicting Inductive Biases of Pre-Trained Models</p> <p><i>Charles Lovering, Rohan Jha, Tal Linzen, Ellie Pavlick</i></p> | 3699 |
| <p>Optimism in Reinforcement Learning with Generalized Linear Function Approximation</p> <p><i>Yining Wang, Ruosong Wang, Simon Shaolei Du, Akshay Krishnamurthy</i></p> | 3726 |
| <p>SCoRe: Pre-Training for Context Representation in Conversational Semantic Parsing</p> <p><i>Tao Yu, Rui Zhang, Alex Polozov, Christopher Meek, Ahmed Hassan Awadallah</i></p> | 3743 |
| <p>A Teacher-Student Framework to Distill Future Trajectories</p> <p><i>Alexander Neitz, Giambattista Parascandolo, Bernhard Schölkopf</i></p> | 3759 |
| <p>Certify Or Predict: Boosting Certified Robustness with Compositional Architectures</p> <p><i>Mark Niklas Mueller, Mislav Balunovic, Martin Vechev</i></p> | 3778 |
| <p>On the Transfer of Disentangled Representations in Realistic Settings</p> <p><i>Andrea Dittadi, Frederik Träuble, Francesco Locatello, Manuel Wuthrich, Vaibhav Agrawal, Ole Winther, Stefan Bauer, Bernhard Schölkopf</i></p> | 3800 |
| <p>Robust Reinforcement Learning on State Observations with Learned Optimal Adversary</p> <p><i>Huan Zhang, Hongge Chen, Duane S Boning, Cho-Jui Hsieh</i></p> | 3822 |
| <p>Learning from Others' Mistakes: Avoiding Dataset Biases Without Modeling Them</p> <p><i>Victor Sanh, Thomas Wolf, Yonatan Belinkov, Alexander M Rush</i></p> | 3838 |

| | |
|--|------|
| <p>Fast and Complete: Enabling Complete Neural Network Verification with Rapid and Massively Parallel Incomplete Verifiers</p> <p><i>Kaidi Xu, Huan Zhang, Shiqi Wang, Yihan Wang, Suman Jana, Xue Lin, Cho-Jui Hsieh</i></p> | 3856 |
| <p>Self-Supervised Adversarial Robustness for the Low-Label, High-Data Regime</p> <p><i>Sven Gowal, Po-Sen Huang, Aaron Van Den Oord, Timothy Mann, Pushmeet Kohli</i></p> | 3871 |
| <p>Modeling the Second Player in Distributionally Robust Optimization</p> <p><i>Paul Michel, Tatsunori Hashimoto, Graham Neubig</i></p> | 3890 |
| <p>Neural Jump Ordinary Differential Equations: Consistent Continuous-Time Prediction and Filtering</p> <p><i>Calypso Herrera, Florian Krach, Josef Teichmann</i></p> | 3906 |
| <p>Gradient Origin Networks</p> <p><i>Sam Bond-Taylor, Chris G. Willcocks</i></p> | 3950 |
| <p>Efficient Generalized Spherical CNNs</p> <p><i>Oliver Cobb, Christopher G. R. Wallis, Augustine N. Mavor-Parker, Augustin Marignier, Matthew A. Price, Mayeul D'Avezac, Jason McEwen</i></p> | 3966 |
| <p>DEBERTA: DECODING-ENHANCED BERT WITH DISENTANGLED ATTENTION</p> <p><i>Pengcheng He, Xiaodong Liu, Jianfeng Gao, Weizhu Chen</i></p> | 3986 |
| <p>Optimizing Memory Placement Using Evolutionary Graph Reinforcement Learning</p> <p><i>Shauharda Khadka, Estelle Aflalo, Mattias Marder, Avrech Ben-David, Santiago Miret, Shie Mannor, Tamir Hazan, Hanlin Tang, Somdeb Majumdar</i></p> | 4007 |
| <p>On the Geometry of Generalization and Memorization in Deep Neural Networks</p> <p><i>Cory Stephenson, Suchismita Padhy, Abhinav Ganesh, Yue Hui, Hanlin Tang, Sueyeon Chung</i></p> | 4024 |
| <p>Continual Learning in Recurrent Neural Networks</p> <p><i>Benjamin Ehret, Christian Henning, Maria Cervera, Alexander Meulemans, Johannes Von Oswald, Benjamin F Grewe</i></p> | 4049 |
| <p>Witches' Brew: Industrial Scale Data Poisoning Via Gradient Matching</p> <p><i>Jonas Geiping, Liam H Fowl, W. Ronny Huang, Wojciech Czaja, Gavin Taylor, Michael Moeller, Tom Goldstein</i></p> | 4063 |
| <p>Overfitting for Fun and Profit: Instance-Adaptive Data Compression</p> <p><i>Ties Van Rozendaal, Iris Am Huijben, Taco Cohen</i></p> | 4087 |
| <p>A Block Minifloat Representation for Training Deep Neural Networks</p> <p><i>Sean Fox, Seyedramin Rasoulnezhad, Julian Faraone, David Boland, Philip Leong</i></p> | 4105 |
| <p>Representation Learning Via Invariant Causal Mechanisms</p> <p><i>Jovana Mitrovic, Brian McWilliams, Jacob C Walker, Lars Holger Buesing, Charles Blundell</i></p> | 4121 |

| | |
|--|------|
| <p>Sparse Encoding for More-Interpretable Feature-Selecting Representations in Probabilistic Matrix Factorization</p> <p><i>Joshua C Chang, Patrick Fletcher, Jungmin Han, Ted L Chang, Shashaank Vattikuti, Bart Desmet, Ayah Zirikly, Carson C Chow</i></p> | 4133 |
| <p>Mapping the Timescale Organization of Neural Language Models</p> <p><i>Hsiang-Yun Sherry Chien, Jinhan Zhang, Christopher Honey</i></p> | 4148 |
| <p>Neural Networks with Late-Phase Weights</p> <p><i>Johannes Von Oswald, Seijin Kobayashi, Joao Sacramento, Alexander Meulemans, Christian Henning, Benjamin F Grewe</i></p> | 4171 |
| <p>Uncertainty-Aware Active Learning for Optimal Bayesian Classifier</p> <p><i>Guang Zhao, Edward Dougherty, Byung-Jun Yoon, Francis Alexander, Xiaoning Qian</i></p> | 4196 |
| <p>ResNet After All: Neural ODEs and Their Numerical Solution</p> <p><i>Katharina Ott, Prateek Katiyar, Philipp Hennig, Michael Tiemann</i></p> | 4216 |
| <p>Generalized Variational Continual Learning</p> <p><i>Noel Loo, Siddharth Swaroop, Richard E Turner</i></p> | 4246 |
| <p>Mind the Gap When Conditioning Amortised Inference in Sequential Latent-Variable Models</p> <p><i>Justin Bayer, Maximilian Soelch, Atanas Mirchev, Baris Kayalibay, Patrick Van Der Smagt</i></p> | 4292 |
| <p>CausalWorld: A Robotic Manipulation Benchmark for Causal Structure and Transfer Learning</p> <p><i>Ossama Ahmed, Frederik Träuble, Anirudh Goyal, Alexander Neitz, Manuel Wuthrich, Yoshua Bengio, Bernhard Schölkopf, Stefan Bauer</i></p> | 4315 |
| <p>Transformer Protein Language Models Are Unsupervised Structure Learners</p> <p><i>Roshan Rao, Joshua Meier, Tom Sercu, Sergey Ovchinnikov, Alexander Rives</i></p> | 4335 |
| <p>Neural ODE Processes</p> <p><i>Alexander Norcliffe, Cristian Bodnar, Ben Day, Jacob Moss, Pietro Liò</i></p> | 4359 |
| <p>The Role of Disentanglement in Generalisation</p> <p><i>Milton Llera Montero, Casimir Jh Ludwig, Rui Ponte Costa, Gaurav Malhotra, Jeffrey Bowers</i></p> | 4378 |
| <p>Learning to Live with Dale's Principle: ANNs with Separate Excitatory and Inhibitory Units</p> <p><i>Jonathan Cornford, Damjan Kalajdzievski, Marco Leite, Amélie Lamarquette, Dimitri Michael Kullmann, Blake Aaron Richards</i></p> | 4395 |
| <p>SALD: Sign Agnostic Learning with Derivatives</p> <p><i>Matan Atzmon, Yaron Lipman</i></p> | 4422 |
| <p>Ringing ReLUs: Harmonic Distortion Analysis of Nonlinear Feedforward Networks</p> <p><i>Christian H. X. Ali Mehmeti-Göpel, David Hartmann, Michael Wand</i></p> | 4439 |
| <p>CoCon: A Self-Supervised Approach for Controlled Text Generation</p> <p><i>Alvin Chan, Yew-Soon Ong, Bill Pung, Aston Zhang, Jie Fu</i></p> | 4461 |

| | |
|--|------|
| <p>Bidirectional Variational Inference for Non-Autoregressive Text-To-Speech</p> <p>Yoonhyung Lee, Joongbo Shin, Kyomin Jung</p> | 4483 |
| <p>Learning Continuous-Time PDEs from Sparse Data with Graph Neural Networks</p> <p>Valerii Iakovlev, Markus Heinonen, Harri Lähdesmäki</p> | 4502 |
| <p>NAS-Bench-ASR: Reproducible Neural Architecture Search for Speech Recognition</p> <p>Abhinav Mehrotra, Alberto Gil C. P. Ramos, Sourav Bhattacharya, Lukasz Dudziak, Ravichander Vipperla, Thomas Chau, Mohamed S Abdelfattah, Samin Ishtiaq, Nicholas Donald Lane</p> | 4517 |
| <p>Collective Robustness Certificates: Exploiting Interdependence in Graph Neural Networks</p> <p>Jan Schuchardt, Aleksandar Bojchevski, Johannes Gasteiger, Stephan Günnemann</p> | 4532 |
| <p>Adversarially Guided Actor-Critic</p> <p>Yannis Flet-Berliac, Johan Ferret, Olivier Pietquin, Philippe Preux, Matthieu Geist</p> | 4552 |
| <p>Training Independent Subnetworks for Robust Prediction</p> <p>Marton Havasi, Rodolphe Jenatton, Stanislav Fort, Jeremiah Zhe Liu, Jasper Snoek, Balaji Lakshminarayanan, Andrew Mingbo Dai, Dustin Tran</p> | 4570 |
| <p>Grounding Language to Autonomously-Acquired Skills Via Goal Generation</p> <p>Ahmed Akakzia, Cédric Colas, Pierre-Yves Oudeyer, Mohamed Chetouani, Olivier Sigaud</p> | 4583 |
| <p>Hopfield Networks is All You Need</p> <p>Hubert Ramsauer, Bernhard Schöfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Lukas Gruber, Markus Holzleitner, Thomas Adler, David Kreil, Michael K Kopp, Günter Klambauer, Johannes Brandstetter, Sepp Hochreiter</p> | 4604 |
| <p>Differentiable Trust Region Layers for Deep Reinforcement Learning</p> <p>Fabian Otto, Philipp Becker, Vien Anh Ngo, Hanna Carolin Maria Ziesche, Gerhard Neumann</p> | 4699 |
| <p>Temporally-Extended ϵ-Greedy Exploration</p> <p>Will Dabney, Georg Ostrovski, Andre Barreto</p> | 4721 |
| <p>Learning Associative Inference Using Fast Weight Memory</p> <p>Imanol Schlag, Tsendsuren Munkhdalai, Jürgen Schmidhuber</p> | 4750 |
| <p>Multiscale Score Matching for Out-Of-Distribution Detection</p> <p>Ahsan Mahmood, Junier Oliva, Martin Andreas Styner</p> | 4778 |
| <p>Learning to Sample with Local and Global Contexts in Experience Replay Buffer</p> <p>Youngmin Oh, Kimin Lee, Jinwoo Shin, Eunho Yang, Sung Ju Hwang</p> | 4795 |
| <p>Parameter-Based Value Functions</p> <p>Francesco Faccio, Louis Kirsch, Jürgen Schmidhuber</p> | 4814 |
| <p>New Bounds for Distributed Mean Estimation and Variance Reduction</p> <p>Peter Davies, Vijaykrishna Gurunathan, Niusha Moshrefi, Saleh Ashkboos, Dan Alistarh</p> | 4850 |

| | |
|--|------|
| <p>Learning to Set Waypoints for Audio-Visual Navigation</p> <p><i>Changan Chen, Sagnik Majumder, Ziad Al-Halah, Ruohan Gao, Santhosh Kumar Ramakrishnan, Kristen Grauman</i></p> | 4861 |
| <p>Disambiguating Symbolic Expressions in Informal Documents</p> <p><i>Dennis Müller, Cezary Kaliszyk</i></p> | 4877 |
| <p>Colorization Transformer</p> <p><i>Manoj Kumar, Dirk Weissenborn, Nal Kalchbrenner</i></p> | 4893 |
| <p>Theoretical Bounds on Estimation Error for Meta-Learning</p> <p><i>James Lucas, Mengye Ren, Irene Raissa Kameni Kameni, Toniann Pitassi, Richard Zemel</i></p> | 4917 |
| <p>Variational Information Bottleneck for Effective Low-Resource Fine-Tuning</p> <p><i>Rabeeh Karimi Mahabadi, Yonatan Belinkov, James Henderson</i></p> | 4944 |
| <p>TropEx: An Algorithm for Extracting Linear Terms in Deep Neural Networks</p> <p><i>Martin Trimmel, Henning Petzka, Cristian Sminchisescu</i></p> | 4957 |
| <p>Seq2Tens: An Efficient Representation of Sequences by Low-Rank Tensor Projections</p> <p><i>Csaba Toth, Patric Bonnier, Harald Oberhauser</i></p> | 4968 |
| <p>Representation Learning for Improved Interpretability and Classification Accuracy of Clinical Factors from EEG</p> <p><i>Garrett Honke, Irina Higgins, Nina Thigpen, Vladimir Miskovic, Katie Link, Sunny Duan, Pramod Gupta, Julia Klawohn, Greg Hajcak</i></p> | 4982 |
| <p>Language-Agnostic Representation Learning of Source Code from Structure and Context</p> <p><i>Daniel Zügner, Tobias Kirschstein, Michele Catasta, Jure Leskovec, Stephan Günnemann</i></p> | 4995 |
| <p>Generalized Multimodal ELBO</p> <p><i>Thomas M. Sutter, Imant Daunhawer, Julia E Vogt</i></p> | 5017 |
| <p>Model-Based Micro-Data Reinforcement Learning: What Are the Crucial Model Properties and Which Model to Choose?</p> <p><i>Balázs Kégl, Gabriel Hurtado, Albert Thomas</i></p> | 5040 |
| <p>Set Prediction Without Imposing Structure as Conditional Density Estimation</p> <p><i>David W Zhang, Gertjan J. Burghouts, Cees G. M. Snoek</i></p> | 5061 |
| <p>Learning Value Functions in Deep Policy Gradients Using Residual Variance</p> <p><i>Yannis Flet-Berliac, Reda Ouhamma, Odalric-Ambrym Maillard, Philippe Preux</i></p> | 5079 |
| <p>IDF++: Analyzing and Improving Integer Discrete Flows for Lossless Compression</p> <p><i>Rianne Van Den Berg, Alexey A. Gritsenko, Mostafa Dehghani, Casper Kaae Sønderby, Tim Salimans</i></p> | 5098 |
| <p>Fully Unsupervised Diversity Denoising with Convolutional Variational Autoencoders</p> <p><i>Mangal Prakash, Alexander Krull, Florian Jug</i></p> | 5117 |

| | |
|--|------|
| <i>Is Attention Better than Matrix Decomposition?</i> | 5161 |
| <i>Zhengyang Geng, Meng-Hao Guo, Hongxu Chen, Xia Li, Ke Wei, Zhouchen Lin</i> | |
| <i>Improving Transformation Invariance in Contrastive Representation Learning</i> | 5181 |
| <i>Adam Foster, Rattana Pukdee, Tom Rainforth</i> | |
| <i>On the Origin of Implicit Regularization in Stochastic Gradient Descent</i> | 5203 |
| <i>Samuel L Smith, Benoit Dherin, David Barrett, Soham De</i> | |
| <i>Transient Non-Stationarity and Generalisation in Deep Reinforcement Learning</i> | 5217 |
| <i>Maximilian Igl, Gregory Farquhar, Jelena Luketina, Wendelin Boehmer, Shimon Whiteson</i> | |
| <i>Lossless Compression of Structured Convolutional Models Via Lifting</i> | 5233 |
| <i>Gustav Sourek, Filip Zelezny, Ondrej Kuzelka</i> | |
| <i>Analyzing the Expressive Power of Graph Neural Networks in a Spectral Perspective</i> | 5244 |
| <i>Muhammet Balçilar, Guillaume Renton, Pierre H eroux, Benoit Ga uz ere, S ebastien Adam, Paul Honeine</i> | |
| <i>A Unifying View on Implicit Bias in Training Linear Neural Networks</i> | 5270 |
| <i>Chulhee Yun, Shankar Krishnan, Hossein Mobahi</i> | |
| <i>Balancing Constraints and Rewards with Meta-Gradient D4PG</i> | 5306 |
| <i>Dan A. Calian, Daniel J Mankowitz, Tom Zahavy, Zhongwen Xu, Junhyuk Oh, Nir Levine, Timothy Mann</i> | |
| <i>Robust Curriculum Learning: From Clean Label Detection to Noisy Label Self-Correction</i> | 5316 |
| <i>Tianyi Zhou, Shengjie Wang, Jeff Bilmes</i> | |
| <i>Clairvoyance: A Pipeline Toolkit for Medical Time Series</i> | 5334 |
| <i>Daniel Jarrett, Jinsung Yoon, Ioana Bica, Zhaozhi Qian, Ari Ercole, Mihaela Van Der Schaar</i> | |
| <i>Plan-Based Relaxed Reward Shaping for Goal-Directed Tasks</i> | 5366 |
| <i>Ingmar Schubert, Ozgur S Oguz, Marc Toussaint</i> | |
| <i>Improving VAEs' Robustness to Adversarial Attack</i> | 5383 |
| <i>Matthew Jf Willetts, Alexander Camuto, Tom Rainforth, S Roberts, Christopher C Holmes</i> | |
| <i>Differentiable Segmentation of Sequences</i> | 5409 |
| <i>Erik Scharw achter, Jonathan Lennartz, Emmanuel M uller</i> | |
| <i>GraPPa: Grammar-Augmented Pre-Training for Table Semantic Parsing</i> | 5426 |
| <i>Tao Yu, Chien-Sheng Wu, Xi Victoria Lin, Bailin Wang, Yi Chern Tan, Xinyi Yang, Dragomir Radev, Richard Socher, Caiming Xiong</i> | |
| <i>Sliced Kernelized Stein Discrepancy</i> | 5442 |
| <i>Wenbo Gong, Yingzhen Li, Jos e Miguel Hern andez-Lobato</i> | |

| | |
|---|------|
| Variational Intrinsic Control Revisited <i>Taehwan Kwon</i> | 5454 |
| HyperDynamics: Meta-Learning Object and Agent Dynamics with Hypernetworks <i>Zhou Xian, Shamit Lal, Hsiao-Yu Tung, Emmanouil Antonios Platanios, Katerina Fragkiadaki</i> | 5469 |
| Towards Resolving the Implicit Bias of Gradient Descent for Matrix Factorization: Greedy Low-Rank Learning <i>Zhiyuan Li, Yuping Luo, Kaifeng Lyu</i> | 5485 |
| Private Post-GAN Boosting <i>Marcel Neunhoffer, Steven Wu, Cynthia Dwork</i> | 5533 |
| Characterizing Signal Propagation to Close the Performance Gap in Unnormalized ResNets <i>Andrew Brock, Soham De, Samuel L Smith</i> | 5550 |
| Prototypical Contrastive Learning of Unsupervised Representations <i>Junnan Li, Pan Zhou, Caiming Xiong, Steven Hoi</i> | 5576 |
| Hyperbolic Neural Networks++ <i>Ryohei Shimizu, Yusuke Mukuta, Tatsuya Harada</i> | 5592 |
| Lipschitz Recurrent Neural Networks <i>N. Benjamin Erichson, Omri Azencot, Alejandro Queiruga, Liam Hodgkinson, Michael W. Mahoney</i> | 5617 |
| A Statistical Theory of Cold Posteriors in Deep Neural Networks <i>Laurence Aitchison</i> | 5638 |
| Boost Then Convolve: Gradient Boosting Meets Graph Neural Networks <i>Sergei Ivanov, Liudmila Prokhorenkova</i> | 5653 |
| Genetic Soft Updates for Policy Evolution in Deep Reinforcement Learning <i>Enrico Marchesini, Davide Corsi, Alessandro Farinelli</i> | 5669 |
| Spatially Structured Recurrent Modules <i>Nasim Rahaman, Anirudh Goyal, Muhammad Waleed Gondal, Manuel Wuthrich, Stefan Bauer, Yash Sharma, Yoshua Bengio, Bernhard Schölkopf</i> | 5684 |
| On the Stability of Fine-Tuning BERT: Misconceptions, Explanations, and Strong Baselines <i>Marius Mosbach, Maksym Andriushchenko, Dietrich Klakow</i> | 5716 |
| End-To-End Egospheric Spatial Memory <i>Daniel James Lenton, Stephen James, Ronald Clark, Andrew Davison</i> | 5735 |
| LEAF: A Learnable Frontend for Audio Classification <i>Neil Zeghidour, Olivier Teboul, Félix De Chaumont Quiry, Marco Tagliasacchi</i> | 5756 |
| Simple Augmentation Goes a Long Way: ADRL for DNN Quantization <i>Lin Ning, Guoyang Chen, Weifeng Zhang, Xipeng Shen</i> | 5772 |

| | |
|--|------|
| The Inductive Bias of ReLU Networks on Orthogonally Separable Data <i>Mary Phuong, Christoph H Lampert</i> | 5784 |
| Monte-Carlo Planning and Learning with Language Action Value Estimates <i>Youngsoo Jang, Seokin Seo, Jongmin Lee, Kee-Eung Kim</i> | 5803 |
| Learning Energy-Based Models by Diffusion Recovery Likelihood <i>Ruiqi Gao, Yang Song, Ben Poole, Ying Nian Wu, Diederik P Kingma</i> | 5821 |
| Capturing Label Characteristics in VAEs <i>Tom Joy, Sebastian Schmon, Philip Torr, Siddharth N, Tom Rainforth</i> | 5846 |
| Linear Mode Connectivity in Multitask and Continual Learning <i>Seyed Iman Mirzadeh, Mehrdad Farajtabar, Dilan Gorur, Razvan Pascanu, Hassan Ghasemzadeh</i> | 5876 |
| Computational Separation Between Convolutional and Fully-Connected Networks <i>Eran Malach, Shai Shalev-Shwartz</i> | 5898 |
| Rethinking Embedding Coupling in Pre-Trained Language Models <i>Hyung Won Chung, Thibault Fevry, Henry Tsai, Melvin Johnson, Sebastian Ruder</i> | 5910 |
| Physics-Aware, Probabilistic Model Order Reduction with Guaranteed Stability <i>Sebastian Kaltenbach, Phaedon Stelios Koutsourelakis</i> | 5927 |
| Disentangling 3D Prototypical Networks for Few-Shot Concept Learning <i>Mihir Prabhudesai, Shamit Lal, Darshan Patil, Hsiao-Yu Tung, Adam W Harley, Katerina Fragkiadaki</i> | 5952 |
| LiftPool: Bidirectional ConvNet Pooling <i>Jiaojiao Zhao, Cees G. M. Snoek</i> | 5975 |
| Latent Convergent Cross Mapping <i>Edward De Brouwer, Adam Arany, Jaak Simm, Yves Moreau</i> | 5990 |
| You Only Need Adversarial Supervision for Semantic Image Synthesis <i>Edgar Schönfeld, Vadim Sushko, Dan Zhang, Juergen Gall, Bernt Schiele, Anna Khoreva</i> | 6002 |
| A Diffusion Theory for Deep Learning Dynamics: Stochastic Gradient Descent Exponentially Favors Flat Minima <i>Zeke Xie, Issei Sato, Masashi Sugiyama</i> | 6034 |
| Robust Learning of Fixed-Structure Bayesian Networks in Nearly-Linear Time <i>Yu Cheng, Honghao Lin</i> | 6062 |
| Activation-Level Uncertainty in Deep Neural Networks <i>Pablo Morales-Alvarez, Daniel Hernández-Lobato, Rafael Molina, José Miguel Hernández-Lobato</i> | 6085 |
| SkipW: Resource Adaptable RNN with Strict Upper Computational Limit <i>Tsiry Mayet, Anne Lambert, Pascal Leguyadec, Françoise Le Bolzer, François Schnitzler</i> | 6106 |

| | |
|--|------|
| Wasserstein-2 Generative Networks <i>Alexander Korotin, Vage Egiazarian, Arip Asadulaev, Alexander Safin, Evgeny Burnaev</i> | 6133 |
| Group Equivariant Stand-Alone Self-Attention for Vision <i>David W. Romero, Jean-Baptiste Cordonnier</i> | 6163 |
| Continuous Wasserstein-2 Barycenter Estimation Without Minimax Optimization <i>Alexander Korotin, Lingxiao Li, Justin Solomon, Evgeny Burnaev</i> | 6185 |
| RNNLogic: Learning Logic Rules for Reasoning on Knowledge Graphs <i>Meng Qu, Junkun Chen, Louis-Pascal Xhonneux, Yoshua Bengio, Jian Tang</i> | 6203 |
| Selective Classification Can Magnify Disparities Across Groups <i>Erik Jones, Shiori Sagawa, Pang Wei Koh, Ananya Kumar, Percy Liang</i> | 6223 |
| FedMix: Approximation of Mixup Under Mean Augmented Federated Learning <i>Tehrim Yoon, Sumin Shin, Sung Ju Hwang, Eunho Yang</i> | 6257 |
| In-N-Out: Pre-Training and Self-Training Using Auxiliary Information for Out-Of-Distribution Robustness <i>Sang Michael Xie, Ananya Kumar, Robbie Jones, Fereshte Khani, Tengyu Ma, Percy Liang</i> | 6276 |
| Sample-Efficient Automated Deep Reinforcement Learning <i>Jörg K. H. Franke, Gregor Koehler, André Biedenkapp, Frank Hutter</i> | 6289 |
| A Temporal Kernel Approach for Deep Learning with Continuous-Time Information <i>Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, Kannan Achan</i> | 6312 |
| Convex Regularization Behind Neural Reconstruction <i>Arda Sahiner, Morteza Mardani, Batu Ozturkler, Mert Pilanci, John M. Pauly</i> | 6345 |
| Vector-Output ReLU Neural Network Problems Are Copositive Programs: Convex Analysis of Two Layer Networks and Polynomial-Time Algorithms <i>Arda Sahiner, Tolga Ergen, John M. Pauly, Mert Pilanci</i> | 6364 |
| Learning Better Structured Representations Using Low-Rank Adaptive Label Smoothing <i>Asish Ghoshal, Xilun Chen, Sonal Gupta, Luke Zettlemoyer, Yashar Mehdad</i> | 6388 |
| Training GANs with Stronger Augmentations Via Contrastive Discriminator <i>Jongheon Jeong, Jinwoo Shin</i> | 6403 |
| Private Image Reconstruction from System Side Channels Using Generative Models <i>Yuanyuan Yuan, Shuai Wang, Junping Zhang</i> | 6426 |
| Learning to Make Decisions Via Submodular Regularization <i>Ayya Alieva, Aiden Aceves, Jialin Song, Stephen Mayo, Yisong Yue, Yuxin Chen</i> | 6458 |
| The Recurrent Neural Tangent Kernel <i>Sina Alemohammad, Zichao Wang, Randall Balestriero, Richard Baraniuk</i> | 6475 |

| | |
|---|------|
| Evaluation of Similarity-Based Explanations <i>Kazuaki Hanawa, Sho Yokoi, Satoshi Hara, Kentaro Inui</i> | 6501 |
| Adaptive Procedural Task Generation for Hard-Exploration Problems <i>Kuan Fang, Yuke Zhu, Silvio Savarese, L. Fei-Fei</i> | 6527 |
| Linear Last-Iterate Convergence in Constrained Saddle-Point Optimization <i>Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, Haipeng Luo</i> | 6547 |
| On Graph Neural Networks Versus Graph-Augmented MLPs <i>Lei Chen, Zhengdao Chen, Joan Bruna</i> | 6586 |
| Solving Compositional Reinforcement Learning Problems Via Task Reduction <i>Yunfei Li, Yilin Wu, Huazhe Xu, Xiaolong Wang, Yi Wu</i> | 6610 |
| Conditional Generative Modeling Via Learning the Latent Space <i>Sameera Ramasinghe, Kanchana Nisal Ranasinghe, Salman Khan, Nick Barnes, Stephen Gould</i> | 6631 |
| DialoGraph: Incorporating Interpretable Strategy-Graph Networks into Negotiation Dialogues <i>Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan Black, Yulia Tsvetkov</i> | 6645 |
| WaNet - Imperceptible Warping-Based Backdoor Attack <i>Tuan Anh Nguyen, Anh Tuan Tran</i> | 6667 |
| Nonseparable Symplectic Neural Networks <i>Shiyong Xiong, Yunjin Tong, Xingzhe He, Shuqi Yang, Cheng Yang, Bo Zhu</i> | 6683 |
| Discovering Diverse Multi-Agent Strategic Behavior Via Reward Randomization <i>Zhenggang Tang, Chao Yu, Boyuan Chen, Huazhe Xu, Xiaolong Wang, Fei Fang, Simon Shaolei Du, Yu Wang, Yi Wu</i> | 6697 |
| Multi-Timescale Representation Learning in LSTM Language Models <i>Shivangi Mahto, Vy Ai Vo, Javier S. Turek, Alexander Huth</i> | 6723 |
| Explaining the Efficacy of Counterfactually Augmented Data <i>Divyansh Kaushik, Amrith Setlur, Eduard H Hovy, Zachary Chase Lipton</i> | 6742 |
| Revisiting Locally Supervised Learning: An Alternative to End-To-End Training <i>Yulin Wang, Zanlin Ni, Shiji Song, Le Yang, Gao Huang</i> | 6775 |
| How Much Over-Parameterization is Sufficient to Learn Deep ReLU Networks? <i>Zixiang Chen, Yuan Cao, Difan Zou, Quanquan Gu</i> | 6796 |
| Blending MPC & Value Function Approximation for Efficient Reinforcement Learning <i>Mohak Bhardwaj, Sanjiban Choudhury, Byron Boots</i> | 6817 |
| Probabilistic Numeric Convolutional Neural Networks <i>Marc Anton Finzi, Roberto Bondesan, Max Welling</i> | 6832 |

| | |
|--|------|
| <p>HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients</p> <p><i>Enmao Diao, Jie Ding, Vahid Tarokh</i></p> | 6855 |
| <p>Semantic Re-Tuning with Contrastive Tension</p> <p><i>Fredrik Carlsson, Amaru Cuba Gyllensten, Evangelia Gogoulou, Erik Ylipää Hellqvist, Magnus Sahlgren</i></p> | 6879 |
| <p>Dataset Meta-Learning from Kernel Ridge-Regression</p> <p><i>Timothy Nguyen, Zhouong Chen, Jaehoon Lee</i></p> | 6900 |
| <p>AUXILIARY TASK UPDATE DECOMPOSITION: THE GOOD, THE BAD AND THE NEUTRAL</p> <p><i>Lucio M. Dery, Yann Dauphin, David Grangier</i></p> | 6925 |
| <p>Fast and Slow Learning of Recurrent Independent Mechanisms</p> <p><i>Kanika Madan, Nan Rosemary Ke, Anirudh Goyal, Bernhard Schölkopf, Yoshua Bengio</i></p> | 6940 |
| <p>Auction Learning as a Two-Player Game</p> <p><i>Jad Rahme, Samy Jelassi, S. Matthew Weinberg</i></p> | 6957 |
| <p>A PAC-Bayesian Approach to Generalization Bounds for Graph Neural Networks</p> <p><i>Renjie Liao, Raquel Urtasun, Richard Zemel</i></p> | 6973 |
| <p>Contextual Transformation Networks for Online Continual Learning</p> <p><i>Quang Pham, Chenghao Liu, Doyen Sahoo, Steven Hoi</i></p> | 6985 |
| <p>Adaptive and Generative Zero-Shot Learning</p> <p><i>Yu-Ying Chou, Hsuan-Tien Lin, Tyng-Luh Liu</i></p> | 7005 |
| <p>Online Adversarial Purification Based on Self-Supervised Learning</p> <p><i>Changhao Shi, Chester Holtz, Gal Mishne</i></p> | 7019 |
| <p>FairFil: Contrastive Neural Debiasing Method for Pretrained Text Encoders</p> <p><i>Pengyu Cheng, Weituo Hao, Siyang Yuan, Shijing Si, Lawrence Carin</i></p> | 7034 |
| <p>Reset-Free Lifelong Learning with Skill-Space Planning</p> <p><i>Kevin Lu, Aditya Grover, Pieter Abbeel, Igor Mordatch</i></p> | 7046 |
| <p>Efficient Empowerment Estimation for Unsupervised Stabilization</p> <p><i>Ruihan Zhao, Kevin Lu, Pieter Abbeel, Stas Tiomkin</i></p> | 7066 |
| <p>MixKD: Towards Efficient Distillation of Large-Scale Language Models</p> <p><i>Kevin J Liang, Weituo Hao, Dinghan Shen, Yufan Zhou, Weizhu Chen, Changyou Chen, Lawrence Carin</i></p> | 7083 |
| <p>CaPC Learning: Confidential and Private Collaborative Learning</p> <p><i>Christopher A. Choquette-Choo, Natalie Dullerud, Adam Dziedzic, Yunxiang Zhang, Somesh Jha, Nicolas Papernot, Xiao Wang</i></p> | 7099 |

| | |
|---|------|
| Multiplicative Filter Networks Rizal Fathony, Anit Kumar Sahu, Devin Willmott, J Zico Kolter | 7122 |
| Planning from Pixels Using Inverse Dynamics Models Keiran Paster, Sheila A. McIlraith, Jimmy Ba | 7133 |
| Semi-Supervised Keypoint Localization Olga Moskvayak, Frederic Maire, Feras Dayoub, Mahsa Baktashmotlagh | 7151 |
| Emergent Road Rules in Multi-Agent Driving Environments Avik Pal, Jonah Philion, Yuan-Hong Liao, Sanja Fidler | 7162 |
| SSD: A Unified Framework for Self-Supervised Outlier Detection Vikash Sehwal, Mung Chiang, Prateek Mittal | 7174 |
| ECONOMIC HYPERPARAMETER OPTIMIZATION WITH BLENDED SEARCH STRATEGY Chi Wang, Qingyun Wu, Silu Huang, Amin Saied | 7191 |
| Inductive Representation Learning in Temporal Networks Via Causal Anonymous Walks Yanbang Wang, Yen-Yu Chang, Yunyu Liu, Jure Leskovec, Pan Li | 7208 |
| Robust Overfitting May Be Mitigated by Properly Learned Smoothing Tianlong Chen, Zhenyu Zhang, Sijia Liu, Shiyu Chang, Zhangyang Wang | 7229 |
| Local Search Algorithms for Rank-Constrained Convex Optimization Kyriakos Axiotis, Maxim Sviridenko | 7248 |
| Learning Task Decomposition with Ordered Memory Policy Network Yuchen Lu, Yikang Shen, Siyuan Zhou, Aaron Courville, Joshua B. Tenenbaum, Chuang Gan | 7267 |
| Property Controllable Variational Autoencoder Via Invertible Mutual Dependence Xiaojie Guo, Yuanqi Du, Liang Zhao | 7288 |
| Grounding Physical Concepts of Objects and Events Through Dynamic Visual Reasoning Zhenfang Chen, Jiayuan Mao, Jiajun Wu, Kwan-Yee Kenneth Wong, Joshua B. Tenenbaum, Chuang Gan | 7308 |
| GradSim: Differentiable Simulation for System Identification and Visuomotor Control J. Krishna Murthy, Miles Macklin, Florian Golemo, Vikram Voleti, Linda Petrini, Martin Weiss, Breandan Considine, Jérôme Parent-Lévesque, Kevin Xie, Kenny Erleben, Liam Paull, Florian Shkurti, Derek Nowrouzezahrai, Sanja Fidler | 7328 |
| Generative Scene Graph Networks Fei Deng, Zhuo Zhi, Donghun Lee, Sungjin Ahn | 7353 |
| Decentralized Attribution of Generative Models Changhoon Kim, Yi Ren, Yezhou Yang | 7376 |

| | |
|---|------|
| <p>Individually Fair Rankings</p> <p><i>Amanda Bower, Hamid Eftekhari, Mikhail Yurochkin, Yuekai Sun</i></p> | 7392 |
| <p>Adaptive Federated Optimization</p> <p><i>Sashank J. Reddi, Zachary Charles, Manzil Zaheer, Zachary Garrett, Keith Rush, Jakub Konecný, Sanjiv Kumar, Hugh Brendan McMahan</i></p> | 7408 |
| <p>GANs Can Play Lottery Tickets Too</p> <p><i>Xuxi Chen, Zhenyu Zhang, Yongduo Sui, Tianlong Chen</i></p> | 7446 |
| <p>Improving Relational Regularized Autoencoders with Spherical Sliced Fused Gromov Wasserstein</p> <p><i>Khai Nguyen, Son Nguyen, Nhat Ho, Tung Pham, Hung Bui</i></p> | 7461 |
| <p>Learning Reasoning Paths Over Semantic Graphs for Video-Grounded Dialogues</p> <p><i>Hung Le, Nancy F. Chen, Steven Hoi</i></p> | 7472 |
| <p>Extreme Memorization Via Scale of Initialization</p> <p><i>Harsh Mehta, Ashok Cutkosky, Behnam Neyshabur</i></p> | 7490 |
| <p>Teaching with Commentaries</p> <p><i>Aniruddh Raghu, Maithra Raghu, Simon Kornblith, David Duvenaud, Geoffrey Hinton</i></p> | 7515 |
| <p>In Defense of Pseudo-Labeling: An Uncertainty-Aware Pseudo-Label Selection Framework for Semi-Supervised Learning</p> <p><i>Mamshad Nayeem Rizve, Kevin Duarte, Yogesh S Rawat, Mubarak Shah</i></p> | 7533 |
| <p>Into the Wild with AudioScope: Unsupervised Audio-Visual Separation of On-Screen Sounds</p> <p><i>Efthymios Tzinis, Scott Wisdom, Aren Jansen, Shawn Hershey, Tal Remez, Dan Ellis, John R. Hershey</i></p> | 7553 |
| <p>Cut Out the Annotator, Keep the Cutout: Better Segmentation with Weak Supervision</p> <p><i>Sarah Hooper, Michael Wornow, Ying Hang Seah, Peter Kellman, Hui Xue, Frederic Sala, Curtis Langlotz, Christopher Re</i></p> | 7580 |
| <p>CoDA: Contrast-Enhanced and Diversity-Promoting Data Augmentation for Natural Language Understanding</p> <p><i>Yanru Qu, Dinghan Shen, Yelong Shen, Sandra Sajeev, Weizhu Chen, Jiawei Han</i></p> | 7597 |
| <p>Deep Learning Meets Projective Clustering</p> <p><i>Alaa Maalouf, Harry Lang, Daniela Rus, Dan Feldman</i></p> | 7611 |
| <p>Learning to Deceive Knowledge Graph Augmented Models Via Targeted Perturbation</p> <p><i>Mrigank Raman, Aaron Chan, Siddhant Agarwal, Peifeng Wang, Hansen Wang, Sungchul Kim, Ryan Rossi, Handong Zhao, Nedim Lipka, Xiang Ren</i></p> | 7631 |
| <p>Knowledge Distillation as Semiparametric Inference</p> <p><i>Tri Dao, Govinda M Kamath, Vasilis Syrgkanis, Lester Mackey</i></p> | 7644 |

| | |
|---|------|
| <p>Meta-Learning with Neural Tangent Kernels</p> <p>Yufan Zhou, Zhenyi Wang, Jiayi Xian, Changyou Chen, Jinhui Xu</p> | 7664 |
| <p>Vulnerability-Aware Poisoning Mechanism for Online RL with Unknown Dynamics</p> <p>Yanchao Sun, Da Huo, Furong Huang</p> | 7693 |
| <p>Understanding and Improving Lexical Choice in Non-Autoregressive Translation</p> <p>Liang Ding, Longyue Wang, Xuebo Liu, Derek F. Wong, Dacheng Tao, Zhaopeng Tu</p> | 7720 |
| <p>Layer-Adaptive Sparsity for the Magnitude-Based Pruning</p> <p>Jaeho Lee, Sejun Park, Sangwoo Mo, Sungsoo Ahn, Jinwoo Shin</p> | 7733 |
| <p>Understanding and Improving Encoder Layer Fusion in Sequence-To-Sequence Learning</p> <p>Xuebo Liu, Longyue Wang, Derek F. Wong, Liang Ding, Lidia S. Chao, Zhaopeng Tu</p> | 7752 |
| <p>SaliencyMix: A Saliency Guided Data Augmentation Strategy for Better Regularization</p> <p>A F M Shahab Uddin, Mst. Sirazam Monira, Wheemyung Shin, Taechoong Chung, Sung-Ho Bae</p> | 7766 |
| <p>Are Wider Nets Better Given the Same Number of Parameters?</p> <p>Anna Golubeva, Guy Gur-Ari, Behnam Neyshabur</p> | 7778 |
| <p>Discovering Non-Monotonic Autoregressive Orderings with Variational Inference</p> <p>Xuanlin Li, Brandon Trabucco, Dong Huk Park, Michael Luo, Sheng Shen, Trevor Darrell, Yang Gao</p> | 7797 |
| <p>CompOFA – Compound Once-For-All Networks for Faster Multi-Platform Deployment</p> <p>Manas Sahni, Shreya Varshini, Alind Khare, Alexey Tumanov</p> | 7836 |
| <p>Representing Partial Programs with Blended Abstract Semantics</p> <p>Maxwell Nye, Yewen Pu, Matthew Bowers, Jacob Andreas, Joshua B. Tenenbaum, Armando Solar-Lezama</p> | 7849 |
| <p>PolarNet: Learning to Optimize Polar Keypoints for Keypoint Based Object Detection</p> <p>Wu Xiongwei, Doyen Sahoo, Steven Hoi</p> | 7865 |
| <p>Neural Architecture Search on ImageNet in Four GPU Hours: A Theoretically Inspired Perspective</p> <p>Wuyang Chen, Xinyu Gong, Zhangyang Wang</p> | 7878 |
| <p>Unsupervised Representation Learning for Time Series with Temporal Neighborhood Coding</p> <p>Sana Tonekaboni, Danny Eytan, Anna Goldenberg</p> | 7893 |
| <p>Do Wide and Deep Networks Learn the Same Things? Uncovering How Neural Network Representations Vary with Width and Depth</p> <p>Thao Nguyen, Maithra Raghu, Simon Kornblith</p> | 7910 |
| <p>Supervised Contrastive Learning for Pre-Trained Language Model Fine-Tuning</p> <p>Beliz Gunel, Jingfei Du, Alexis Conneau, Veselin Stoyanov</p> | 7935 |

| | |
|---|------|
| <p>Early Stopping in Deep Networks: Double Descent and How to Eliminate it</p> <p>Reinhard Heckel, Fatih Furkan Yilmaz</p> | 7950 |
| <p>Contrastive Syn-To-Real Generalization</p> <p>Wuyang Chen, Zhiding Yu, Shalini De Mello, Sifei Liu, Jose M. Alvarez, Zhangyang Wang, Anima Anandkumar</p> | 7980 |
| <p>Benchmarks for Deep Off-Policy Evaluation</p> <p>Justin Fu, Mohammad Norouzi, Ofir Nachum, George Tucker, Ziyu Wang, Alexander Novikov, Mengjiao Yang, Michael R Zhang, Yutian Chen, Aviral Kumar, Cosmin Paduraru, Sergey Levine, Thomas Paine</p> | 7992 |
| <p>Pre-Training Text-To-Text Transformers for Concept-Centric Common Sense</p> <p>Wangchunshu Zhou, Dong-Ho Lee, Ravi Kiran Selvam, Seyeon Lee, Xiang Ren</p> | 8019 |
| <p>Combining Label Propagation and Simple Models Out-Performs Graph Neural Networks</p> <p>Qian Huang, Horace He, Abhay Singh, Ser-Nam Lim, Austin Benson</p> | 8034 |
| <p>Learning Long-Term Visual Dynamics with Region Proposal Interaction Networks</p> <p>Haozhi Qi, Xiaolong Wang, Deepak Pathak, Yi Ma, Jitendra Malik</p> | 8055 |
| <p>Chaos of Learning Beyond Zero-Sum and Coordination Via Game Decompositions</p> <p>Yun Kuen Cheung, Yixin Tao</p> | 8071 |
| <p>Control-Aware Representations for Model-Based Reinforcement Learning</p> <p>Brandon Cui, Yinlam Chow, Mohammad Ghavamzadeh</p> | 8093 |
| <p>Provably Robust Classification of Adversarial Examples with Detection</p> <p>Fatemeh Sheikholeslami, Ali Lotfi, J Zico Kolter</p> | 8103 |
| <p>Return-Based Contrastive Representation Learning for Reinforcement Learning</p> <p>Guoqing Liu, Chuheng Zhang, Li Zhao, Tao Qin, Jinhua Zhu, Li Jian, Nenghai Yu, Tie-Yan Liu</p> | 8119 |
| <p>Adversarially-Trained Deep Nets Transfer Better: Illustration on Image Classification</p> <p>Francisco Utrera, Evan Kravitz, N. Benjamin Erichson, Rajiv Khanna, Michael W. Mahoney</p> | 8139 |
| <p>Learning Structural Edits Via Incremental Tree Transformations</p> <p>Ziyu Yao, Frank F. Xu, Pengcheng Yin, Huan Sun, Graham Neubig</p> | 8159 |
| <p>Cross-Attentional Audio-Visual Fusion for Weakly-Supervised Action Localization</p> <p>Jun-Tae Lee, Mihir Jain, Hyungwoo Park, Sungrack Yun</p> | 8179 |
| <p>Improved Estimation of Concentration Under L_p-Norm Distance Metrics Using Half Spaces</p> <p>Jack Prescott, Xiao Zhang, David Evans</p> | 8196 |
| <p>Beyond Categorical Label Representations for Image Classification</p> <p>Boyuan Chen, Yu Li, Sunand Raghupathi, Hod Lipson</p> | 8211 |

| | |
|---|------|
| Fantastic Four: Differentiable and Efficient Bounds on Singular Values of Convolution Layers Sahil Singla, Soheil Feizi | 8226 |
| Accelerating Convergence of Replica Exchange Stochastic Gradient MCMC Via Variance Reduction Wei Deng, Qi Feng, Georgios P. Karagiannis, Guang Lin, Faming Liang | 8245 |
| IsarStep: A Benchmark for High-Level Mathematical Reasoning Wenda Li, Lei Yu, Yuhuai Wu, Lawrence C. Paulson | 8275 |
| Factorizing Declarative and Procedural Knowledge in Structured, Dynamical Environments Anirudh Goyal, Alex Lamb, Phanideep Gampa, Philippe Beaudoin, Charles Blundell, Sergey Levine, Yoshua Bengio, Michael Curtis Mozer | 8290 |
| Provable Rich Observation Reinforcement Learning with Combinatorial Latent States Dipendra Misra, Qinghua Liu, Chi Jin, John Langford | 8314 |
| LowKey: Leveraging Adversarial Attacks to Protect Social Media Users from Facial Recognition Valeriia Cherepanova, Micah Goldblum, Harrison Foley, Shiyuan Duan, John P Dickerson, Gavin Taylor, Tom Goldstein | 8357 |
| Neural Networks for Learning Counterfactual G-Invariances from Single Environments S Chandra Mouli, Bruno Ribeiro | 8372 |
| Simple Spectral Graph Convolution Hao Zhu, Piotr Koniusz | 8402 |
| Anatomy of Catastrophic Forgetting: Hidden Representations and Task Semantics Vinay Venkatesh Ramasesh, Ethan Dyer, Maithra Raghu | 8417 |
| On Fast Adversarial Robustness Adaptation in Model-Agnostic Meta-Learning Ren Wang, Kaidi Xu, Sijia Liu, Pin-Yu Chen, Tsui-Wei Weng, Chuang Gan, Meng Wang | 8448 |
| The Geometry of Integration in Text Classification RNNs Kyle Aitken, Vinay Venkatesh Ramasesh, Ankush Garg, Yuan Cao, David Sussillo, Niru Maheswaranathan | 8464 |
| Towards Robust Neural Networks Via Close-Loop Control Zhuotong Chen, Qianxiao Li, Zheng Zhang | 8495 |
| Projected Latent Markov Chain Monte Carlo: Conditional Sampling of Normalizing Flows Chris Cannella, Mohammadreza Soltani, Vahid Tarokh | 8517 |
| Understanding the Failure Modes of Out-Of-Distribution Generalization Vaishnavh Nagarajan, Anders Andreassen, Behnam Neyshabur | 8544 |
| Usable Information and Evolution of Optimal Representations During Training Michael Kleinman, Alessandro Achille, Daksh Idnani, Jonathan Kao | 8578 |

| | |
|---|------|
| <p>Adaptive Extra-Gradient Methods for Min-Max Optimization and Games</p> <p><i>Kimón Antonakopoulos, Verónica Belmega, Panayotis Mertikopoulos</i></p> | 8597 |
| <p>Shapley Explainability on the Data Manifold</p> <p><i>Christopher Frye, Damien De Mijolla, Tom Begley, Laurence Cowton, Megan Stanley, Ilya Feige</i></p> | 8609 |
| <p>Reinforcement Learning with Random Delays</p> <p><i>Yann Bouteiller, Simon Ramstedt, Giovanni Beltrame, Christopher Pal, Jonathan Binas</i></p> | 8623 |
| <p>Shape Or Texture: Understanding Discriminative Features in CNNs</p> <p><i>Md Amirul Islam, Matthew Kowal, Patrick Esser, Sen Jia, Björn Ommer, Konstantinos G. Derpanis, Neil Bruce</i></p> | 8644 |
| <p>NOVAS: Non-Convex Optimization Via Adaptive Stochastic Search for End-To-End Learning and Control</p> <p><i>Ioannis Exarchos, Marcus Aloysius Pereira, Ziyi Wang, Evangelos Theodorou</i></p> | 8659 |
| <p>Negative Data Augmentation</p> <p><i>Abhishek Sinha, Kumar Ayush, Jiaming Song, Burak Uzcent, Hongxia Jin, Stefano Ermon</i></p> | 8676 |
| <p>Molecule Optimization by Explainable Evolution</p> <p><i>Binghong Chen, Tianzhe Wang, Chengtao Li, Hanjun Dai, Le Song</i></p> | 8694 |
| <p>Estimating Lipschitz Constants of Monotone Deep Equilibrium Models</p> <p><i>Chirag Pabbaraju, Ezra Winston, J Zico Kolter</i></p> | 8709 |
| <p>Implicit Gradient Regularization</p> <p><i>David Barrett, Benoit Dherin</i></p> | 8730 |
| <p>Faster Binary Embeddings for Preserving Euclidean Distances</p> <p><i>Jinjie Zhang, Rayan Saab</i></p> | 8755 |
| <p>Scalable Transfer Learning with Expert Models</p> <p><i>Joan Puigcerver, Carlos Riquelme Ruiz, Basil Mustafa, Cedric Renggli, André Susano Pinto, Sylvain Gelly, Daniel Keysers, Neil Houlsby</i></p> | 8773 |
| <p>Signatory: Differentiable Computations of the Signature and Logsignature Transforms, on Both CPU and GPU</p> <p><i>Patrick Kidger, Terry Lyons</i></p> | 8785 |
| <p>Distance-Based Regularisation of Deep Networks for Fine-Tuning</p> <p><i>Henry Gouk, Timothy Hospedales, Massimiliano Pontil</i></p> | 8812 |
| <p>Decoupling Global and Local Representations Via Invertible Generative Flows</p> <p><i>Xuezhe Ma, Xiang Kong, Shanghang Zhang, Eduard H Hovy</i></p> | 8823 |
| <p>Understanding Over-Parameterization in Generative Adversarial Networks</p> <p><i>Yogesh Balaji, Mohammadmahdi Sajedi, Neha Mukund Kalibhat, Mucong Ding, Dominik Stöger, Mahdi Soltanolkotabi, Soheil Feizi</i></p> | 8846 |

| | |
|---|------|
| Filtered Inner Product Projection for Crosslingual Embedding Alignment Vin Sachidananda, Ziyi Yang, Chenguang Zhu | 8884 |
| Deep Partition Aggregation: Provable Defenses Against General Poisoning Attacks Alexander Levine, Soheil Feizi | 8909 |
| HalentNet: Multimodal Trajectory Forecasting with Hallucinative Intents Deyao Zhu, Mohamed Zahran, Li Erran Li, Mohamed Elhoseiny | 8929 |
| INT: An Inequality Benchmark for Evaluating Generalization in Theorem Proving Yuhuai Wu, Albert Jiang, Jimmy Ba, Roger Baker Grosse | 8950 |
| Bayesian Few-Shot Classification with One-Vs-Each Pólya-Gamma Augmented Gaussian Processes Jake Snell, Richard Zemel | 8985 |
| A Hypergradient Approach to Robust Regression Without Correspondence Yujia Xie, Yixiu Mao, Simiao Zuo, Hongteng Xu, Xiaojing Ye, Tuo Zhao, Hongyuan Zha | 9011 |
| C-Learning: Horizon-Aware Cumulative Accessibility Estimation Panteha Naderian, Gabriel Loaiza-Ganem, Harry J. Braviner, Anthony L. Caterini, Jesse C. Cresswell, Tong Li, Animesh Garg | 9033 |
| MIROSTAT: A NEURAL TEXT DECODING ALGORITHM THAT DIRECTLY CONTROLS PERPLEXITY Sourya Basu, Govardana Sachitanandam Ramachandran, Nitish Shirish Keskar, Lav R. Varshney | 9052 |
| Global Optimality of Softmax Policy Gradient with Single Hidden Layer Neural Networks in the Mean-Field Regime Andrea Agazzi, Jianfeng Lu | 9077 |
| The Deep Bootstrap Framework: Good Online Learners Are Good Offline Generalizers Preetum Nakkiran, Behnam Neyshabur, Hanie Sedghi | 9101 |
| A Better Alternative to Error Feedback for Communication-Efficient Distributed Learning Samuel Horváth, Peter Richtarik | 9129 |
| DrNAS: Dirichlet Neural Architecture Search Xiangning Chen, Ruo Chen Wang, Minhao Cheng, Xiaocheng Tang, Cho-Jui Hsieh | 9147 |
| Offline Model-Based Optimization Via Normalized Maximum Likelihood Estimation Justin Fu, Sergey Levine | 9164 |
| Viewmaker Networks: Learning Views for Unsupervised Representation Learning Alex Tamkin, Mike Wu, Noah Goodman | 9183 |
| Robust Pruning at Initialization Soufiane Hayou, Jean-Francois Ton, Arnaud Doucet, Yee Whye Teh | 9196 |

| | |
|--|------|
| <p>Single-Photon Image Classification</p> <p><i>Thomas Fischbacher, Luciano Sbaiz</i></p> | 9235 |
| <p>Can a Fruit Fly Learn Word Embeddings?</p> <p><i>Yuchen Liang, Chaitanya Ryali, Benjamin Hoover, Leopold Grinberg, Saket Navlakha, Mohammed J Zaki, Dmitry Krotov</i></p> | 9250 |
| <p>A Wigner-Eckart Theorem for Group Equivariant Convolution Kernels</p> <p><i>Leon Lang, Maurice Weiler</i></p> | 9267 |
| <p>Non-Asymptotic Confidence Intervals of Off-Policy Evaluation: Primal and Dual Bounds</p> <p><i>Yihao Feng, Ziyang Tang, Na Zhang, Qiang Liu</i></p> | 9364 |
| <p>Model Patching: Closing the Subgroup Performance Gap with Data Augmentation</p> <p><i>Karan Goel, Albert Gu, Yixuan Li, Christopher Re</i></p> | 9387 |
| <p>SOLAR: Sparse Orthogonal Learned and Random Embeddings</p> <p><i>Tharun Medini, Beidi Chen, Anshumali Shrivastava</i></p> | 9419 |
| <p>Neural Representation and Generation for RNA Secondary Structures</p> <p><i>Zichao Yan, William L. Hamilton, Mathieu Blanchette</i></p> | 9435 |
| <p>A Mathematical Exploration of Why Language Models Help Solve Downstream Tasks</p> <p><i>Nikunj Saunshi, Sadhika Malladi, Sanjeev Arora</i></p> | 9459 |
| <p>Noise Or Signal: The Role of Image Backgrounds in Object Recognition</p> <p><i>Kai Yuanqing Xiao, Logan Engstrom, Andrew Ilyas, Aleksander Madry</i></p> | 9494 |
| <p>Does Enhanced Shape Bias Improve Neural Network Robustness to Common Corruptions?</p> <p><i>Chaithanya Kumar Mummadi, Ranjitha Subramaniam, Robin Huttmacher, Julien Vitay, Volker Fischer, Jan Hendrik Metzen</i></p> | 9522 |
| <p>Long Live the Lottery: The Existence of Winning Tickets in Lifelong Learning</p> <p><i>Tianlong Chen, Zhenyu Zhang, Sijia Liu, Shiyu Chang, Zhangyang Wang</i></p> | 9533 |
| <p>Exploring the Uncertainty Properties of Neural Networks' Implicit Priors in the Infinite-Width Limit</p> <p><i>Ben Adlam, Jaehoon Lee, Lechao Xiao, Jeffrey Pennington, Jasper Snoek</i></p> | 9552 |
| <p>Hierarchical Autoregressive Modeling for Neural Video Compression</p> <p><i>Ruihan Yang, Yibo Yang, Joseph Marino, Stephan Mandt</i></p> | 9579 |
| <p>DeLighT: Deep and Light-Weight Transformer</p> <p><i>Sachin Mehta, Marjan Ghazvininejad, Srinivasan Iyer, Luke Zettlemoyer, Hannaneh Hajishirzi</i></p> | 9594 |
| <p>Learning a Minimax Optimizer: A Pilot Study</p> <p><i>Jiayi Shen, Xiaohan Chen, Howard Heaton, Tianlong Chen, Jialin Liu, Wotao Yin, Zhangyang Wang</i></p> | 9613 |

| | |
|--|------|
| Fuzzy Tiling Activations: A Simple Approach to Learning Sparse Representations Online Yangchen Pan, Kirby Banman, Martha White | 9629 |
| FedBE: Making Bayesian Model Ensemble Applicable to Federated Learning Hong-You Chen, Wei-Lun Chao | 9659 |
| Are Neural Nets Modular? Inspecting Functional Modularity Through Differentiable Weight Masks Róbert Csordás, Sjoerd Van Steenkiste, Jürgen Schmidhuber | 9680 |
| Calibration Tests Beyond Classification David Widmann, Fredrik Lindsten, Dave Zachariah | 9710 |
| Meta Back-Translation Hieu Pham, Xinyi Wang, Yiming Yang, Graham Neubig | 9747 |
| Attentional Constellation Nets for Few-Shot Learning Weijian Xu, Yifan Xu, Huaijin Wang, Zhuowen Tu | 9757 |
| Uncertainty Estimation in Autoregressive Structured Prediction Andrey Malinin, Mark Gales | 9773 |
| Measuring Massive Multitask Language Understanding Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, Jacob Steinhardt | 9804 |
| No MCMC for Me: Amortized Sampling for Fast and Stable Training of Energy-Based Models Will Sussman Grathwohl, Jacob Jin Kelly, Milad Hashemi, Mohammad Norouzi, Kevin Swersky, David Duvenaud | 9831 |
| Aligning AI with Shared Human Values Dan Hendrycks, Collin Burns, Steven Basart, Andrew Critch, Jerry Li, Dawn Song, Jacob Steinhardt | 9866 |
| Class Normalization for (Continual)? Generalized Zero-Shot Learning Ivan Skorokhodov, Mohamed Elhoseiny | 9895 |
| Learning Explanations that Are Hard to Vary Giambattista Parascandolo, Alexander Neitz, Antonio Orvieto, Luigi Gresele, Bernhard Schölkopf | 9926 |
| Degree-Quant: Quantization-Aware Training for Graph Neural Networks Shyam Anil Tailor, Javier Fernandez-Marques, Nicholas Donald Lane | 9950 |
| Byzantine-Resilient Non-Convex Stochastic Gradient Descent Zeyuan Allen-Zhu, Faeze Ebrahimiaghazani, Jerry Li, Dan Alistarh | 9972 |
| Teaching Temporal Logics to Neural Networks Christopher Hahn, Frederik Schmitt, Jens U. Kreber, Markus Norman Rabe, Bernd Finkbeiner | 9996 |

| | |
|---|-------|
| <i>Bypassing the Ambient Dimension: Private SGD with Gradient Subspace Identification</i> Yingxue Zhou, Steven Wu, Arindam Banerjee | 10014 |
| <i>Generative Time-Series Modeling with Fourier Flows</i> Ahmed Alaa, Alex James Chan, Mihaela Van Der Schaar | 10042 |
| <i>Overparameterisation and Worst-Case Generalisation: Friend Or Foe?</i> Aditya Krishna Menon, Ankit Singh Rawat, Sanjiv Kumar | 10053 |
| <i>Improving Zero-Shot Voice Style Transfer Via Disentangled Representation Learning</i> Siyang Yuan, Pengyu Cheng, Ruiyi Zhang, Weituo Hao, Zhe Gan, Lawrence Carin | 10073 |
| <i>Meta-Learning with Negative Learning Rates</i> Alberto Bernacchia | 10085 |
| <i>Learning to Recombine and Resample Data for Compositional Generalization</i> Ekin Akyürek, Afra Feyza Akyürek, Jacob Andreas | 10108 |
| <i>Fooling a Complete Neural Network Verifier</i> Dániel Zombori, Balázs Bánhelyi, Tibor Csendes, István Megyeri, Márk Jelasity | 10130 |
| <i>UMEC: Unified Model and Embedding Compression for Efficient Recommendation Systems</i> Jiayi Shen, Haotao Wang, Shupeng Gui, Jianchao Tan, Zhangyang Wang, Ji Liu | 10146 |
| <i>Evaluations and Methods for Explanation Through Robustness Analysis</i> Cheng-Yu Hsieh, Chih-Kuan Yeh, Xuanqing Liu, Pradeep Kumar Ravikumar, Seungyeon Kim, Sanjiv Kumar, Cho-Jui Hsieh | 10159 |
| <i>Learning and Evaluating Representations for Deep One-Class Classification</i> Kihyuk Sohn, Chun-Liang Li, Jinsung Yoon, Minhó Jin, Tomas Pfister | 10189 |
| <i>Conditional Negative Sampling for Contrastive Learning of Visual Representations</i> Mike Wu, Milan Mosse, Chengxu Zhuang, Daniel Yamins, Noah Goodman | 10221 |
| <i>Communication in Multi-Agent Reinforcement Learning: Intention Sharing</i> Woojun Kim, Jongeui Park, Youngchul Sung | 10233 |
| <i>Neural Thompson Sampling</i> Weitong Zhang, Dongruo Zhou, Lihong Li, Quanquan Gu | 10248 |
| <i>Optimal Regularization Can Mitigate Double Descent</i> Preetum Nakkiran, Prayaag Venkat, Sham M. Kakade, Tengyu Ma | 10274 |
| <i>Separation and Concentration in Deep Networks</i> John Zarka, Florentin Guth, Stéphane Mallat | 10297 |
| <i>Group Equivariant Generative Adversarial Networks</i> Neel Dey, Antong Chen, Soheil Ghafurian | 10308 |

| | |
|--|-------|
| <p>InfoBERT: Improving Robustness of Language Models from an Information Theoretic Perspective</p> <p><i>Boxin Wang, Shuohang Wang, Yu Cheng, Zhe Gan, Ruoxi Jia, Bo Li, Jingjing Liu</i></p> | 10336 |
| <p>Using Latent Space Regression to Analyze and Leverage Compositionality in GANs</p> <p><i>Lucy Chai, Jonas Wulff, Phillip Isola</i></p> | 10359 |
| <p>Go with the Flow: Adaptive Control for Neural ODEs</p> <p><i>Mathieu Chalvidal, Matthew Ricci, Rufin Vanrullen, Thomas Serre</i></p> | 10389 |
| <p>A Learning Theoretic Perspective on Local Explainability</p> <p><i>Jeffrey Li, Vaishnavh Nagarajan, Gregory Plumb, Ameet Talwalkar</i></p> | 10411 |
| <p>CopulaGNN: Towards Integrating Representational and Correlational Roles of Graphs in Graph Neural Networks</p> <p><i>Jiaqi Ma, Bo Chang, Xuefei Zhang, Qiaozhu Mei</i></p> | 10436 |
| <p>Gradient Descent on Neural Networks Typically Occurs at the Edge of Stability</p> <p><i>Jeremy Cohen, Simran Kaur, Yuanzhi Li, J Zico Kolter, Ameet Talwalkar</i></p> | 10449 |
| <p>AutoLRS: Automatic Learning-Rate Schedule by Bayesian Optimization on the Fly</p> <p><i>Yuchen Jin, Tianyi Zhou, Liangyu Zhao, Yibo Zhu, Chuanxiong Guo, Marco Canini, Arvind Krishnamurthy</i></p> | 10529 |
| <p>Autoregressive Dynamics Models for Offline Policy Evaluation and Optimization</p> <p><i>Michael R Zhang, Thomas Paine, Ofir Nachum, Cosmin Paduraru, George Tucker, Ziyu Wang, Mohammad Norouzi</i></p> | 10551 |
| <p>On the Role of Planning in Model-Based Deep Reinforcement Learning</p> <p><i>Jessica B Hamrick, Abram L. Friesen, Feryal Behbahani, Arthur Guez, Fabio Viola, Sims Witherspoon, Thomas Anthony, Lars Holger Buesing, Petar Velickovic, Theophane Weber</i></p> | 10568 |
| <p>Zero-Cost Proxies for Lightweight NAS</p> <p><i>Mohamed S Abdelfattah, Abhinav Mehrotra, Lukasz Dudziak, Nicholas Donald Lane</i></p> | 10605 |
| <p>Personalized Federated Learning with First Order Model Optimization</p> <p><i>Michael Zhang, Karan Sapra, Sanja Fidler, Serena Yeung, Jose M. Alvarez</i></p> | 10622 |
| <p>Deconstructing the Regularization of BatchNorm</p> <p><i>Yann Dauphin, Ekin Dogus Cubuk</i></p> | 10639 |
| <p>No Cost Likelihood Manipulation at Test Time for Making Better Mistakes in Deep Networks</p> <p><i>Shyamgopal Karthik, Ameya Prabhu, Puneet K. Dokania, Vineet Gandhi</i></p> | 10651 |
| <p>Intrinsic-Extrinsic Convolution and Pooling for Learning on 3D Protein Structures</p> <p><i>Pedro Hermosilla, Marco Schäfer, Matej Lang, Gloria Fackelmann, Pere-Pau Vázquez, Barbora Kozlikova, Michael Krone, Tobias Ritschel, Timo Ropinski</i></p> | 10664 |

| | |
|---|-------|
| <p>Generative Language-Grounded Policy in Vision-And-Language Navigation with Bayes' Rule</p> <p>Shuhei Kurita, Kyunghyun Cho</p> | 10680 |
| <p>Learning a Latent Search Space for Routing Problems Using Variational Autoencoders</p> <p>André Hottung, Bhanu Bhandari, Kevin Tierney</p> | 10698 |
| <p>Mastering Atari with Discrete World Models</p> <p>Danijar Hafner, Timothy P Lillicrap, Mohammad Norouzi, Jimmy Ba</p> | 10711 |
| <p>Spatial Dependency Networks: Neural Layers for Improved Generative Image Modeling</p> <p>Đorđe Miladinovic, Aleksandar Stanic, Stefan Bauer, Jürgen Schmidhuber, Joachim M. Buhmann</p> | 10737 |
| <p>Efficient Transformers in Reinforcement Learning Using Actor-Learner Distillation</p> <p>Emilio Parisotto, Russ Salakhutdinov</p> | 10757 |
| <p>IOT: Instance-Wise Layer Reordering for Transformer Structures</p> <p>Jinhua Zhu, Lijun Wu, Yingce Xia, Shufang Xie, Tao Qin, Wengang Zhou, Houqiang Li, Tie-Yan Liu</p> | 10772 |
| <p>Federated Learning Via Posterior Averaging: A New Perspective and Practical Algorithms</p> <p>Maruan Al-Shedivat, Jennifer Gillenwater, Eric Xing, Afshin Rostamizadeh</p> | 10790 |
| <p>Extracting Strong Policies for Robotics Tasks from Zero-Order Trajectory Optimizers</p> <p>Cristina Pinneri, Shambhuraj Sawant, Sebastian Blaes, Georg Martius</p> | 10813 |
| <p>BREEDS: Benchmarks for Subpopulation Shift</p> <p>Shibani Santurkar, Dimitris Tsipras, Aleksander Madry</p> | 10827 |
| <p>On the Impossibility of Global Convergence in Multi-Loss Optimization</p> <p>Alistair Letcher</p> | 10860 |
| <p>Efficient Wasserstein Natural Gradients for Reinforcement Learning</p> <p>Ted Moskovitz, Michael Arbel, Ferenc Huszar, Arthur Gretton</p> | 10886 |
| <p>Multi-Level Local SGD: Distributed SGD for Heterogeneous Hierarchical Networks</p> <p>Timothy Castiglia, Anirban Das, Stacy Patterson</p> | 10901 |
| <p>Kanerva++: Extending the Kanerva Machine with Differentiable, Locally Block Allocated Latent Memory</p> <p>Jason Ramapuram, Yan Wu, Alexandros Kalousis</p> | 10937 |
| <p>EVALUATION OF NEURAL ARCHITECTURES TRAINED WITH SQUARE LOSS VS CROSS-ENTROPY IN CLASSIFICATION TASKS</p> <p>Like Hui, Mikhail Belkin</p> | 10956 |
| <p>Self-Supervised Learning from a Multi-View Perspective</p> <p>Yao-Hung Hubert Tsai, Yue Wu, Ruslan Salakhutdinov, Louis-Philippe Morency</p> | 10973 |

| | |
|--|-------|
| <p>Interpretable Neural Architecture Search Via Bayesian Optimisation with Weisfeiler-Lehman Kernels</p> <p><i>Binxin Ru, Xingchen Wan, Xiaowen Dong, Michael Osborne</i></p> | 10991 |
| <p>CT-Net: Channel Tensorization Network for Video Classification</p> <p><i>Kunchang Li, Xianhang Li, Yali Wang, Jun Wang, Yu Qiao</i></p> | 11014 |
| <p>Intraclass Clustering: An Implicit Learning Ability that Regularizes DNNs</p> <p><i>Simon Carbonnelle, Christophe De Vleeschouwer</i></p> | 11027 |
| <p>Learning Robust State Abstractions for Hidden-Parameter Block MDPs</p> <p><i>Amy Zhang, Shagun Sodhani, Khimya Khetarpal, Joelle Pineau</i></p> | 11041 |
| <p>Isometric Transformation Invariant and Equivariant Graph Convolutional Networks</p> <p><i>Masanobu Horie, Naoki Morita, Toshiaki Hishinuma, Yu Ihara, Naoto Mitsume</i></p> | 11063 |
| <p>Learning Safe Multi-Agent Control with Decentralized Neural Barrier Certificates</p> <p><i>Zengyi Qin, Kaiqing Zhang, Yuxiao Chen, Jingkai Chen, Chuchu Fan</i></p> | 11087 |
| <p>ChipNet: Budget-Aware Pruning with Heaviside Continuous Approximations</p> <p><i>Rishabh Tiwari, Udbhav Bamba, Arnav Chavan, Deepak Gupta</i></p> | 11103 |
| <p>AdaSpeech: Adaptive Text to Speech for Custom Voice</p> <p><i>Mingjian Chen, Xu Tan, Bohan Li, Yanqing Liu, Tao Qin, Sheng Zhao, Tie-Yan Liu</i></p> | 11121 |
| <p>Estimating and Evaluating Regression Predictive Uncertainty in Deep Object Detectors</p> <p><i>Ali Harakeh, Steven L. Waslander</i></p> | 11131 |
| <p>Domain-Robust Visual Imitation Learning with Mutual Information Constraints</p> <p><i>Edoardo Cedin, Oya Celiktutan</i></p> | 11154 |
| <p>Clustering-Friendly Representation Learning Via Instance Discrimination and Feature Decorrelation</p> <p><i>Yaling Tao, Kentaro Takagi, Kouta Nakata</i></p> | 11177 |
| <p>Progressive Skeletonization: Trimming More Fat from a Network at Initialization</p> <p><i>Pau De Jorge, Amartya Sanyal, Harkirat Behl, Philip Torr, Grégory Rogez, Puneet K. Dokania</i></p> | 11192 |
| <p>On the Curse of Memory in Recurrent Neural Networks: Approximation and Optimization Analysis</p> <p><i>Zhong Li, Jiequn Han, Weinan E, Qianxiao Li</i></p> | 11213 |
| <p>Do Not Let Privacy Overbill Utility: Gradient Embedding Perturbation for Private Learning</p> <p><i>Da Yu, Huishuai Zhang, Wei Chen, Tie-Yan Liu</i></p> | 11256 |
| <p>More Or Less: When and How to Build Convolutional Neural Network Ensembles</p> <p><i>Abdul Wasay, Stratos Idreos</i></p> | 11274 |

| | |
|--|-------|
| <p>Revisiting Hierarchical Approach for Persistent Long-Term Video Prediction</p> <p>Wonkwang Lee, Whie Jung, Han Zhang, Ting Chen, Jing Yu Koh, Thomas Huang, Hyungsuk Yoon, Honglak Lee, Seunghoon Hong</p> | 11291 |
| <p>Symmetry-Aware Actor-Critic for 3D Molecular Design</p> <p>Gregor N. C. Simm, Robert Pinsler, Gábor Csányi, José Miguel Hernández-Lobato</p> | 11318 |
| <p>Learnable Embedding Sizes for Recommender Systems</p> <p>Siyi Liu, Chen Gao, Yihong Chen, Depeng Jin, Yong Li</p> | 11336 |
| <p>Learning Neural Generative Dynamics for Molecular Conformation Generation</p> <p>Minkai Xu, Shitong Luo, Yoshua Bengio, Jian Peng, Jian Tang</p> | 11353 |
| <p>Learning the Pareto Front with Hypernetworks</p> <p>Aviv Navon, Aviv Shamsian, Ethan Fetaya, Gal Chechik</p> | 11370 |
| <p>Ask Your Humans: Using Human Instructions to Improve Generalization in Reinforcement Learning</p> <p>Valerie Chen, Abhinav Gupta, Kenneth Marino</p> | 11388 |
| <p>Taming GANs with Lookahead-Minmax</p> <p>Tatjana Chavdarova, Matteo Pagliardini, Sebastian U Stich, François Fleuret, Martin Jaggi</p> | 11410 |
| <p>Uncertainty in Gradient Boosting Via Ensembles</p> <p>Andrey Malinin, Liudmila Prokhorenkova, Aleksei Ustimenko</p> | 11448 |
| <p>Adversarial Score Matching and Improved Sampling for Image Generation</p> <p>Alexia Jolicoeur-Martineau, Rémi Piché-Taillefer, Ioannis Mitliagkas, Remi Tachet Des Combes</p> | 11465 |
| <p>FastSpeech 2: Fast and High-Quality End-To-End Text to Speech</p> <p>Yi Ren, Chenxu Hu, Xu Tan, Tao Qin, Sheng Zhao, Zhou Zhao, Tie-Yan Liu</p> | 11474 |
| <p>Interpretable Models for Granger Causality Using Self-Explaining Neural Networks</p> <p>Ricards Marcinkevics, Julia E Vogt</p> | 11489 |
| <p>Learning Deep Features in Instrumental Variable Regression</p> <p>Liyuan Xu, Yutian Chen, Siddarth Srinivasan, Nando De Freitas, Arnaud Doucet, Arthur Gretton</p> | 11512 |
| <p>Statistical Inference for Individual Fairness</p> <p>Subha Maity, Songkai Xue, Mikhail Yurochkin, Yuekai Sun</p> | 11540 |
| <p>Randomized Ensembled Double Q-Learning: Learning Fast Without a Model</p> <p>Xinyue Chen, Che Wang, Zijian Zhou, Keith W. Ross</p> | 11559 |
| <p>Lifelong Learning of Compositional Structures</p> <p>Jorge A Mendez, Eric Eaton</p> | 11584 |
| <p>QPLEX: Duplex Dueling Multi-Agent Q-Learning</p> <p>Jianhao Wang, Zhizhou Ren, Terry Liu, Yang Yu, Chongjie Zhang</p> | 11609 |

| | |
|--|-------|
| <p>Meta-Learning of Structured Task Distributions in Humans and Machines</p> <p><i>Sreejan Kumar, Ishita Dasgupta, Jonathan Cohen, Nathaniel Daw, Thomas Griffiths</i></p> | 11636 |
| <p>PC2WF: 3D Wireframe Reconstruction from Raw Point Clouds</p> <p><i>Yujia Liu, Stefano D'Aronco, Konrad Schindler, Jan Dirk Wegner</i></p> | 11649 |
| <p>Neural Attention Distillation: Erasing Backdoor Triggers from Deep Neural Networks</p> <p><i>Yige Li, Xixiang Lyu, Nodens Koren, Lingjuan Lyu, Bo Li, Xingjun Ma</i></p> | 11661 |
| <p>HyperGrid Transformers: Towards a Single Model for Multiple Tasks</p> <p><i>Yi Tay, Zhe Zhao, Dara Bahri, Donald Metzler, Da-Cheng Juan</i></p> | 11680 |
| <p>Long Range Arena : A Benchmark for Efficient Transformers</p> <p><i>Yi Tay, Mostafa Dehghani, Samira Abnar, Yikang Shen, Dara Bahri, Philip Pham, Jinfeng Rao, Liu Yang, Sebastian Ruder, Donald Metzler</i></p> | 11694 |
| <p>Acting in Delayed Environments with Non-Stationary Markov Policies</p> <p><i>Esther Derman, Gal Dalal, Shie Mannor</i></p> | 11713 |
| <p>Neurally Augmented ALISTA</p> <p><i>Freya Behrens, Jonathan Sauder, Peter Jung</i></p> | 11724 |
| <p>Relating by Contrasting: A Data-Efficient Framework for Multimodal Generative Models</p> <p><i>Yuge Shi, Brooks Paige, Philip Torr, Siddharth N</i></p> | 11737 |
| <p>Categorical Normalizing Flows Via Continuous Transformations</p> <p><i>Phillip Lippe, Efstratios Gavves</i></p> | 11762 |
| <p>Prototypical Representation Learning for Relation Extraction</p> <p><i>Ning Ding, Xiaobin Wang, Yao Fu, Guangwei Xu, Rui Wang, Pengjun Xie, Ying Shen, Fei Huang, Hai-Tao Zheng, Rui Zhang</i></p> | 11789 |
| <p>Generalized Energy Based Models</p> <p><i>Michael Arbel, Liang Zhou, Arthur Gretton</i></p> | 11805 |
| <p>Predicting Classification Accuracy When Adding New Unobserved Classes</p> <p><i>Yuli Slavutsky, Yuval Benjamini</i></p> | 11841 |
| <p>In Search of Lost Domain Generalization</p> <p><i>Ishaan Gulrajani, David Lopez-Paz</i></p> | 11856 |
| <p>Estimating Informativeness of Samples with Smooth Unique Information</p> <p><i>Hrayr Harutyunyan, Alessandro Achille, Giovanni Paolini, Orchid Majumder, Avinash Ravichandran, Rahul Bhotika, Stefano Soatto</i></p> | 11885 |
| <p>Identifying Physical Law of Hamiltonian Systems Via Meta-Learning</p> <p><i>Seungjun Lee, Haesang Yang, Woojae Seong</i></p> | 11907 |

| | |
|--|-------|
| Adapting to Reward Progressivity Via Spectral Reinforcement Learning <i>Michael Dann, John Thangarajah</i> | 11921 |
| Understanding the Effects of Data Parallelism and Sparsity on Neural Network Training <i>Namhoon Lee, Thalayasingam Ajanthan, Philip Torr, Martin Jaggi</i> | 11937 |
| Primal Wasserstein Imitation Learning <i>Robert Dadashi, Leonard Hussenot, Matthieu Geist, Olivier Pietquin</i> | 11958 |
| Prediction and Generalisation Over Directed Actions by Grid Cells <i>Changmin Yu, Timothy Behrens, Neil Burgess</i> | 11977 |
| Drop-Bottleneck: Learning Discrete Compressed Representation for Noise-Robust Exploration <i>Jaekyeom Kim, Minjung Kim, Dongyeon Woo, Gunhee Kim</i> | 11998 |
| BOIL: Towards Representation Change for Few-Shot Learning <i>Jaehoon Oh, Hyungjun Yoo, Changhwan Kim, Se-Young Yun</i> | 12016 |
| MultiModalQA: Complex Question Answering Over Text, Tables and Images <i>Alon Talmor, Ori Yoran, Amnon Catav, Dan Lahav, Yizhong Wang, Akari Asai, Gabriel Ilharco, Hannaneh Hajishirzi, Jonathan Berant</i> | 12040 |
| Neural Gradients Are Near-Lognormal: Improved Quantized and Sparse Training <i>Brian Chmiel, Liad Ben-Uri, Moran Shkolnik, Elad Hoffer, Ron Banner, Daniel Soudry</i> | 12052 |
| Group Equivariant Conditional Neural Processes <i>Makoto Kawano, Wataru Kumagai, Akiyoshi Sannai, Yusuke Iwasawa, Yutaka Matsuo</i> | 12063 |
| Deployment-Efficient Reinforcement Learning Via Model-Based Offline Optimization <i>Tatsuya Matsushima, Hiroki Furuta, Yutaka Matsuo, Ofir Nachum, Shixiang Gu</i> | 12081 |
| Efficient Conformal Prediction Via Cascaded Inference with Expanded Admission <i>Adam Fisch, Tal Schuster, Tommi S. Jaakkola, Regina Barzilay</i> | 12108 |
| Reducing the Computational Cost of Deep Generative Models with Binary Neural Networks <i>Thomas Bird, Friso Kingma, David Barber</i> | 12129 |
| Large-Width Functional Asymptotics for Deep Gaussian Neural Networks <i>Daniele Bracale, Stefano Favaro, Sandra Fortini, Stefano Peluchetti</i> | 12138 |
| MetaNorm: Learning to Normalize Few-Shot Batches Across Domains <i>Yingjun Du, Xiantong Zhen, Ling Shao, Cees G. M. Snoek</i> | 12148 |
| Representation Balancing Offline Model-Based Reinforcement Learning <i>Byung-Jun Lee, Jongmin Lee, Kee-Eung Kim</i> | 12171 |
| Local Convergence Analysis of Gradient Descent Ascent with Finite Timescale Separation <i>Tanner Fiez, Lillian J Ratliff</i> | 12193 |

| | |
|--|-------|
| <p><i>The Importance of Pessimism in Fixed-Dataset Policy Optimization</i></p> <p><i>Jacob Buckman, Carles Gelada, Marc G Bellemare</i></p> | 12254 |
| <p><i>Uncertainty Estimation and Calibration with Finite-State Probabilistic RNNs</i></p> <p><i>Cheng Wang, Carolin Lawrence, Mathias Niepert</i></p> | 12285 |
| <p><i>Empirical Or Invariant Risk Minimization? a Sample Complexity Perspective</i></p> <p><i>Kartik Ahuja, Jun Wang, Amit Dhurandhar, Karthikeyan Shanmugam, Kush R. Varshney</i></p> | 12308 |
| <p><i>Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval</i></p> <p><i>Lee Xiong, Chenyan Xiong, Ye Li, Kwok-Fung Tang, Jialin Liu, Paul N. Bennett, Junaid Ahmed, Arnold Overwijk</i></p> | 12357 |
| <p><i>FairBatch: Batch Selection for Model Fairness</i></p> <p><i>Yuji Roh, Kangwook Lee, Steven Euijong Whang, Changho Suh</i></p> | 12373 |
| <p><i>An Unsupervised Deep Learning Approach for Real-World Image Denoising</i></p> <p><i>Dihan Zheng, Sia Huat Tan, Xiaowen Zhang, Zuoqiang Shi, Kaisheng Ma, Chenglong Bao</i></p> | 12394 |
| <p><i>When Optimizing f-Divergence is Robust with Label Noise</i></p> <p><i>Jiaheng Wei, Yang Liu</i></p> | 12417 |
| <p><i>Meta-Learning Symmetries by Reparameterization</i></p> <p><i>Allan Zhou, Tom Knowles, Chelsea Finn</i></p> | 12428 |
| <p><i>A Geometric Analysis of Deep Generative Image Models and Its Applications</i></p> <p><i>Binxu Wang, Carlos R Ponce</i></p> | 12447 |
| <p><i>What Makes Instance Discrimination Good for Transfer Learning?</i></p> <p><i>Nanxuan Zhao, Zhirong Wu, Rynson W. H. Lau, Stephen Lin</i></p> | 12472 |
| <p><i>Answering Complex Open-Domain Questions with Multi-Hop Dense Retrieval</i></p> <p><i>Wenhan Xiong, Xiang Li, Srini Iyer, Jingfei Du, Patrick Lewis, William Yang Wang, Yashar Mehdad, Scott Yih, Sebastian Riedel, Douwe Kiela, Barlas Oguz</i></p> | 12489 |
| <p><i>On Dyadic Fairness: Exploring and Mitigating Bias in Graph Connections</i></p> <p><i>Peizhao Li, Yifei Wang, Han Zhao, Pengyu Hong, Hongfu Liu</i></p> | 12508 |
| <p><i>Spatio-Temporal Graph Scattering Transform</i></p> <p><i>Chao Pan, Siheng Chen, Antonio Ortega</i></p> | 12526 |
| <p><i>MODALS: Modality-Agnostic Automated Data Augmentation in the Latent Space</i></p> <p><i>Tsz-Him Cheung, Dit-Yan Yeung</i></p> | 12544 |
| <p><i>ALFWorld: Aligning Text and Embodied Environments for Interactive Learning</i></p> <p><i>Mohit Shridhar, Xingdi Yuan, Marc-Alexandre Cote, Yonatan Bisk, Adam Trischler, Matthew Hausknecht</i></p> | 12562 |

| | |
|--|-------|
| <p>Latent Skill Planning for Exploration and Transfer</p> <p>Kevin Xie, Homanga Bharadhwaj, Danijar Hafner, Animesh Garg, Florian Shkurti</p> | 12583 |
| <p>Evaluating the Disentanglement of Deep Generative Models Through Manifold Topology</p> <p>Sharon Zhou, Eric Zelikman, Fred Lu, Andrew Y. Ng, Gunnar E. Carlsson, Stefano Ermon</p> | 12599 |
| <p>Combining Ensembles and Data Augmentation Can Harm Your Calibration</p> <p>Yeming Wen, Ghassen Jerfel, Rafael Muller, Michael W Dusenberry, Jasper Snoek, Balaji Lakshminarayanan, Dustin Tran</p> | 12611 |
| <p>CoCo: Controllable Counterfactuals for Evaluating Dialogue State Trackers</p> <p>Shiyang Li, Semih Yavuz, Kazuma Hashimoto, Jia Li, Tong Niu, Nazneen Rajani, Xifeng Yan, Yingbo Zhou, Caiming Xiong</p> | 12632 |
| <p>AdaGCN: Adaboosting Graph Convolutional Networks into Deep Models</p> <p>Ke Sun, Zhanxing Zhu, Zhouchen Lin</p> | 12657 |
| <p>Learning What to Do by Simulating the Past</p> <p>David Lindner, Rohin Shah, Pieter Abbeel, Anca Dragan</p> | 12672 |
| <p>CcGAN: Continuous Conditional Generative Adversarial Networks for Image Generation</p> <p>Xin Ding, Yongwei Wang, Zuheng Xu, William J Welch, Z. Jane Wang</p> | 12696 |
| <p>ANOCE: Analysis of Causal Effects with Multiple Mediators Via Constrained Structural Learning</p> <p>Hengrui Cai, Rui Song, Wenbin Lu</p> | 12726 |
| <p>Direction Matters: On the Implicit Bias of Stochastic Gradient Descent with Moderate Learning Rate</p> <p>Jingfeng Wu, Difan Zou, Vladimir Braverman, Quanquan Gu</p> | 12759 |
| <p>Robust Early-Learning: Hindering the Memorization of Noisy Labels</p> <p>Xiaobo Xia, Tongliang Liu, Bo Han, Chen Gong, Nannan Wang, Zongyuan Ge, Yi Chang</p> | 12802 |
| <p>A Design Space Study for LISTA and Beyond</p> <p>Tianjian Meng, Xiaohan Chen, Yifan Jiang, Zhangyang Wang</p> | 12817 |
| <p>PAC Confidence Predictions for Deep Neural Network Classifiers</p> <p>Sangdon Park, Shuo Li, Insup Lee, Osbert Bastani</p> | 12834 |
| <p>X2T: Training an X-To-Text Typing Interface with Online Learning from User Feedback</p> <p>Jensen Gao, Siddharth Reddy, Glen Berseth, Nicholas Hardy, Nikhilesh Natraj, Karunesh Ganguly, Anca Dragan, Sergey Levine</p> | 12861 |
| <p>Discrete Graph Structure Learning for Forecasting Multiple Time Series</p> <p>Chao Shang, Jie Chen, Jinbo Bi</p> | 12878 |
| <p>Task-Agnostic Morphology Evolution</p> <p>Donald Joseph Hejna III, Pieter Abbeel, Lerrel Pinto</p> | 12892 |

| | |
|---|-------|
| Deep Equals Shallow for ReLU Networks in Kernel Regimes <i>Alberto Bietti, Francis Bach</i> | 12913 |
| Loss Function Discovery for Object Detection Via Convergence-Simulation Driven Search <i>Peidong Liu, Gengwei Zhang, Bochao Wang, Hang Xu, Xiaodan Liang, Yong Jiang, Zhenguo Li</i> | 12935 |
| Effective Distributed Learning with Random Features: Improved Bounds and Algorithms <i>Yong Liu, Jiankun Liu, Shuqiang Wang</i> | 12946 |
| On InstaHide, Phase Retrieval, and Sparse Matrix Factorization <i>Sitan Chen, Xiaoxiao Li, Zhao Song, Danyang Zhuo</i> | 12974 |
| Anchor & Transform: Learning Sparse Embeddings for Large Vocabularies <i>Paul Pu Liang, Manzil Zaheer, Yuan Wang, Amr Ahmed</i> | 13000 |
| Achieving Linear Speedup with Partial Worker Participation in Non-IID Federated Learning <i>Haibo Yang, Minghong Fang, Jia Liu</i> | 13031 |
| Deep Neural Tangent Kernel and Laplace Kernel Have the Same RKHS <i>Lin Chen, Sheng Xu</i> | 13054 |
| Fast Convergence of Stochastic Subgradient Method Under Interpolation <i>Huang Fang, Zhenan Fan, Michael Friedlander</i> | 13076 |
| Generating Adversarial Computer Programs Using Optimized Obfuscations <i>Shashank Srikant, Sijia Liu, Tamara Mitrovska, Shiyu Chang, Quanfu Fan, Gaoyuan Zhang, Una-May O'Reilly</i> | 13089 |
| C-Learning: Learning to Achieve Goals Via Recursive Classification <i>Benjamin Eysenbach, Ruslan Salakhutdinov, Sergey Levine</i> | 13105 |
| Off-Dynamics Reinforcement Learning: Training for Transfer with Domain Classifiers <i>Benjamin Eysenbach, Shreyas Chaudhari, Swapnil Asawa, Sergey Levine, Ruslan Salakhutdinov</i> | 13130 |
| Multi-Resolution Modeling of a Discrete Stochastic Process Identifies Causes of Cancer <i>Adam Uri Yaari, Maxwell Sherman, Oliver Clarke Priebe, Po-Ru Loh, Boris Katz, Andrei Barbu, Bonnie Berger</i> | 13150 |
| On the Critical Role of Conventions in Adaptive Human-AI Collaboration <i>Andy Shih, Arjun Sawhney, Jovana Kondic, Stefano Ermon, Dorsa Sadigh</i> | 13174 |
| WaveGrad: Estimating Gradients for Waveform Generation <i>Nanxin Chen, Yu Zhang, Heiga Zen, Ron J Weiss, Mohammad Norouzi, William Chan</i> | 13188 |
| A Critique of Self-Expressive Deep Subspace Clustering <i>Benjamin David Haeffele, Chong You, Rene Vidal</i> | 13203 |
| Explaining by Imitating: Understanding Decisions by Interpretable Policy Learning <i>Alihan Hüyük, Daniel Jarrett, Cem Tekin, Mihaela Van Der Schaar</i> | 13223 |

| | |
|--|-------|
| For Self-Supervised Learning, Rationality Implies Generalization, Provably Yamini Bansal, Gal Kaplun, Boaz Barak | 13247 |
| Directed Acyclic Graph Neural Networks Veronika Thost, Jie Chen | 13272 |
| My Body is a Cage: The Role of Morphology in Graph-Based Incompatible Control Vitaly Kurin, Maximilian Igl, Tim Rocktäschel, Wendelin Boehmer, Shimon Whiteson | 13288 |
| Incremental Few-Shot Learning Via Vector Quantization in Deep Embedded Space Kuilin Chen, Chi-Guhn Lee | 13308 |
| Revisiting Few-Sample BERT Fine-Tuning Tianyi Zhang, Felix Wu, Arzoo Katiyar, Kilian Q Weinberger, Yoav Artzi | 13324 |
| Linear Convergent Decentralized Optimization with Compression Xiaorui Liu, Yao Li, Rongrong Wang, Jiliang Tang, Ming Yan | 13346 |
| Learning from Demonstration with Weakly Supervised Disentanglement Yordan Hristov, Subramanian Ramamoorthy | 13376 |
| Incorporating Symmetry into Deep Dynamics Models for Improved Generalization Rui Wang, Robin Walters, Rose Yu | 13394 |
| The Risks of Invariant Risk Minimization Elan Rosenfeld, Pradeep Kumar Ravikumar, Andrej Risteski | 13414 |
| Scaling Symbolic Methods Using Gradients for Neural Model Explanation Subham Sekhar Sahoo, Subhashini Venugopalan, Li Li, Rishabh Singh, Patrick Riley | 13441 |
| Contextual Dropout: An Efficient Sample-Dependent Dropout Module Xinjie Fan, Shujian Zhang, Korawat Tanwisuth, Xiaoning Qian, Mingyuan Zhou | 13459 |
| Contrastive Learning with Hard Negative Samples Joshua David Robinson, Ching-Yao Chuang, Suvrit Sra, Stefanie Jegelka | 13486 |
| Debiasing Concept-Based Explanations with Causal Analysis Mohammad Taha Bahadori, David Heckerman | 13515 |
| Risk-Averse Offline Reinforcement Learning Núria Armengol Urpí, Sebastian Curi, Andreas Krause | 13526 |
| Parameter Efficient Multimodal Transformers for Video Representation Learning Sangho Lee, Youngjae Yu, Gunhee Kim, Thomas Breuel, Jan Kautz, Yale Song | 13543 |
| Tradeoffs in Data Augmentation: An Empirical Study Raphael Gontijo-Lopes, Sylvia Smullin, Ekin Dogus Cubuk, Ethan Dyer | 13560 |

| | |
|---|-------|
| <p>Concept Learners for Few-Shot Learning</p> <p><i>Kaidi Cao, Maria Brbic, Jure Leskovec</i></p> | 13587 |
| <p>Neural Mechanics: Symmetry and Broken Conservation Laws in Deep Learning Dynamics</p> <p><i>Daniel Kunin, Javier Sagastuy-Brena, Surya Ganguli, Daniel Lk Yamins, Hidenori Tanaka</i></p> | 13604 |
| <p>Fair Mixup: Fairness Via Interpolation</p> <p><i>Ching-Yao Chuang, Youssef Mroueh</i></p> | 13634 |
| <p>Heating Up Decision Boundaries: Isocapacitory Saturation, Adversarial Scenarios and Generalization Bounds</p> <p><i>Bogdan Georgiev, Lukas Franken, Mayukh Mukherjee</i></p> | 13649 |
| <p>Influence Functions in Deep Learning Are Fragile</p> <p><i>Samyadeep Basu, Phil Pope, Soheil Feizi</i></p> | 13685 |
| <p>Few-Shot Learning Via Learning the Representation, Provably</p> <p><i>Simon Shaolei Du, Wei Hu, Sham M. Kakade, Jason D. Lee, Qi Lei</i></p> | 13707 |
| <p>Self-Supervised Learning of Compressed Video Representations</p> <p><i>Youngjae Yu, Sangho Lee, Gunhee Kim, Yale Song</i></p> | 13736 |
| <p>Flowtron: An Autoregressive Flow-Based Generative Network for Text-To-Speech Synthesis</p> <p><i>Rafael Valle, Kevin J. Shih, Ryan Prenger, Bryan Catanzaro</i></p> | 13759 |
| <p>R-GAP: Recursive Gradient Attack on Privacy</p> <p><i>Junyi Zhu, Matthew B. Blaschko</i></p> | 13776 |
| <p>Why Resampling Outperforms Reweighting for Correcting Sampling Bias with Stochastic Gradients</p> <p><i>Jing An, Lexing Ying, Yuhua Zhu</i></p> | 13793 |
| <p>Learning with Instance-Dependent Label Noise: A Sample Sieve Approach</p> <p><i>Hao Cheng, Zhaowei Zhu, Xingyu Li, Yifei Gong, Xing Sun, Yang Liu</i></p> | 13812 |
| <p>Unsupervised Audiovisual Synthesis Via Exemplar Autoencoders</p> <p><i>Kangle Deng, Aayush Bansal, Deva Ramanan</i></p> | 13839 |
| <p>Single-Timescale Actor-Critic Provably Finds Globally Optimal Policy</p> <p><i>Zuyue Fu, Zhuoran Yang, Zhaoran Wang</i></p> | 13860 |
| <p>Wandering Within a World: Online Contextualized Few-Shot Learning</p> <p><i>Mengye Ren, Michael Louis Iuzzolino, Michael Curtis Mozer, Richard Zemel</i></p> | 13902 |
| <p>Neural Delay Differential Equations</p> <p><i>Qunxi Zhu, Yao Guo, Wei Lin</i></p> | 13923 |
| <p>Dance Revolution: Long-Term Dance Generation with Music Via Curriculum Learning</p> <p><i>Ruozi Huang, Huang Hu, Wei Wu, Kei Sawada, Mi Zhang, Daxin Jiang</i></p> | 13943 |

| | |
|---|-------|
| <p>Learning Parametrised Graph Shift Operators</p> <p><i>George Dasoulas, Johannes F. Lutzeyer, Michalis Vazirgiannis</i></p> | 13957 |
| <p>DOP: Off-Policy Multi-Agent Decomposed Policy Gradients</p> <p><i>Yihan Wang, Beining Han, Tonghan Wang, Heng Dong, Chongjie Zhang</i></p> | 13974 |
| <p>Unsupervised Discovery of 3D Physical Objects from Video</p> <p><i>Yilun Du, Kevin A. Smith, Tomer Ullman, Joshua B. Tenenbaum, Jiajun Wu</i></p> | 13998 |
| <p>Exploring Balanced Feature Spaces for Representation Learning</p> <p><i>Bingyi Kang, Yu Li, Sa Xie, Zehuan Yuan, Jiashi Feng</i></p> | 14013 |
| <p>Auxiliary Learning by Implicit Differentiation</p> <p><i>Aviv Navon, Idan Achituve, Haggai Maron, Gal Chechik, Ethan Fetaya</i></p> | 14028 |
| <p>On the Bottleneck of Graph Neural Networks and Its Practical Implications</p> <p><i>Uri Alon, Eran Yahav</i></p> | 14048 |
| <p>The Role of Momentum Parameters in the Optimal Convergence of Adaptive Polyak's Heavy-Ball Methods</p> <p><i>Wei Tao, Sheng Long, Gaowei Wu, Qing Tao</i></p> | 14064 |
| <p>Entropic Gradient Descent Algorithms and Wide Flat Minima</p> <p><i>Fabrizio Pittorino, Carlo Lucibello, Christoph Feinauer, Gabriele Perugini, Carlo Baldassi, Elizaveta Demyanenko, Riccardo Zecchina</i></p> | 14082 |
| <p>SEDONA: Search for Decoupled Neural Networks Toward Greedy Block-Wise Learning</p> <p><i>Myeongjang Pyeon, Jihwan Moon, Taeyoung Hahn, Gunhee Kim</i></p> | 14099 |
| <p>Not-MIWAE: Deep Generative Modelling with Missing Not at Random Data</p> <p><i>Niels Bruun Ipsen, Pierre-Alexandre Mattei, Jes Frellesen</i></p> | 14116 |
| <p>Reweighting Augmented Samples by Minimizing the Maximal Expected Loss</p> <p><i>Mingyang Yi, Lu Hou, Lifeng Shang, Xin Jiang, Qun Liu, Zhi-Ming Ma</i></p> | 14134 |
| <p>Learning to Represent Action Values as a Hypergraph on the Action Vertices</p> <p><i>Arash Tavakoli, Mehdi Fatemi, Petar Kormushev</i></p> | 14148 |
| <p>Shapley Explanation Networks</p> <p><i>Rui Wang, Xiaoqian Wang, David I. Inouye</i></p> | 14167 |
| <p>NBDT: Neural-Backed Decision Tree</p> <p><i>Alvin Wan, Lisa Dunlap, Daniel Ho, Jihan Yin, Scott Lee, Suzanne Petryk, Sarah Adel Bargal, Joseph E. Gonzalez</i></p> | 14193 |
| <p>Denoising Diffusion Implicit Models</p> <p><i>Jiaming Song, Chenlin Meng, Stefano Ermon</i></p> | 14205 |

| | |
|--|-------|
| <p>GShard: Scaling Giant Models with Conditional Computation and Automatic Sharding</p> <p><i>Dmitry Lepikhin, Hyoukjoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang, Maxim Krikun, Noam Shazeer, Zhifeng Chen</i></p> | 14225 |
| <p>What Should Not Be Contrastive in Contrastive Learning</p> <p><i>Tete Xiao, Xiaolong Wang, Alexei A Efros, Trevor Darrell</i></p> | 14248 |
| <p>On Data-Augmentation and Consistency-Based Semi-Supervised Learning</p> <p><i>Atin Ghosh, Alexandre H. Thiery</i></p> | 14261 |
| <p>Diverse Video Generation Using a Gaussian Process Trigger</p> <p><i>Gaurav Shrivastava, Abhinav Shrivastava</i></p> | 14274 |
| <p>Monotonic Kronecker-Factored Lattice</p> <p><i>William Taylor Bakst, Nobuyuki Morioka, Erez Louidor</i></p> | 14291 |
| <p>Auto Seg-Loss: Searching Metric Surrogates for Semantic Segmentation</p> <p><i>Hao Li, Chenxin Tao, Xizhou Zhu, Xiaogang Wang, Gao Huang, Jifeng Dai</i></p> | 14307 |
| <p>Combining Physics and Machine Learning for Network Flow Estimation</p> <p><i>Arlei Lopes Da Silva, Furkan Kocayusufoglu, Saber Jafarpour, Francesco Bullo, Ananthram Swami, Ambuj Singh</i></p> | 14318 |
| <p>NeMo: Neural Mesh Models of Contrastive Features for Robust 3D Pose Estimation</p> <p><i>Angtian Wang, Adam Kortylewski, Alan Yuille</i></p> | 14336 |
| <p>How to Find Your Friendly Neighborhood: Graph Attention Design with Self-Supervision</p> <p><i>Dongkwan Kim, Alice Oh</i></p> | 14353 |
| <p>MiCE: Mixture of Contrastive Experts for Unsupervised Image Clustering</p> <p><i>Tsung Wei Tsai, Chongxuan Li, Jun Zhu</i></p> | 14378 |
| <p>Anytime Sampling for Autoregressive Models Via Ordered Autoencoding</p> <p><i>Yilun Xu, Yang Song, Sahaj Garg, Linyuan Gong, Rui Shu, Aditya Grover, Stefano Ermon</i></p> | 14403 |
| <p>Initialization and Regularization of Factorized Neural Layers</p> <p><i>Mikhail Khodak, Neil A. Tenenholz, Lester Mackey, Nicolo Fusi</i></p> | 14426 |
| <p>CO2: Consistent Contrast for Unsupervised Visual Representation Learning</p> <p><i>Chen Wei, Huiyu Wang, Wei Shen, Alan Yuille</i></p> | 14446 |
| <p>DC3: A Learning Method for Optimization with Hard Constraints</p> <p><i>Priya L. Donti, David Rolnick, J Zico Kolter</i></p> | 14459 |
| <p>Enforcing Robust Control Guarantees Within Neural Network Policies</p> <p><i>Priya L. Donti, Melrose Roderick, Mahyar Fazlyab, J Zico Kolter</i></p> | 14476 |
| <p>Robust and Generalizable Visual Representation Learning Via Random Convolutions</p> <p><i>Zhenlin Xu, Deyi Liu, Junlin Yang, Colin Raffel, Marc Niethammer</i></p> | 14502 |

| | |
|--|-------|
| <p>WrapNet: Neural Net Inference with Ultra-Low-Precision Arithmetic</p> <p><i>Renkun Ni, Hong-Min Chu, Oscar Castaneda, Ping-Yeh Chiang, Christoph Studer, Tom Goldstein</i></p> | 14521 |
| <p>Rank the Episodes: A Simple Approach for Exploration in Procedurally-Generated Environments</p> <p><i>Daochen Zha, Wenye Ma, Lei Yuan, Xia Hu, Ji Liu</i></p> | 14537 |
| <p>Stochastic Security: Adversarial Defense Using Long-Run Dynamics of Energy-Based Models</p> <p><i>Mitch Hill, Jonathan Craig Mitchell, Song-Chun Zhu</i></p> | 14562 |
| <p>Nearest Neighbor Machine Translation</p> <p><i>Urvashi Khandelwal, Angela Fan, Dan Jurafsky, Luke Zettlemoyer, Mike Lewis</i></p> | 14580 |
| <p>Learning Hyperbolic Representations of Topological Features</p> <p><i>Panagiotis Kyriakis, Iordanis Fostropoulos, Paul Bogdan</i></p> | 14594 |
| <p>Selectivity Considered Harmful: Evaluating the Causal Impact of Class Selectivity in DNNs</p> <p><i>Matthew L Leavitt, Ari S. Morcos</i></p> | 14609 |
| <p>Integrating Categorical Semantics into Unsupervised Domain Translation</p> <p><i>Samuel Lavoie-Marchildon, Faruk Ahmed, Aaron Courville</i></p> | 14647 |
| <p>Zero-Shot Synthesis with Group-Supervised Learning</p> <p><i>Yunhao Ge, Sami Abu-El-Hajja, Gan Xin, Laurent Itti</i></p> | 14669 |
| <p>VA-RED2: Video Adaptive Redundancy Reduction</p> <p><i>Bowen Pan, Rameswar Panda, Camilo Luciano Fosco, Chung-Ching Lin, Alex J Andonian, Yue Meng, Kate Saenko, Aude Oliva, Rogerio Feris</i></p> | 14685 |
| <p>BERTology Meets Biology: Interpreting Attention in Protein Language Models</p> <p><i>Jesse Vig, Ali Madani, Lav R. Varshney, Caiming Xiong, Richard Socher, Nazneen Rajani</i></p> | 14704 |
| <p>Trajectory Prediction Using Equivariant Continuous Convolution</p> <p><i>Robin Walters, Jinxi Li, Rose Yu</i></p> | 14728 |
| <p>Unsupervised Meta-Learning Through Latent-Space Interpolation in Generative Models</p> <p><i>Siavash Khodadadeh, Sharare Zehtabian, Saeed Vahidian, Weijia Wang, Bill Lin, Ladislau Boloni</i></p> | 14745 |
| <p>Heteroskedastic and Imbalanced Deep Learning with Adaptive Regularization</p> <p><i>Kaidi Cao, Yining Chen, Junwei Lu, Nikos Archiga, Adrien Gaidon, Tengyu Ma</i></p> | 14760 |
| <p>Implicit Under-Parameterization Inhibits Data-Efficient Deep Reinforcement Learning</p> <p><i>Aviral Kumar, Rishabh Agarwal, Dibya Ghosh, Sergey Levine</i></p> | 14776 |
| <p>Bowtie Networks: Generative Modeling for Joint Few-Shot Recognition and Novel-View Synthesis</p> <p><i>Zhipeng Bao, Yu-Xiong Wang, Martial Hebert</i></p> | 14812 |

| | |
|---|-------|
| <p>Saliency is a Possible Red Herring When Diagnosing Poor Generalization</p> <p><i>Joseph D Viviano, Becks Simpson, Francis Dutil, Yoshua Bengio, Joseph Paul Cohen</i></p> | 14830 |
| <p>What Can You Learn from Your Muscles? Learning Visual Representation from Human Interactions</p> <p><i>Kiana Ehsani, Daniel Gordon, Thomas Hai Dang Nguyen, Roozbeh Mottaghi, Ali Farhadi</i></p> | 14855 |
| <p>Conservative Safety Critics for Exploration</p> <p><i>Homanga Bharadhwaj, Aviral Kumar, Nicholas Rhinehart, Sergey Levine, Florian Shkurti, Animesh Garg</i></p> | 14870 |
| <p>Multi-Time Attention Networks for Irregularly Sampled Time Series</p> <p><i>Satya Narayan Shukla, Benjamin Marlin</i></p> | 14897 |
| <p>Graph Traversal with Tensor Functionals: A Meta-Algorithm for Scalable Learning</p> <p><i>Elan Sopher Markowitz, Keshav Balasubramanian, Mehrnoosh Mirtaheri, Sami Abu-El-Haija, Bryan Perozzi, Greg Ver Steeg, Aram Galstyan</i></p> | 14912 |
| <p>What They Do When in Doubt: A Study of Inductive Biases in Seq2seq Learners</p> <p><i>Eugene Kharitonov, Rahma Chaabouni</i></p> | 14929 |
| <p>Tomographic Auto-Encoder: Unsupervised Bayesian Recovery of Corrupted Data</p> <p><i>Francesco Tonolini, Pablo Garcia Moreno, Andreas Damianou, Roderick Murray-Smith</i></p> | 14951 |
| <p>OPAL: Offline Primitive Discovery for Accelerating Offline Reinforcement Learning</p> <p><i>Anurag Ajay, Aviral Kumar, Pulkit Agrawal, Sergey Levine, Ofir Nachum</i></p> | 14975 |
| <p>Knowledge Distillation Via Softmax Regression Representation Learning</p> <p><i>Jing Yang, Brais Martinez, Adrian Bulat, Georgios Tzimiropoulos</i></p> | 14998 |
| <p>Large Associative Memory Problem in Neurobiology and Machine Learning</p> <p><i>Dmitry Krotov, John J. Hopfield</i></p> | 15012 |
| <p>Remembering for the Right Reasons: Explanations Reduce Catastrophic Forgetting</p> <p><i>Sayna Ebrahimi, Suzanne Petryk, Akash Gokul, William Gan, Joseph E. Gonzalez, Marcus Rohrbach, Trevor Darrell</i></p> | 15025 |
| <p>Deep Networks and the Multiple Manifold Problem</p> <p><i>Sam Buchanan, Dar Gilboa, John Wright</i></p> | 15040 |
| <p>Interpreting Knowledge Graph Relation Representation from Word Embeddings</p> <p><i>Carl Allen, Ivana Balazevic, Timothy Hospedales</i></p> | 15292 |
| <p>Learning Perturbation Sets for Robust Machine Learning</p> <p><i>Eric Wong, J Zico Kolter</i></p> | 15308 |
| <p>A Trainable Optimal Transport Embedding for Feature Aggregation and Its Relationship to Attention</p> <p><i>Grégoire Mialon, Dexiong Chen, Alexandre D'Aspremont, Julien Mairal</i></p> | 15342 |

| | |
|--|-------|
| <i>RODE: Learning Roles to Decompose Multi-Agent Tasks</i> | 15365 |
| <i>Tonghan Wang, Tarun Gupta, Anuj Mahajan, Bei Peng, Shimon Whiteson, Chongjie Zhang</i> | |
| <i>Scalable Bayesian Inverse Reinforcement Learning</i> | 15389 |
| <i>Alex James Chan, Mihaela Van Der Schaar</i> | |
| <i>Learning "What-If" Explanations for Sequential Decision-Making</i> | 15403 |
| <i>Ioana Bica, Daniel Jarrett, Alihan Hüyük, Mihaela Van Der Schaar</i> | |
| <i>Learning Advanced Mathematical Computations from Examples</i> | 15422 |
| <i>Francois Charton, Amaury Hayat, Guillaume Lample</i> | |
| <i>Repurposing Pretrained Models for Robust Out-Of-Domain Few-Shot Learning</i> | 15447 |
| <i>Namyeong Kwon, Hwidong Na, Gabriel Huang, Simon Lacoste-Julien</i> | |
| <i>Empirical Analysis of Unlabeled Entity Problem in Named Entity Recognition</i> | 15467 |
| <i>Yangming Li, Lemao Liu, Shuming Shi</i> | |
| <i>Calibration of Neural Networks Using Splines</i> | 15479 |
| <i>Kartik Gupta, Amir Rahimi, Thalaisyasingam Ajanthan, Thomas Mensink, Cristian Sminchisescu, Richard Hartley</i> | |
| <i>Probing BERT in Hyperbolic Spaces</i> | 15498 |
| <i>Boli Chen, Yao Fu, Guangwei Xu, Pengjun Xie, Chuanqi Tan, Mosha Chen, Liping Jing</i> | |
| <i>Refining Deep Generative Models Via Discriminator Gradient Flow</i> | 15522 |
| <i>Abdul Fatir Ansari, Ming Liang Ang, Harold Soh</i> | |
| <i>Coping with Label Shift Via Distributionally Robust Optimisation</i> | 15546 |
| <i>Jingzhao Zhang, Aditya Krishna Menon, Andreas Veit, Srinadh Bhojanapalli, Sanjiv Kumar, Suvrit Sra</i> | |
| <i>Variational State-Space Models for Localisation and Dense 3D Mapping in 6 DoF</i> | 15565 |
| <i>Atanas Mirchev, Baris Kayalibay, Patrick Van Der Smagt, Justin Bayer</i> | |
| <i>Few-Shot Bayesian Optimization with Deep Kernel Surrogates</i> | 15588 |
| <i>Martin Wistuba, Josif Grabocka</i> | |
| <i>-Mix: A Domain-Agnostic Strategy for Contrastive Representation Learning</i> | 15603 |
| <i>Kibok Lee, Yian Zhu, Kihyuk Sohn, Chun-Liang Li, Jinwoo Shin, Honglak Lee</i> | |
| <i>Graph Information Bottleneck for Subgraph Recognition</i> | 15622 |
| <i>Junchi Yu, Tingyang Xu, Yu Rong, Yatao Bian, Junzhou Huang, Ran He</i> | |
| <i>Rethinking Positional Encoding in Language Pre-Training</i> | 15641 |
| <i>Guolin Ke, Di He, Tie-Yan Liu</i> | |
| <i>Practical Massively Parallel Monte-Carlo Tree Search Applied to Molecular Design</i> | 15655 |
| <i>Xiufeng Yang, Tanuj Aasawat, Kazuki Yoshizoe</i> | |

| | |
|---|-------|
| <p>When Does Preconditioning Help Or Hurt Generalization?</p> <p><i>Shun-Ichi Amari, Jimmy Ba, Roger Baker Grosse, Xuechen Li, Atsushi Nitanda, Taiji Suzuki, Denny Wu, Ji Xu</i></p> | 15674 |
| <p>ARMOURED: Adversarially Robust MOdels Using Unlabeled Data by REgularizing Diversity</p> <p><i>Kangkang Lu, Cuong Manh Nguyen, Xun Xu, Kiran Krishnamachari, Yu Jing Goh, Chuan-Sheng Foo</i></p> | 15714 |
| <p>Learning Energy-Based Generative Models Via Coarse-To-Fine Expanding and Sampling</p> <p><i>Yang Zhao, Jianwen Xie, Ping Li</i></p> | 15735 |
| <p>Multi-Prize Lottery Ticket Hypothesis: Finding Accurate Binary Neural Networks by Pruning a Randomly Weighted Network</p> <p><i>James Diffenderfer, Bhavya Kailkhura</i></p> | 15764 |
| <p>Learning Accurate Entropy Model with Global Reference for Image Compression</p> <p><i>Yichen Qian, Zhiyu Tan, Xiuyu Sun, Ming Lin, Dongyang Li, Zhenhong Sun, Li Hao, Rong Jin</i></p> | 15787 |
| <p>Convex Potential Flows: Universal Probability Distributions with Optimal Transport and Convex Optimization</p> <p><i>Chin-Wei Huang, Ricky T. Q. Chen, Christos Tsirigotis, Aaron Courville</i></p> | 15804 |
| <p>Greedy-GQ with Variance Reduction: Finite-Time Analysis and Improved Complexity</p> <p><i>Shaocong Ma, Ziyi Chen, Yi Zhou, Shaofeng Zou</i></p> | 15817 |
| <p>Large Batch Simulation for Deep Reinforcement Learning</p> <p><i>Brennan Shacklett, Erik Wijmans, Aleksei Petrenko, Manolis Savva, Dhruv Batra, Vladlen Koltun, Kayvon Fatahalian</i></p> | 15847 |
| <p>Hopper: Multi-Hop Transformer for Spatiotemporal Reasoning</p> <p><i>Honglu Zhou, Asim Kadav, Farley Lai, Alexandru Niculescu-Mizil, Martin Renqiang Min, Mubbasir Kapadia, Hans Peter Graf</i></p> | 15864 |
| <p>Efficient Reinforcement Learning in Factored MDPs with Application to Constrained RL</p> <p><i>Xiaoyu Chen, Jiachen Hu, Lihong Li, Liwei Wang</i></p> | 15893 |
| <p>Unbiased Teacher for Semi-Supervised Object Detection</p> <p><i>Yen-Cheng Liu, Chih-Yao Ma, Zijian He, Chia-Wen Kuo, Kan Chen, Peizhao Zhang, Bichen Wu, Zsolt Kira, Peter Vajda</i></p> | 15939 |
| <p>MELR: Meta-Learning Via Modeling Episode-Level Relationships for Few-Shot Learning</p> <p><i>Nanyi Fei, Zhiwu Lu, Tao Xiang, Songfang Huang</i></p> | 15956 |
| <p>Partitioned Learned Bloom Filters</p> <p><i>Kapil Vaidya, Eric Knorr, Michael Mitzenmacher, Tim Kraska</i></p> | 15976 |
| <p>Wasserstein Embedding for Graph Learning</p> <p><i>Soheil Kolouri, Navid Naderializadeh, Gustavo K. Rohde, Heiko Hoffmann</i></p> | 15993 |

| | |
|---|-------|
| High-Capacity Expert Binary Networks <i>Adrian Bulat, Brais Martinez, Georgios Tzimiropoulos</i> | 16013 |
| SAFENet: A Secure, Accurate and Fast Neural Network Inference <i>Qian Lou, Yilin Shen, Hongxia Jin, Lei Jiang</i> | 16032 |
| Learning Manifold Patch-Based Representations of Man-Made Shapes <i>Dmitriy Smirnov, Mikhail Bessmeltsev, Justin Solomon</i> | 16045 |
| Universal Approximation Power of Deep Residual Neural Networks Via Nonlinear Control Theory <i>Paulo Tabuada, Bahman Ghahsifard</i> | 16069 |
| Learning Neural Event Functions for Ordinary Differential Equations <i>Ricky T. Q. Chen, Brandon Amos, Maximilian Nickel</i> | 16080 |
| Neural Spatio-Temporal Point Processes <i>Ricky T. Q. Chen, Brandon Amos, Maximilian Nickel</i> | 16095 |
| Proximal Gradient Descent-Ascent: Variable Convergence Under K Geometry <i>Ziyi Chen, Yi Zhou, Tengyu Xu, Yingbin Liang</i> | 16114 |
| Adaptive Universal Generalized PageRank Graph Neural Network <i>Eli Chien, Jianhao Peng, Pan Li, Olgica Milenkovic</i> | 16138 |
| Open Question Answering Over Tables and Text <i>Wenhu Chen, Ming-Wei Chang, Eva Schlinger, William Yang Wang, William W. Cohen</i> | 16162 |
| Text Generation by Learning from Demonstrations <i>Richard Yuanzhe Pang, He He</i> | 16180 |
| Tilted Empirical Risk Minimization <i>Tian Li, Ahmad Beirami, Maziar Sanjabi, Virginia Smith</i> | 16202 |
| Explainable Subgraph Reasoning for Forecasting on Temporal Knowledge Graphs <i>Zhen Han, Peng Chen, Yunpu Ma, Volker Tresp</i> | 16246 |
| Bayesian Context Aggregation for Neural Processes <i>Michael Volpp, Fabian Flürenbrock, Lukas Grossberger, Christian Daniel, Gerhard Neumann</i> | 16270 |
| Conformation-Guided Molecular Representation with Hamiltonian Neural Networks <i>Ziyao Li, Shuwen Yang, Guojie Song, Lingsheng Cai</i> | 16299 |
| Learning with AMiGo: Adversarially Motivated Intrinsic Goals <i>Andres Campero, Roberta Raileanu, Heinrich Kuttler, Joshua B. Tenenbaum, Tim Rocktäschel, Edward Grefenstette</i> | 16310 |

| | |
|--|-------|
| <p>Training with Quantization Noise for Extreme Model Compression</p> <p><i>Pierre Stock, Angela Fan, Benjamin Graham, Edouard Grave, Rémi Gribonval, Herve Jegou, Armand Joulin</i></p> | 16328 |
| <p>Interpreting and Boosting Dropout from a Game-Theoretic View</p> <p><i>Hao Zhang, Sen Li, Yinchao Ma, Mingjie Li, Yichen Xie, Quanshi Zhang</i></p> | 16348 |
| <p>VTNet: Visual Transformer Network for Object Goal Navigation</p> <p><i>Heming Du, Xin Yu, Liang Zheng</i></p> | 16368 |
| <p>Exemplary Natural Images Explain CNN Activations Better than State-Of-The-Art Feature Visualization</p> <p><i>Judy Borowski, Roland Simon Zimmermann, Judith Schepers, Robert Geirhos, Thomas S. A. Wallis, Matthias Bethge, Wieland Brendel</i></p> | 16384 |
| <p>Multi-Class Uncertainty Calibration Via Mutual Information Maximization-Based Binning</p> <p><i>Kanil Patel, William H. Beluch, Bin Yang, Michael Pfeiffer, Dan Zhang</i></p> | 16425 |
| <p>A Discriminative Gaussian Mixture Model with Sparsity</p> <p><i>Hideaki Hayashi, Seiichi Uchida</i></p> | 16457 |
| <p>Trusted Multi-View Classification</p> <p><i>Zongbo Han, Changqing Zhang, Huazhu Fu, Joey Tianyi Zhou</i></p> | 16471 |
| <p>IEPT: Instance-Level and Episode-Level Pretext Tasks for Few-Shot Learning</p> <p><i>Manli Zhang, Jianhong Zhang, Zhiwu Lu, Tao Xiang, Mingyu Ding, Songfang Huang</i></p> | 16482 |
| <p>Deep Encoder, Shallow Decoder: Reevaluating Non-Autoregressive Machine Translation</p> <p><i>Jungo Kasai, Nikolaos Pappas, Hao Peng, James Cross, Noah Smith</i></p> | 16498 |
| <p>Effective Abstract Reasoning with Dual-Contrast Network</p> <p><i>Tao Zhuo, Mohan Kankanhalli</i></p> | 16514 |
| <p>On Position Embeddings in BERT</p> <p><i>Benyou Wang, Lifeng Shang, Christina Lioma, Xin Jiang, Hao Yang, Qun Liu, Jakob Grue Simonsen</i></p> | 16527 |
| <p>Neural Pruning Via Growing Regularization</p> <p><i>Huan Wang, Can Qin, Yulun Zhang, Yun Fu</i></p> | 16548 |
| <p>Mixed-Features Vectors and Subspace Splitting</p> <p><i>Alejandro Pimentel-Alarcón, Daniel L. Pimentel-Alarcón</i></p> | 16564 |
| <p>Hierarchical Reinforcement Learning by Discovering Intrinsic Options</p> <p><i>Jesse Zhang, Haonan Yu, Wei Xu</i></p> | 16579 |
| <p>Sharper Generalization Bounds for Learning with Gradient-Dominated Objective Functions</p> <p><i>Yunwen Lei, Yiming Ying</i></p> | 16597 |

| | |
|--|-------|
| Representation Learning for Sequence Data with Deep Autoencoding Predictive Components <i>Junwen Bai, Weiran Wang, Yingbo Zhou, Caiming Xiong</i> | 16620 |
| Average-Case Acceleration for Bilinear Games and Normal Matrices <i>Carles Domingo-Enrich, Fabian Pedregosa, Damien Scieur</i> | 16639 |
| Learning Task-General Representations with Generative Neuro-Symbolic Modeling <i>Reuben Feinman, Brenden M. Lake</i> | 16661 |
| Is Label Smoothing Truly Incompatible with Knowledge Distillation: An Empirical Study <i>Zhiqiang Shen, Zechun Liu, Dejia Xu, Zitian Chen, Kwang-Ting Cheng, Marios Savvides</i> | 16681 |
| The Unreasonable Effectiveness of Patches in Deep Convolutional Kernels Methods <i>Louis Thiry, Michael Arbel, Eugene Belilovsky, Edouard Oyallon</i> | 16698 |
| Graph Coarsening with Neural Networks <i>Chen Cai, Ding kang Wang, Yusu Wang</i> | 16712 |
| On the Universality of the Double Descent Peak in Ridgeless Regression <i>David Holzmüller</i> | 16739 |
| Deep Repulsive Clustering of Ordered Data Based on Order-Identity Decomposition <i>Seon-Ho Lee, Chang-Su Kim</i> | 16785 |
| Rapid Task-Solving in Novel Environments <i>Samuel Ritter, Ryan Faulkner, Laurent Sartran, Adam Santoro, Matthew Botvinick, David Raposo</i> | 16812 |
| DINO: A Conditional Energy-Based GAN for Domain Translation <i>Konstantinos Vougioukas, Stavros Petridis, Maja Pantic</i> | 16827 |
| Removing Undesirable Feature Contributions Using Out-Of-Distribution Data <i>Saehyung Lee, Changhwa Park, Hyungyu Lee, Jihun Yi, Jonghyun Lee, Sungroh Yoon</i> | 16846 |
| Accurate Learning of Graph Representations with Graph Multiset Pooling <i>Jinheon Baek, Minki Kang, Sung Ju Hwang</i> | 16864 |
| Federated Semi-Supervised Learning with Inter-Client Consistency & Disjoint Learning <i>Wonyong Jeong, Jaehong Yoon, Eunho Yang, Sung Ju Hwang</i> | 16886 |
| Contrastive Learning with Adversarial Perturbations for Conditional Text Generation <i>Seanie Lee, Dong Bok Lee, Sung Ju Hwang</i> | 16901 |
| Optimal Conversion of Conventional Artificial Neural Networks to Spiking Neural Networks <i>Shikuang Deng, Shi Gu</i> | 16926 |
| Efficient Continual Learning with Modular Networks and Task-Driven Priors <i>Tom Veniat, Ludovic Denoyer, Marcaurelio Ranzato</i> | 16940 |

| | |
|---|-------|
| <p>On the Universality of Rotation Equivariant Point Cloud Networks</p> <p>Nadav Dym, Haggai Maron</p> | 16970 |
| <p>Neural Learning of One-Of-Many Solutions for Combinatorial Problems in Structured Output Spaces</p> <p>Yatin Nandwani, Deepanshu Jindal, Mausam ., Parag Singla</p> | 16992 |
| <p>GAN2GAN: Generative Noise Learning for Blind Denoising with Single Noisy Images</p> <p>Sungmin Cha, Taeon Park, Byeongjoon Kim, Jongduk Baek, Taesup Moon</p> | 17013 |
| <p>CPR: Classifier-Projection Regularization for Continual Learning</p> <p>Sungmin Cha, Hsiang Hsu, Taebaek Hwang, Flavio Calmon, Taesup Moon</p> | 17024 |
| <p>On the Dynamics of Training Attention Models</p> <p>Haoye Lu, Yongyi Mao, Amiya Nayak</p> | 17036 |
| <p>Model-Based Offline Planning</p> <p>Arthur Argenson, Gabriel Dulac-Arnold</p> | 17053 |
| <p>Modelling Hierarchical Structure Between Dialogue Policy and Natural Language Generator with Option Framework for Task-Oriented Dialogue System</p> <p>Jianhong Wang, Yuan Zhang, Tae-Kyun Kim, Yunjie Gu</p> | 17078 |
| <p>Distilling Knowledge from Reader to Retriever for Question Answering</p> <p>Gautier Izacard, Edouard Grave</p> | 17098 |
| <p>Efficient Certified Defenses Against Patch Attacks on Image Classifiers</p> <p>Jan Hendrik Metzen, Maksym Yatsura</p> | 17111 |
| <p>Taking Notes on the Fly Helps Language Pre-Training</p> <p>Qiyu Wu, Chen Xing, Yatao Li, Guolin Ke, Di He, Tie-Yan Liu</p> | 17127 |
| <p>Graph Edit Networks</p> <p>Benjamin Paassen, Daniele Grattarola, Daniele Zambon, Cesare Alippi, Barbara Eva Hammer</p> | 17141 |
| <p>FOCAL: Efficient Fully-Offline Meta-Reinforcement Learning Via Distance Metric Learning and Behavior Regularization</p> <p>Lanqing Li, Rui Yang, Dijun Luo</p> | 17175 |
| <p>Effective and Efficient Vote Attack on Capsule Networks</p> <p>Jindong Gu, Baoyuan Wu, Volker Tresp</p> | 17198 |
| <p>Rapid Neural Architecture Search by Learning to Generate Graphs from Datasets</p> <p>Hayeon Lee, Eunyoung Hyung, Sung Ju Hwang</p> | 17214 |
| <p>Impact of Representation Learning in Linear Bandits</p> <p>Jiaqi Yang, Wei Hu, Jason D. Lee, Simon Shaolei Du</p> | 17230 |

| | |
|--|-------|
| <p>Creative Sketch Generation</p> <p>Songwei Ge, Vedanuj Goswami, Larry Zitnick, Devi Parikh</p> | 17255 |
| <p>Self-Supervised Representation Learning with Relative Predictive Coding</p> <p>Yao-Hung Hubert Tsai, Martin Q. Ma, Muqiao Yang, Han Zhao, Louis-Philippe Morency, Ruslan Salakhutdinov</p> | 17281 |
| <p>One Network Fits All? Modular Versus Monolithic Task Formulations in Neural Networks</p> <p>Atish Agarwala, Abhimanyu Das, Brendan Juba, Rina Panigrahy, Vatsal Sharan, Xin Wang, Qiuyi Zhang</p> | 17304 |
| <p>Conditionally Adaptive Multi-Task Learning: Improving Transfer Learning in NLP Using Fewer Parameters & Less Data</p> <p>Jonathan Pilault, Amine El Hattami, Christopher Pal</p> | 17334 |
| <p>A Universal Representation Transformer Layer for Few-Shot Image Classification</p> <p>Lu Liu, William L. Hamilton, Guodong Long, Jing Jiang, Hugo Larochelle</p> | 17355 |
| <p>Isometric Propagation Network for Generalized Zero-Shot Learning</p> <p>Lu Liu, Tianyi Zhou, Guodong Long, Jing Jiang, Xuanyi Dong, Chengqi Zhang</p> | 17367 |
| <p>Towards Impartial Multi-Task Learning</p> <p>Liyang Liu, Yi Li, Zhanghui Kuang, Jing-Hao Xue, Yimin Chen, Wenming Yang, Qingmin Liao, Wayne Zhang</p> | 17380 |
| <p>On Learning Universal Representations Across Languages</p> <p>Xiangpeng Wei, Rongxiang Weng, Yue Hu, Luxi Xing, Heng Yu, Weihua Luo</p> | 17401 |
| <p>Isotropy in the Contextual Embedding Space: Clusters and Manifolds</p> <p>Xingyu Cai, Jiaji Huang, Yuchen Bian, Kenneth Church</p> | 17419 |
| <p>MoPro: Webly Supervised Learning with Momentum Prototypes</p> <p>Junnan Li, Caiming Xiong, Steven Hoi</p> | 17441 |
| <p>GraphCodeBERT: Pre-Training Code Representations with Data Flow</p> <p>Daya Guo, Shuo Ren, Shuai Lu, Zhangyin Feng, Duyu Tang, Shujie Liu, Long Zhou, Nan Duan, Alexey Svyatkovskiy, Shengyu Fu, Michele Tufano, Shao Kun Deng, Colin Clement, Dawn Drain, Neel Sundaresan, Jian Yin, Daxin Jiang, Ming Zhou</p> | 17456 |
| <p>Enjoy Your Editing: Controllable GANs for Image Editing Via Latent Space Navigation</p> <p>Peiye Zhuang, Oluwasanmi O Koyejo, Alex Schwing</p> | 17474 |
| <p>Fourier Neural Operator for Parametric Partial Differential Equations</p> <p>Zongyi Li, Nikola Borislavov Kovachki, Kamyar Aizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, Anima Anandkumar</p> | 17490 |
| <p>Training BatchNorm and Only BatchNorm: On the Expressive Power of Random Features in CNNs</p> <p>Jonathan Frankle, David J. Schwab, Ari S. Morcos</p> | 17506 |

| | |
|--|-------|
| <p>Towards Faster and Stabilized GAN Training for High-Fidelity Few-Shot Image Synthesis</p> <p>Bingchen Liu, Yizhe Zhu, Kunpeng Song, Ahmed Elgammal</p> | 17534 |
| <p>Targeted Attack Against Deep Neural Networks Via Flipping Limited Weight Bits</p> <p>Jiawang Bai, Baoyuan Wu, Yong Zhang, Yiming Li, Zhifeng Li, Shu-Tao Xia</p> | 17547 |
| <p>FedBN: Federated Learning on Non-IID Features Via Local Batch Normalization</p> <p>Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, Qi Dou</p> | 17566 |
| <p>AdamP: Slowing Down the Slowdown for Momentum Optimizers on Scale-Invariant Weights</p> <p>Byeongho Heo, Sanghyuk Chun, Seong Joon Oh, Dongyoon Han, Sangdoon Yun, Gyuwan Kim, Youngjung Uh, Jung-Woo Ha</p> | 17593 |
| <p>Learning Subgoal Representations with Slow Dynamics</p> <p>Siyuan Li, Lulu Zheng, Jianhao Wang, Chongjie Zhang</p> | 17620 |
| <p>A Unified Approach to Interpreting and Boosting Adversarial Transferability</p> <p>Xin Wang, Jie Ren, Shuyun Lin, Xiangming Zhu, Yisen Wang, Quanshi Zhang</p> | 17638 |
| <p>Perceptual Adversarial Robustness: Defense Against Unseen Threat Models</p> <p>Cassidy Laidlaw, Sahil Singla, Soheil Feizi</p> | 17673 |
| <p>Better Fine-Tuning by Reducing Representational Collapse</p> <p>Armen Aghajanyan, Akshat Shrivastava, Anchit Gupta, Naman Goyal, Luke Zettlemoyer, Sonal Gupta</p> | 17698 |
| <p>Learning N:M Fine-Grained Structured Sparse Neural Networks from Scratch</p> <p>Aojun Zhou, Yukun Ma, Junnan Zhu, Jianbo Liu, Zhijie Zhang, Kun Yuan, Wenxiu Sun, Hongsheng Li</p> | 17710 |
| <p>Active Contrastive Learning of Audio-Visual Video Representations</p> <p>Shuang Ma, Zhaoyang Zeng, Daniel McDuff, Yale Song</p> | 17725 |
| <p>PseudoSeg: Designing Pseudo Labels for Semantic Segmentation</p> <p>Yuliang Zou, Zizhao Zhang, Han Zhang, Chun-Liang Li, Xiao Bian, Jia-Bin Huang, Tomas Pfister</p> | 17744 |
| <p>Counterfactual Generative Networks</p> <p>Axel Sauer, Andreas Geiger</p> | 17762 |
| <p>Contemplating Real-World Object Classification</p> <p>Ali Borji</p> | 17787 |
| <p>Explainable Deep One-Class Classification</p> <p>Philipp Liznerski, Lukas Ruff, Robert A. Vandermeulen, Billy Joe Franks, Marius Kloft, Klaus Robert Muller</p> | 17834 |
| <p>Batch Reinforcement Learning Through Continuation Method</p> <p>Yijie Guo, Shengyu Feng, Nicolas Le Roux, Ed Chi, Honglak Lee, Minmin Chen</p> | 17859 |
| <p>Protecting DNNs from Theft Using an Ensemble of Diverse Models</p> <p>Sanjay Kariyappa, Atul Prakash, Moinuddin K Qureshi</p> | 17870 |

| | |
|---|-------|
| Domain Generalization with MixStyle Kaiyang Zhou, Yongxin Yang, Yu Qiao, Tao Xiang | 17883 |
| Efficient Inference of Flexible Interaction in Spiking-Neuron Networks Feng Zhou, Yixuan Zhang, Jun Zhu | 17898 |
| DICE: Diversity in Deep Ensembles Via Conditional Redundancy Adversarial Estimation Alexandre Rame, Matthieu Cord | 17914 |
| Universal Weakly Supervised Segmentation by Pixel-To-Segment Contrastive Learning Tsung-Wei Ke, Jyh-Jing Hwang, Stella Yu | 17945 |
| Shape-Texture Debiased Neural Network Training Yingwei Li, Qihang Yu, Mingxing Tan, Jieru Mei, Peng Tang, Wei Shen, Alan Yuille, Cihang Xie | 17963 |
| Interactive Weak Supervision: Learning Useful Heuristics for Data Labeling Benedikt Boecking, Willie Neiswanger, Eric Xing, Artur Dubrawski | 17975 |
| MoVie: Revisiting Modulated Convolutions for Visual Counting and Beyond Duy Kien Nguyen, Vedanuj Goswami, Xinlei Chen | 18002 |
| Improve Object Detection with Feature-Based Knowledge Distillation: Towards Accurate and Efficient Detectors Linfeng Zhang, Kaisheng Ma | 18018 |
| Revisiting Dynamic Convolution Via Matrix Decomposition Yunsheng Li, Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Ye Yu, Lu Yuan, Zicheng Liu, Mei Chen, Nuno Vasconcelos | 18032 |
| DynaTune: Dynamic Tensor Program Optimization in Deep Neural Network Compilation Minjia Zhang, Menghao Li, Chi Wang, Mingqin Li | 18043 |
| MALI: A Memory Efficient and Reverse Accurate Integrator for Neural ODEs Juntang Zhuang, Nicha C Dvornek, Sekhar Tatikonda, James S Duncan | 18058 |
| Bag of Tricks for Adversarial Training Tianyu Pang, Xiao Yang, Yinpeng Dong, Hang Su, Jun Zhu | 18080 |
| Learning to Generate 3D Shapes with Generative Cellular Automata Dongsu Zhang, Changwoon Choi, Jeonghwan Kim, Young Min Kim | 18101 |
| SEED: Self-Supervised Distillation for Visual Representation Zhiyuan Fang, Jianfeng Wang, Lijuan Wang, Lei Zhang, Yezhou Yang, Zicheng Liu | 18123 |
| PDE-Driven Spatiotemporal Disentanglement J r mie Don , Jean-Yves Franceschi, Sylvain Lamprier, Patrick Gallinari | 18144 |
| Rethinking Soft Labels for Knowledge Distillation: A Bias-Variance Tradeoff Perspective Helong Zhou, Liangchen Song, Jiajie Chen, Ye Zhou, Guoli Wang, Junsong Yuan, Qian Zhang | 18173 |

| | |
|--|-------|
| <p>Distributed Momentum for Byzantine-Resilient Stochastic Gradient Descent</p> <p><i>El Mahdi El Mhamdi, Rachid Guerraoui, Sébastien Rouault</i></p> | 18188 |
| <p>Scaling the Convex Barrier with Active Sets</p> <p><i>Alessandro De Palma, Harkirat Behl, Rudy R Bunel, Philip Torr, M. Pawan Kumar</i></p> | 18225 |
| <p>BSQ: Exploring Bit-Level Sparsity for Mixed-Precision Neural Network Quantization</p> <p><i>Huanrui Yang, Lin Duan, Yiran Chen, Hai Li</i></p> | 18252 |
| <p>Dual-Mode ASR: Unify and Improve Streaming ASR with Full-Context Modeling</p> <p><i>Jiahui Yu, Wei Han, Anmol Gulati, Chung-Cheng Chiu, Bo Li, Tara N Sainath, Yonghui Wu, Ruoming Pang</i></p> | 18268 |
| <p>Policy-Driven Attack: Learning to Query for Hard-Label Black-Box Adversarial Examples</p> <p><i>Ziang Yan, Yiwon Guo, Jian Liang, Changshui Zhang</i></p> | 18281 |
| <p>PSTNet: Point Spatio-Temporal Convolution on Point Cloud Sequences</p> <p><i>Hehe Fan, Xin Yu, Yuhang Ding, Yi Yang, Mohan Kankanhalli</i></p> | 18296 |
| <p>Network Pruning that Matters: A Case Study on Retraining Variants</p> <p><i>Duong Hoang Le, Binh-Son Hua</i></p> | 18319 |
| <p>EEC: Learning to Encode and Regenerate Images for Continual Learning</p> <p><i>Ali Ayub, Alan Wagner</i></p> | 18337 |
| <p>DARTS-: Robustly Stepping Out of Performance Collapse Without Indicators</p> <p><i>Xiangxiang Chu, Xiaoxing Wang, Bo Zhang, Shun Lu, Xiaolin Wei, Junchi Yan</i></p> | 18353 |
| <p>BiPointNet: Binary Neural Network for Point Clouds</p> <p><i>Haotong Qin, Zhongang Cai, Mingyuan Zhang, Yifu Ding, Haiyu Zhao, Shuai Yi, Xianglong Liu, Hao Su</i></p> | 18375 |
| <p>AdaFuse: Adaptive Temporal Fusion Network for Efficient Action Recognition</p> <p><i>Yue Meng, Rameswar Panda, Chung-Ching Lin, Prasanna Sattigeri, Leonid Karlinsky, Kate Saenko, Aude Oliva, Rogerio Feris</i></p> | 18399 |
| <p>Generating Furry Cars: Disentangling Object Shape and Appearance Across Multiple Domains</p> <p><i>Utkarsh Ojha, Krishna Kumar Singh, Yong Jae Lee</i></p> | 18412 |
| <p>Pruning Neural Networks at Initialization: Why Are We Missing the Mark?</p> <p><i>Jonathan Frankle, Gintare Karolina Dziugaite, Daniel Roy, Michael Carbin</i></p> | 18431 |
| <p>BRECQ: Pushing the Limit of Post-Training Quantization by Block Reconstruction</p> <p><i>Yuhang Li, Ruihao Gong, Xu Tan, Yang Yang, Peng Hu, Qi Zhang, Fengwei Yu, Wei Wang, Shi Gu</i></p> | 18472 |