

# **International Conference on Learning Representations (ICLR 2019)**

New Orleans, Louisiana, USA  
6 – 9 May 2019

Volume 1 of 12

ISBN: 978-1-7138-7273-3

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2019) by International Conference on Learning Representations  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact International Conference on Learning Representations  
at the address below.

International Conference on Learning Representations  
2710 E Corridor Drive  
Appleton, WI 54913

[www.iclr.cc](http://www.iclr.cc)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# TABLE OF CONTENTS

## VOLUME 1

### POSTER PRESENTATIONS

REGULARIZED LEARNING FOR DOMAIN ADAPTATION UNDER LABEL SHIFTS.....	1
<i>Kamyar Azizzadenesheli, Anqi Liu, Fanny Yang, Animashree Anandkumar</i>	
TOWARDS ROBUST, LOCALLY LINEAR DEEP NETWORKS .....	26
<i>Guang-He Lee, David Alvarez-Melis, Tommi S. Jaakkola</i>	
THE LIMITATIONS OF ADVERSARIAL TRAINING AND THE BLIND-SPOT ATTACK .....	47
<i>Huan Zhang, Hongge Chen, Zhao Song, Duane Boning, Inderjit S. Dhillon, Cho-Jui Hsieh</i>	
LEARNING TO LEARN WITHOUT FORGETTING BY MAXIMIZING TRANSFER AND MINIMIZING INTERFERENCE .....	63
<i>Matthew Riemer, Ignacio Cases, Robert Ajemian, Miao Liu, Irina Rish, Yuhai Tu, and Gerald Tesauro</i>	
GLOBAL-TO-LOCAL MEMORY POINTER NETWORKS FOR TASK-ORIENTED DIALOGUE .....	94
<i>Chien-Sheng Wu, Richard Socher, Caiming Xiong</i>	
RETHINKING THE VALUE OF NETWORK PRUNING .....	113
<i>Zhuang Liu, Mingjie Sun, Tinghui Zhou, Gao Huang, Trevor Darrell</i>	
NEURAL TTS STYLIZATION WITH ADVERSARIAL AND COLLABORATIVE GAMES .....	134
<i>Shuang Ma, Daniel Mcduff, Yale Song</i>	
ON THE UNIVERSAL APPROXIMABILITY AND COMPLEXITY BOUNDS OF QUANTIZED RELU NEURAL NETWORKS .....	150
<i>Yukun Ding, Jinglan Liu, Jinjun Xiong, Yiyu Shi</i>	
POINCARÉ GLOVE: HYPERBOLIC WORD EMBEDDINGS .....	169
<i>Alexandru Tifrea, Gary Becigneul, Octavian-Eugen Ganea,</i>	
EIDETIC 3D LSTM: A MODEL FOR VIDEO PREDICTION AND BEYOND .....	193
<i>Yunbo Wang, Lu Jiang, Ming-Hsuan Yang, Li-Jia Li, Mingsheng Long, Li Fei-Fei</i>	
TOWARDS GAN BENCHMARKS WHICH REQUIRE GENERALIZATION .....	207
<i>Ishaan Gulrajani, Colin Raffel, Luke Metz</i>	
THERE ARE MANY CONSISTENT EXPLANATIONS OF UNLABELED DATA: WHY YOU SHOULD AVERAGE .....	221
<i>Ben Athiwaratkun, Marc Finzi, Pavel Izmailov, Andrew Gordon Wilson</i>	
SYNTHETIC DATASETS FOR NEURAL PROGRAM SYNTHESIS.....	243
<i>Richard Shin, Neel Kant, Kavi Gupta, Chris Bender, Brandon Trabucco, Rishabh Singh, Dawn Song</i>	
STOCHASTIC PREDICTION OF MULTI-AGENT INTERACTIONS FROM PARTIAL OBSERVATIONS .....	259
<i>Chen Sun, Per Karlsson, Jiajun Wu, Joshua B Tenenbaum, Kevin Murphy</i>	
DYREP: LEARNING REPRESENTATIONS OVER DYNAMIC GRAPHS .....	274
<i>Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, Hongyuan Zha</i>	

LABEL SUPER-RESOLUTION NETWORKS.....	299
<i>Kolya Malkin, Caleb Robinson, Le Hou, Rachel Soobitsky, Jacob Czawlytko, Dimitris Samaras, Joel Saltz, Lucas Joppa, Nebojsa Jojic</i>	
MULTI-STEP RETRIEVER-READER INTERACTION FOR SCALABLE OPEN-DOMAIN QUESTION ANSWERING .....	321
<i>Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Andrew McCallum</i>	
CAPSULE GRAPH NEURAL NETWORK .....	334
<i>Zhang Xinyi, Lihui Chen</i>	
LEARNING A META-SOLVER FOR SYNTAX-GUIDED PROGRAM SYNTHESIS.....	350
<i>Xujie Si, Yuan Yang, Hanjun Dai, Mayur Naik, Le Song</i>	
STOCHASTIC OPTIMIZATION OF SORTING NETWORKS VIA CONTINUOUS RELAXATIONS .....	361
<i>Aditya Grover, Eric Wang, Aaron Zweig, Stefano Ermon</i>	
ENERGY-CONSTRAINED COMPRESSION FOR DEEP NEURAL NETWORKS VIA WEIGHTED SPARSE PROJECTION AND LAYER INPUT MASKING.....	384
<i>Haichuan Yang, Yuhao Zhu, Ji Liu</i>	
COMPOSING COMPLEX SKILLS BY LEARNING TRANSITION POLICIES .....	404
<i>Youngwoon Lee, Shao-Hua Sun, Sriram Somasundaram, Edward S. Hu, Joseph J. Lim</i>	
REVEALING INTERPRETABLE OBJECT REPRESENTATIONS FROM HUMAN BEHAVIOR.....	423
<i>Charles Y. Zheng, Francisco Pereira, Chris I. Baker, Martin N. Hebart</i>	
PROXYLESSNAS: DIRECT NEURAL ARCHITECTURE SEARCH ON TARGET TASK AND HARDWARE .....	439
<i>Han Cai, Ligeng Zhu, Song Han</i>	
A GENERATIVE MODEL FOR ELECTRON PATHS .....	452
<i>John Bradshaw, Matt J. Kusner, Brooks Paige, Marwin H. S. Segler, José Miguel Hernández-Lobato</i>	
LEARNING TO INFER AND EXECUTE 3D SHAPE PROGRAMS .....	471
<i>Yonglong Tian, Andrew Luo, Xingyuan Sun, Kevin Ellis, William T. Freeman, Joshua B. Tenenbaum, Jiajun Wu</i>	
MUSIC TRANSFORMER: GENERATING MUSIC WITH LONG-TERM STRUCTURE .....	492
<i>Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Ian Simon, Curtis Hawthorne, Noam Shazeer, Andrew M. Dai, Matthew D. Hoffman, Monica Dinulescu, Douglas Eck</i>	
MODELING THE LONG TERM FUTURE IN MODEL-BASED REINFORCEMENT LEARNING.....	507
<i>Nan Rosemary Ke, Amanpreet Singh, Ahmed Touati, Anirudh Goyal, Yoshua Bengio, Devi Parikh, Dhruv Batra</i>	
MODEL-PREDICTIVE POLICY LEARNING WITH UNCERTAINTY REGULARIZATION FOR DRIVING IN DENSE TRAFFIC .....	521
<i>Mikael Henaff, Alfredo Canziani, Yann LeCun</i>	
LEARNING TO SCREEN FOR FAST SOFTMAX INFERENCE ON LARGE VOCABULARY NEURAL NETWORKS.....	541
<i>Patrick Chen, Si Si, Sanjiv Kumar, Yang Li, Cho-Jui Hsieh</i>	
INTERPOLATION-PREDICTION NETWORKS FOR IRREGULARLY SAMPLED TIME SERIES.....	557
<i>Satya Narayan Shukla, Benjamin Marlin</i>	

CONTINGENCY-AWARE EXPLORATION IN REINFORCEMENT LEARNING .....	571
<i>Jongwook Choi, Yijie Guo, Marcin Moczulski, Junhyuk Oh, Neal Wu, Mohammad Norouzi, Honglak Lee</i>	
NEURAL GRAPH EVOLUTION: TOWARDS EFFICIENT AUTOMATIC ROBOT DESIGN.....	590
<i>Tingwu Wang, Yuhao Zhou, Sanja Fidler, Jimmy Ba</i>	
SELFLESS SEQUENTIAL LEARNING .....	607
<i>Rahaf Aljundi, Marcus Rohrbach, Tinne Tuytelaars</i>	
LEARNING PARTICLE DYNAMICS FOR MANIPULATING RIGID BODIES, DEFORMABLE OBJECTS, AND FLUIDS.....	624
<i>Yunzhu Li, Jiajun Wu, Russ Tedrake, Joshua B. Tenenbaum, Antonio Torralba</i>	
REPRESENTING FORMAL LANGUAGES: A COMPARISON BETWEEN FINITE AUTOMATA AND RECURRENT NEURAL NETWORKS .....	639
<i>Joshua J. Michalek, Ameesh Shah, Abhinav Verma, Richard G. Baraniuk, Swarat Chaudhuri, Ankit B. Patel</i>	
DISJOINT MAPPING NETWORK FOR CROSS-MODAL MATCHING OF VOICES AND FACES.....	654
<i>Yandong Wen, Mahmoud Al Ismail, Weiyang Liu, Bhiksha Raj, Rita Singh</i>	
LEARNING PROCEDURAL ABSTRACTIONS AND EVALUATING DISCRETE LATENT TEMPORAL STRUCTURE .....	671
<i>Karan Goel, Emma Brunskill</i>	
LEARNING TO DESIGN RNA .....	689
<i>Frederic Runge, Danny Stoll, Stefan Falkner, Frank Hutter</i>	
COST-SENSITIVE ROBUSTNESS AGAINST ADVERSARIAL EXAMPLES.....	718
<i>Xiao Zhang, David Evans</i>	
COMBINATORIAL ATTACKS ON BINARIZED NEURAL NETWORKS.....	732
<i>Elias B Khalil, Amrita Gupta, Bistra Dilkina</i>	
A VARIATIONAL INEQUALITY PERSPECTIVE ON GENERATIVE ADVERSARIAL NETWORKS.....	745
<i>Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, Simon Lacoste-Julien</i>	
MULTIPLE-ATTRIBUTE TEXT REWRITING .....	783
<i>Guillaume Lample, Sandeep Subramanian, Eric Smith, Ludovic Denoyer, Marc'Aurelio Ranzato, Y-Lan Boureau</i>	

## VOLUME 2

MULTI-AGENT DUAL LEARNING.....	803
<i>Yiren Wang, Yingce Xia, Tianyu He, Fei Tian, Tao Qin, ChengXiang Zhai, Tie-Yan Liu</i>	
LEARNING SPARSE RELATIONAL TRANSITION MODELS.....	818
<i>Victoria Xia, Zi Wang, Kelsey Allen, Tom Silver, Leslie Pack Kaelbling</i>	
SPIGAN: PRIVILEGED ADVERSARIAL LEARNING FROM SIMULATION.....	832
<i>Kuan-Hui Lee, German Ros, Jie Li, Adrien Gaidon</i>	
UNIVERSAL STAGewise LEARNING FOR NON-CONVEX PROBLEMS WITH CONVERGENCE ON AVERAGED SOLUTIONS .....	846
<i>Zaiyi Chen, Zhuoning Yuan, Jinfeng Yi, Bowen Zhou, Enhong Chen, Tianbao Yang</i>	

REASONING ABOUT PHYSICAL INTERACTIONS WITH OBJECT-ORIENTED PREDICTION AND PLANNING.....	869
<i>Michael Janner, Sergey Levine, William T. Freeman, Joshua B. Tenenbaum, Chelsea Finn, Jiajun Wu</i>	
POSTERIOR ATTENTION MODELS FOR SEQUENCE TO SEQUENCE LEARNING.....	881
<i>Shiv Shankar, Sunita Sarawagi</i>	
NOODL: PROVABLE ONLINE DICTIONARY LEARNING AND SPARSE CODING .....	892
<i>Sirisha Rambhatla, Xingguo Li, Jarvis Haupt</i>	
RELGAN: RELATIONAL GENERATIVE ADVERSARIAL NETWORKS FOR TEXT GENERATION.....	934
<i>Weili Nie, Nina Narodytska, Ankit Patel</i>	
DO DEEP GENERATIVE MODELS KNOW WHAT THEY DON'T KNOW?.....	954
<i>Eric Nalisnick, Akihiro Matsukawa, Yee Whye Teh, Dilan Gorur, Balaji Lakshminarayanan</i>	
K FOR THE PRICE OF 1: PARAMETER-EFFICIENT MULTI-TASK AND TRANSFER LEARNING.....	973
<i>Pramod Kaushik Mudrakarta, Mark Sandler, Andrey Zhmoginov, Andrew Howard</i>	
MISGAN: LEARNING FROM INCOMPLETE DATA WITH GENERATIVE ADVERSARIAL NETWORKS.....	988
<i>Steven Cheng-Xian Li, Bo Jiang, Benjamin Marlin</i>	
LEARNABLE EMBEDDING SPACE FOR EFFICIENT NEURAL ARCHITECTURE COMPRESSION .....	1008
<i>Shengcao Cao, Xiaofang Wang, Kris M. Kitani</i>	
GUIDING POLICIES WITH LANGUAGE VIA META-LEARNING.....	1025
<i>John D. Co-Reyes, Abhishek Gupta, Suvansh Sanjeev, Nick Altieri, Jacob Andreas, John DeNero, Pieter Abbeel, Sergey Levine</i>	
ACTIVE LEARNING WITH PARTIAL FEEDBACK.....	1042
<i>Peiyun Hu, Zachary C. Lipton, Anima Anandkumar, Deva Ramanan</i>	
ON THE SENSITIVITY OF ADVERSARIAL ROBUSTNESS TO INPUT DATA DISTRIBUTIONS.....	1056
<i>Gavin Weiguang Ding, Kry Yik Chau Lui, Xiaomeng Jin, Luyu Wang, Ruitong Huang</i>	
EFFICIENT MULTI-OBJECTIVE NEURAL ARCHITECTURE SEARCH VIA LAMARCKIAN EVOLUTION.....	1075
<i>Thomas Elsken, Jan Hendrik Metzen, Frank Hutter</i>	
DISTRIBUTIONAL CONCAVITY REGULARIZATION FOR GANS .....	1098
<i>Shoichiro Yamaguchi, Masanori Koyama</i>	
CHARACTERIZING AUDIO ADVERSARIAL EXAMPLES USING TEMPORAL DEPENDENCY .....	1121
<i>Zhuolin Yang, Bo Li, Pin-Yu Chen, Dawn Song</i>	
GANSYNTH: ADVERSARIAL NEURAL AUDIO SYNTHESIS .....	1136
<i>Jesse Engel, Kumar Krishna Agrawal, Shuo Chen, Ishaan Gulrajani, Chris Donahue, Adam Roberts</i>	
TOWARD UNDERSTANDING THE IMPACT OF STALENESS IN DISTRIBUTED MACHINE LEARNING.....	1153
<i>Wei Dai, Yi Zhou, Nanqing Dong, Hao Zhang, Eric Xing</i>	
BEYOND GREEDY RANKING: SLATE OPTIMIZATION VIA LIST-CVAE.....	1173
<i>Ray Jiang, Sven Gowal, Yuqiu Qian, Timothy Mann, Danilo J. Rezende</i>	

G-SGD: OPTIMIZING RELU NEURAL NETWORKS IN ITS POSITIVELY SCALE-INVARIANT SPACE .....	1185
<i>Qi Meng, Shuxin Zheng, Huishuai Zhang, Wei Chen, Qiwei Ye, Zhi-Ming Ma, Nenghai Yu, Tie-Yan Liu</i>	
SPHERICAL CNNs ON UNSTRUCTURED GRIDS .....	1204
<i>Chiyu Max Jiang, Jingwei Huang, Karthik Kashinath, Prabhat, Philip Marcus, Matthias Niessner</i>	
BOOSTING ROBUSTNESS CERTIFICATION OF NEURAL NETWORKS .....	1220
<i>Gagandeep Singh, Timon Gehr, Markus Püschel, Martin Vechev</i>	
TWO-TIMESCALE NETWORKS FOR NONLINEAR VALUE FUNCTION APPROXIMATION .....	1232
<i>Wesley Chung, Somjit Nath, Ajin Joseph, Martha White</i>	
DIFFERENTIABLE PERTURB-AND-PARSE: SEMI-SUPERVISED PARSING WITH A STRUCTURED VARIATIONAL AUTOENCODER .....	1264
<i>Caio Corro, Ivan Titov</i>	
ALGORITHMIC FRAMEWORK FOR MODEL-BASED DEEP REINFORCEMENT LEARNING WITH THEORETICAL GUARANTEES.....	1288
<i>Yuping Luo, Huazhe Xu, Yuanzhi Li, Yuandong Tian, Trevor Darrell, Tengyu Ma</i>	
TREE-STRUCTURED RECURRENT SWITCHING LINEAR DYNAMICAL SYSTEMS FOR MULTI-SCALE MODELING.....	1315
<i>Josue Nassar, Scott Linderman, Monica Bugallo, Il Memming Park</i>	
DIAGNOSING AND ENHANCING VAE MODELS .....	1332
<i>Bin Dai, David Wipf</i>	
EFFICIENTLY TESTING LOCAL OPTIMALITY AND ESCAPING SADDLES FOR RELU NETWORKS.....	1344
<i>Chulhee Yun, Suvrit Sra, Ali Jadbabaie</i>	
DON'T LET YOUR DISCRIMINATOR BE FOOLED .....	1367
<i>Brady Zhou, Philipp Krähenbühl</i>	
RIGOROUS AGENT EVALUATION: AN ADVERSARIAL APPROACH TO UNCOVER CATASTROPHIC FAILURES .....	1384
<i>Jonathan Uesato, Ananya Kumar, Csaba Szepesvari, Tom Erez, Avraham Ruderman, Keith Anderson, Krishnamurthy (Dj) Dvijotham, Nicolas Heess, Pushmeet Kohli</i>	
COMPETITIVE EXPERIENCE REPLAY .....	1404
<i>Hao Liu, Alexander Trott, Richard Socher, Caiming Xiong</i>	
A MAX-AFFINE SPLINE PERSPECTIVE OF RECURRENT NEURAL NETWORKS .....	1420
<i>Zichao Wang, Randall Balestriero, Richard Baraniuk</i>	
TOP-DOWN NEURAL MODEL FOR FORMULAE .....	1440
<i>Karel Chvalovský</i>	
FEATURE-WISE BIAS AMPLIFICATION .....	1449
<i>Klas Leino, Emily Black, Matt Fredrikson, Shayak Sen, Anupam Datta</i>	
AD-VAT: AN ASYMMETRIC DUELING MECHANISM FOR LEARNING VISUAL ACTIVE TRACKING .....	1462
<i>Fangwei Zhong, Peng Sun, Wenhan Luo, Tingyun Yan, Yizhou Wang</i>	
LEARNING TO LEARN WITH CONDITIONAL CLASS DEPENDENCIES.....	1478
<i>Xiang Jiang, Mohammad Havaei, Farshid Varno, Gabriel Chartrand, Nicolas Chapados, Stan Matwin</i>	

GAN DISSECTION: VISUALIZING AND UNDERSTANDING GENERATIVE ADVERSARIAL NETWORKS.....	1489
<i>David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, Antonio Torralba</i>	

TIMBRETRON: A WAVENET(CYCLEGAN(CQT(AUDIO))) PIPELINE FOR MUSICAL TIMBRE TRANSFER.....	1508
<i>Sicong Huang, Qiyang Li, Cem Anil, Xuchan Bao, Sageev Oore, Roger B. Grosse</i>	

A MEAN FIELD THEORY OF BATCH NORMALIZATION.....	1525
<i>Greg Yang, Jeffrey Pennington, Vinay Rao, Jascha Sohl-Dickstein, Samuel S. Schoenholz</i>	

### VOLUME 3

A CLOSER LOOK AT FEW-SHOT CLASSIFICATION.....	1620
<i>Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, Jia-Bin Huang</i>	

STCN: STOCHASTIC TEMPORAL CONVOLUTIONAL NETWORKS.....	1636
<i>Emre Aksan, Otmar Hilliges</i>	

ROTATE: KNOWLEDGE GRAPH EMBEDDING BY RELATIONAL ROTATION IN COMPLEX SPACE.....	1649
<i>Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, Jian Tang</i>	

LEARNING TO NAVIGATE THE WEB.....	1667
<i>Izzeddin Gur, Ulrich Rueckert, Aleksandra Faust, Dilek Hakkani-Tur</i>	

MODELING UNCERTAINTY WITH HEDGED INSTANCE EMBEDDINGS.....	1679
<i>Seong Joon Oh, Kevin P. Murphy, Jiyan Pan, Joseph Roth, Florian Schroff, Andrew C. Gallagher</i>	

AUTOMATICALLY COMPOSING REPRESENTATION TRANSFORMATIONS AS A MEANS FOR GENERALIZATION.....	1695
<i>Michael Chang, Abhishek Gupta, Sergey Levine, Thomas L. Griffiths</i>	

SYSTEMATIC GENERALIZATION: WHAT IS REQUIRED AND CAN IT BE LEARNED?.....	1718
<i>Dzmitry Bahdanau, Shikhar Murty, Michael Noukhovitch, Thien Huu Nguyen, Harm de Vries, Aaron Courville</i>	

ADAPTIVE INPUT REPRESENTATIONS FOR NEURAL LANGUAGE MODELING.....	1734
<i>Alexei Baevski, Michael Auli</i>	

OPTIMAL CONTROL VIA NEURAL NETWORKS: A CONVEX APPROACH.....	1747
<i>Yize Chen, Yuanyuan Shi, Baosen Zhang</i>	

AN EMPIRICAL STUDY OF EXAMPLE FORGETTING DURING DEEP NEURAL NETWORK LEARNING.....	1768
<i>Mariya Toneva, Alessandro Sordani, Remi Tachet des Combes, Adam Trischler, Yoshua Bengio, Geoffrey J. Gordon</i>	

GLUE: A MULTI-TASK BENCHMARK AND ANALYSIS PLATFORM FOR NATURAL LANGUAGE UNDERSTANDING.....	1786
<i>Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, Samuel R. Bowman</i>	

PLAN ONLINE, LEARN OFFLINE: EFFICIENT LEARNING AND EXPLORATION VIA MODEL-BASED CONTROL.....	1806
<i>Kendall Lowrey, Aravind Rajeswaran, Sham Kakade, Emanuel Todorov, Igor Mordatch</i>	

LEARNING GRID CELLS AS VECTOR REPRESENTATION OF SELF-POSITION COUPLED WITH MATRIX REPRESENTATION OF SELF-MOTION.....	1821
<i>Ruiqi Gao, Jianwen Xie, Song-Chun Zhu, Ying Nian Wu</i>	



PREVENTING POSTERIOR COLLAPSE WITH DELTA-VAES .....	1845
<i>Ali Razavi, Aaron van den Oord, Ben Poole, Oriol Vinyals</i>	
DEEP ONLINE LEARNING VIA META-LEARNING: CONTINUAL ADAPTATION FOR MODEL-BASED RL .....	1869
<i>Anusha Nagabandi, Chelsea Finn, Sergey Levine</i>	
VALUE PROPAGATION NETWORKS.....	1884
<i>Nantas Nardelli, Gabriel Synnaeve, Zeming Lin, Pushmeet Kohli, Philip H. S. Torr, Nicolas Usunier</i>	
EFFICIENT AUGMENTATION VIA DATA SUBSAMPLING .....	1897
<i>Michael Kuchnik, Virginia Smith</i>	
ALISTA: ANALYTIC WEIGHTS ARE AS GOOD AS LEARNED WEIGHTS IN LISTA .....	1919
<i>Jialin Liu, Xiaohan Chen, Zhangyang Wang, Wotao Yin</i>	
RECALL TRACES: BACKTRACKING MODELS FOR EFFICIENT REINFORCEMENT LEARNING.....	1952
<i>Anirudh Goyal, Philemon Brakel, William Fedus, Soumye Singhal, Timothy Lillicrap, Sergey Levine, Hugo Larochelle, Yoshua Bengio</i>	
FIXUP INITIALIZATION: RESIDUAL LEARNING WITHOUT NORMALIZATION .....	1972
<i>Hongyi Zhang, Yann N. Dauphin, Tengyu Ma</i>	
DIVERSITY-SENSITIVE CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS .....	1988
<i>Dingdong Yang, Seunghoon Hong, Yunseok Jang, Tianchen Zhao, Honglak Lee</i>	
INFORMATION ASYMMETRY IN KL-REGULARIZED RL .....	2011
<i>Alexandre Galashov, Siddhant M. Jayakumar, Leonard Hasenclever, Dhruva Tirumala, Jonathan Schwarz, Guillaume Desjardins, Wojciech M. Czarnecki, Yee Whye Teh, Razvan Pascanu, Nicolas Heess</i>	
FLOWQA: GRASPING FLOW IN HISTORY FOR CONVERSATIONAL MACHINE COMPREHENSION .....	2036
<i>Hsin-Yuan Huang, Eunsol Choi, Wen-tau Yih</i>	
PROBABILISTIC PLANNING WITH SEQUENTIAL MONTE CARLO METHODS .....	2051
<i>Alexandre Piche, Valentin Thomas, Cyril Ibrahim, Yoshua Bengio, Chris Pal</i>	
SPREADING VECTORS FOR SIMILARITY SEARCH .....	2068
<i>Alexandre Sablayrolles, Matthijs Douze, Cordelia Schmid, Hervé Jégou</i>	
LANCZOSNET: MULTI-SCALE DEEP GRAPH CONVOLUTIONAL NETWORKS.....	2081
<i>Renjie Liao, Zhizhen Zhao, Raquel Urtasun, Richard Zemel</i>	
LEARNING WHAT YOU CAN DO BEFORE DOING ANYTHING .....	2098
<i>Oleh Rybkin, Karl Pertsch, Konstantinos G. Derpanis, Kostas Daniilidis, Andrew Jaegle</i>	
LAGGING INFERENCE NETWORKS AND POSTERIOR COLLAPSE IN VARIATIONAL AUTOENCODERS.....	2123
<i>Junxian He, Daniel Spokoyny, Graham Neubig, Taylor Berg-Kirkpatrick</i>	
PREFERENCES IMPLICIT IN THE STATE OF THE WORLD .....	2138
<i>Rohin Shah, Dmitrii Krasheninnikov, Jordan Alexander, Pieter Abbeel, Anca Dragan</i>	
A DIRECT APPROACH TO ROBUST DEEP LEARNING USING ADVERSARIAL NETWORKS.....	2154
<i>Huaxia Wang, Chun-Nam Yu</i>	
IMPROVING THE GENERALIZATION OF ADVERSARIAL TRAINING WITH DOMAIN ADAPTATION .....	2169
<i>Chuanbiao Song, Kun He, Liwei Wang, John E. Hopcroft</i>	

SUBGRADIENT DESCENT LEARNS ORTHOGONAL DICTIONARIES .....	2183
<i>Yu Bai, Qijia Jiang, Ju Sun</i>	
IMPROVING DIFFERENTIABLE NEURAL COMPUTERS THROUGH MEMORY MASKING, DE-ALLOCATION, AND LINK DISTRIBUTION SHARPNESS CONTROL .....	2220
<i>Robert Csordas, Juergen Schmidhuber</i>	
UNSUPERVISED CONTROL THROUGH NON-PARAMETRIC DISCRIMINATIVE REWARDS .....	2233
<i>David Warde-Farley, Tom Van de Wiele, Tejas Kulkarni, Catalin Ionescu, Steven Hansen, Volodymyr Mnih</i>	
SELF-TUNING NETWORKS: BILEVEL OPTIMIZATION OF HYPERPARAMETERS USING STRUCTURED BEST-RESPONSE FUNCTIONS .....	2250
<i>Matthew Mackay, Paul Vicol, Jonathan Lorraine, David Duvenaud, Roger Grosse</i>	
EXPLAINING IMAGE CLASSIFIERS BY COUNTERFACTUAL GENERATION .....	2275
<i>Chun-Hao Chang, Elliot Creager, Anna Goldenberg, David Duvenaud</i>	
PREDICTING THE GENERALIZATION GAP IN DEEP NETWORKS WITH MARGIN DISTRIBUTIONS .....	2294
<i>Yiding Jiang, Dilip Krishnan, Hossein Mobahi, Samy Bengio</i>	
MODE NORMALIZATION .....	2313
<i>Lucas Deecke, Iain Murray, Hakan Bilen</i>	
KERNEL CHANGE-POINT DETECTION WITH AUXILIARY DEEP GENERATIVE MODELS .....	2325
<i>Wei-Cheng Chang, Chun-Liang Li, Yiming Yang, Barnabás Póczos</i>	
TOWARDS METAMERISM VIA FOVEATED STYLE TRANSFER .....	2339
<i>Arturo Deza, Aditya Jonnalagadda, Miguel P. Eckstein</i>	

#### VOLUME 4

LEARNING TO SOLVE CIRCUIT-SAT: AN UNSUPERVISED DIFFERENTIABLE APPROACH .....	2360
<i>Saeed Amizadeh, Sergiy Matusevych, Markus Weimer</i>	
VARIANCE REDUCTION FOR REINFORCEMENT LEARNING IN INPUT-DRIVEN ENVIRONMENTS .....	2374
<i>Hongzi Mao, Shaileshh Bojja Venkatakrishnan, Malte Schwarzkopf, Mohammad Alizadeh</i>	
ROBUSTNESS MAY BE AT ODDS WITH ACCURACY .....	2394
<i>Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, Aleksander Madry</i>	
SUPERVISED COMMUNITY DETECTION WITH LINE GRAPH NEURAL NETWORKS .....	2417
<i>Zhengdao Chen, Lisha Li, Joan Bruna</i>	
HARMONIC UNPAIRED IMAGE-TO-IMAGE TRANSLATION .....	2440
<i>Rui Zhang, Tomas Pfister, Jia Li</i>	
LEARNING NEURAL PDE SOLVERS WITH CONVERGENCE GUARANTEES .....	2459
<i>Jun-Ting Hsieh, Shengjia Zhao, Stephan Eismann, Lucia Mirabella, Stefano Ermon</i>	
META-LEARNING WITH DIFFERENTIABLE CLOSED-FORM SOLVERS .....	2473
<i>Luca Bertinetto, Joao F. Henriques, Philip Torr, Andrea Vedaldi</i>	
BUILDING DYNAMIC KNOWLEDGE GRAPHS FROM TEXT USING MACHINE READING COMPREHENSION .....	2488
<i>Rajarshi Das, Tsendsuren Munkhdalai, Xingdi Yuan, Adam Trischler, Andrew McCallum</i>	
PRIOR CONVICTIONS: BLACK-BOX ADVERSARIAL ATTACKS WITH BANDITS AND PRIORS .....	2500
<i>Andrew Ilyas, Logan Engstrom, Aleksander Madry</i>	

DHER: HINDSIGHT EXPERIENCE REPLAY FOR DYNAMIC GOALS .....	2525
<i>Meng Fang, Cheng Zhou, Bei Shi, Boqing Gong, Jia Xu, Tong Zhang</i>	
STOCHASTIC GRADIENT/MIRROR DESCENT: MINIMAX OPTIMALITY AND IMPLICIT REGULARIZATION .....	2537
<i>Navid Azizan, Babak Hassibi</i>	
UNSUPERVISED LEARNING OF THE SET OF LOCAL MAXIMA .....	2555
<i>Lior Wolf, Sagie Benaim, Tomer Galanti</i>	
NEURAL LOGIC MACHINES .....	2574
<i>Honghua Dong, Jiayuan Mao, Tian Lin, Chong Wang, Lihong Li, Denny Zhou</i>	
DEFENSIVE QUANTIZATION: WHEN EFFICIENCY MEETS ROBUSTNESS.....	2596
<i>Ji Lin, Chuang Gan, Song Han</i>	
DIMENSIONALITY REDUCTION FOR REPRESENTING THE KNOWLEDGE OF PROBABILISTIC MODELS .....	2610
<i>Marc T Law, Jake Snell, Amir-massoud Farahmand, Raquel Urtasun, Richard S Zemel</i>	
BIOLOGICALLY-PLAUSIBLE LEARNING ALGORITHMS CAN SCALE TO LARGE DATASETS .....	2644
<i>Will Xiao, Honglin Chen, Qianli Liao, Tomaso Poggio</i>	
GENERATING MULTI-AGENT TRAJECTORIES USING PROGRAMMATIC WEAK SUPERVISION .....	2653
<i>Eric Zhan, Stephan Zheng, Yisong Yue, Long Sha, Patrick Lucey</i>	
PREDICT THEN PROPAGATE: GRAPH NEURAL NETWORKS MEET PERSONALIZED PAGERANK.....	2667
<i>Johannes Gasteiger, Aleksandar Bojchevski, Stephan Günnemann</i>	
ENVIRONMENT PROBING INTERACTION POLICIES .....	2682
<i>Wenxuan Zhou, Lerrel Pinto, Abhinav Gupta</i>	
TIME-AGNOSTIC PREDICTION: PREDICTING PREDICTABLE VIDEO FRAMES .....	2695
<i>Dinesh Jayaraman, Frederik Ebert, Alexei Efros, Sergey Levine</i>	
UNSUPERVISED ADVERSARIAL IMAGE RECONSTRUCTION .....	2715
<i>Arthur Pajot, Emmanuel de Bezenac, Patrick Gallinari</i>	
DEEP DECODER: CONCISE IMAGE REPRESENTATIONS FROM UNTRAINED NON- CONVOLUTIONAL NETWORKS.....	2736
<i>Reinhard Heckel, Paul Hand</i>	
MINIMAL IMAGES IN DEEP NEURAL NETWORKS: FRAGILE OBJECT RECOGNITION IN NATURAL IMAGES .....	2750
<i>Sanjana Srivastava, Guy Ben-Yosef, Xavier Boix</i>	
OVERCOMING THE DISENTANGLEMENT VS RECONSTRUCTION TRADE-OFF VIA JACOBIAN SUPERVISION.....	2768
<i>José Lezama</i>	
PROBGAN: TOWARDS PROBABILISTIC GAN WITH THEORETICAL GUARANTEES.....	2786
<i>Hao He, Hao Wang, Guang-He Lee, Yonglong Tian</i>	
THEORETICAL ANALYSIS OF AUTO RATE-TUNING BY BATCH NORMALIZATION .....	2814
<i>Sanjeev Arora, Zhiyuan Li, Kaifeng Lyu</i>	
THREE MECHANISMS OF WEIGHT DECAY REGULARIZATION .....	2836
<i>Guodong Zhang, Chaoqi Wang, Bowen Xu, Roger Grosse</i>	

APPROXIMATING CNNs WITH BAG-OF-LOCAL-FEATURES MODELS WORKS SURPRISINGLY WELL ON IMAGENET .....	2853
<i>Wieland Brendel, Matthias Bethge</i>	
LEARNING EXPLORATION POLICIES FOR NAVIGATION .....	2868
<i>Tao Chen, Saurabh Gupta, Abhinav Gupta</i>	
LEARNING PROGRAMMATICALLY STRUCTURED REPRESENTATIONS WITH PERCEPTOR GRADIENTS .....	2882
<i>Svetlin Penkov, Subramanian Ramamoorthy</i>	
LEARNING MIXED-CURVATURE REPRESENTATIONS IN PRODUCT SPACES .....	2898
<i>Albert Gu, Frederic Sala, Beliz Gunel, Christopher Ré</i>	
STABLE RECURRENT MODELS .....	2919
<i>John Miller, Moritz Hardt</i>	
DEEP ANOMALY DETECTION WITH OUTLIER EXPOSURE.....	2942
<i>Dan Hendrycks, Mantas Mazeika, Thomas Dietterich</i>	
EXEMPLAR GUIDED UNSUPERVISED IMAGE-TO-IMAGE TRANSLATION WITH SEMANTIC CONSISTENCY .....	2960
<i>Liqian Ma, Xu Jia, Stamatios Georgoulis, Tinne Tuytelaars, Luc Van Gool</i>	
DOUBLY REPARAMETERIZED GRADIENT ESTIMATORS FOR MONTE CARLO OBJECTIVES .....	2977
<i>George Tucker, Dieterich Lawson, Shixiang Gu, Chris J. Maddison</i>	
ADVERSARIAL ATTACKS ON GRAPH NEURAL NETWORKS VIA META LEARNING.....	2991
<i>Daniel Zügner, Stephan Günnemann</i>	
ADAPTIVE GRADIENT METHODS WITH DYNAMIC BOUND OF LEARNING RATE .....	3006
<i>Liangchen Luo, Yuanhao Xiong, Yan Liu, Xu Sun</i>	
RANDOM MESH PROJECTORS FOR INVERSE PROBLEMS .....	3027
<i>Konik Kothari, Sidharth Gupta, Maarten v. de Hoop, Ivan Dokmanic</i>	
A STATISTICAL APPROACH TO ASSESSING NEURAL NETWORK ROBUSTNESS.....	3045
<i>Stefan Webb, Tom Rainforth, Yee Whye Teh, M. Pawan Kumar</i>	
LEARNING ACTIONABLE REPRESENTATIONS WITH GOAL CONDITIONED POLICIES .....	3060
<i>Dibya Ghosh, Abhishek Gupta, Sergey Levine</i>	
LEARNING IMPLICITLY RECURRENT CNNs THROUGH PARAMETER SHARING.....	3078
<i>Pedro Savarese, Michael Maire</i>	
TRANSFER LEARNING FOR SEQUENCES VIA LEARNING TO COLLOCATE .....	3093
<i>Wanyun Cui, Guangyu Zheng, Zhiqiang Shen, Sihang Jiang, Wei Wang</i>	
HOW TO TRAIN YOUR MAML.....	3103
<i>Antreas Antoniou, Harrison Edwards, Amos Storkey</i>	
VARIATIONAL DISCRIMINATOR BOTTLENECK: IMPROVING IMITATION LEARNING, INVERSE RL, AND GANS BY CONSTRAINING INFORMATION FLOW .....	3114
<i>Xue Bin Peng, Angjoo Kanazawa, Sam Toyer, Pieter Abbeel, Sergey Levine</i>	

## VOLUME 5

LEARNING WHAT AND WHERE TO ATTEND .....	3141
<i>Drew Linsley, Dan Shiebler, Sven Eberhardt, Thomas Serre</i>	

LEARNING PROTEIN SEQUENCE EMBEDDINGS USING INFORMATION FROM STRUCTURE.....	3162
<i>Tristan Bepler, Bonnie Berger</i>	
WHAT DO YOU LEARN FROM CONTEXT? PROBING FOR SENTENCE STRUCTURE IN CONTEXTUALIZED WORD REPRESENTATIONS .....	3179
<i>Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R. Bowman, Dipanjan Das, Ellie Pavlick</i>	
DEEP LEARNING GENERALIZES BECAUSE THE PARAMETER-FUNCTION MAP IS BIASED TOWARDS SIMPLE FUNCTIONS .....	3196
<i>Guillermo Valle-Perez, Chico Q. Camargo, Ard A. Louis</i>	
CAMOU: LEARNING PHYSICAL VEHICLE CAMOUFLAGES TO ADVERSARIALLY ATTACK DETECTORS IN THE WILD.....	3231
<i>Yang Zhang, Hassan Foroosh, Philip David, Boqing Gong</i>	
MAE: MUTUAL POSTERIOR-DIVERGENCE REGULARIZATION FOR VARIATIONAL AUTOENCODERS .....	3251
<i>Xuezhe Ma, Chunting Zhou, Eduard Hovy</i>	
A CONVERGENCE ANALYSIS OF GRADIENT DESCENT FOR DEEP LINEAR NEURAL NETWORKS.....	3267
<i>Sanjeev Arora, Nadav Cohen, Noah Golowich, Wei Hu</i>	
DON'T SETTLE FOR AVERAGE, GO FOR THE MAX: FUZZY SETS AND MAX-POOLED WORD VECTORS.....	3300
<i>Vitalii Zhelezniak, Aleksandar Savkov, April Shen, Francesco Moramarco, Jack Flann, Nils Y. Hammerla</i>	
THE ROLE OF OVER-PARAMETRIZATION IN GENERALIZATION OF NEURAL NETWORKS.....	3317
<i>Behnam Neyshabur, Zhiyuan Li, Srinadh Bhojanapalli, Yann LeCun, Nathan Srebro</i>	
ON THE CONVERGENCE OF A CLASS OF ADAM-TYPE ALGORITHMS FOR NON- CONVEX OPTIMIZATION .....	3337
<i>Xiangyi Chen, Sijia Liu, Ruoyu Sun, Mingyi Hong</i>	
GRADIENT DESCENT ALIGNS THE LAYERS OF DEEP LINEAR NETWORKS .....	3367
<i>Ziwei Ji, Matus Telgarsky</i>	
HIERARCHICAL RL USING AN ENSEMBLE OF PROPRIOCEPTIVE PERIODIC POLICIES .....	3387
<i>Kenneth Marino, Abhinav Gupta, Rob Fergus, Arthur Szlam</i>	
LEARNING TO SIMULATE.....	3400
<i>Nataniel Ruiz, Samuel Schuster, Manmohan Chandraker</i>	
MULTILINGUAL NEURAL MACHINE TRANSLATION WITH SOFT DECOUPLED ENCODING .....	3412
<i>Xinyi Wang, Hieu Pham, Philip Arthur, Graham Neubig</i>	
META-LEARNING WITH LATENT EMBEDDING OPTIMIZATION .....	3425
<i>Andrei A. Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, Raia Hadsell</i>	
VISUAL SEMANTIC NAVIGATION USING SCENE PRIORS .....	3442
<i>Wei Yang, Xiaolong Wang, Ali Farhadi, Abhinav Gupta, Roozbeh Mottaghi</i>	
ADASHIFT: DECORRELATION AND CONVERGENCE OF ADAPTIVE LEARNING RATE METHODS.....	3455
<i>Zhiming Zhou, Qingru Zhang, Guansong Lu, Hongwei Wang, Weinan Zhang, Yong Yu</i>	

SLICED WASSERSTEIN AUTO-ENCODERS .....	3481
<i>Soheil Kolouri, Phillip E. Pope, Charles E. Martin, Gustavo K. Rohde</i>	
AMORTIZED BAYESIAN META-LEARNING .....	3500
<i>Sachin Ravi, Alex Beatson</i>	
LOCAL SGD CONVERGES FAST AND COMMUNICATES LITTLE .....	3514
<i>Sebastian U. Stich</i>	
TEXTTOVEC: DEEP CONTEXTUALIZED NEURAL AUTOREGRESSIVE TOPIC MODELS OF LANGUAGE WITH DISTRIBUTED COMPOSITIONAL PRIOR .....	3531
<i>Pankaj Gupta, Yatin Chaudhary, Florian Buettner, Hinrich Schuetze</i>	
NEURAL PROGRAM REPAIR BY JOINTLY LEARNING TO LOCALIZE AND REPAIR .....	3548
<i>Marko Vasic, Aditya Kanade, Petros Maniatis, David Bieber, Rishabh Singh</i>	
HUMAN-LEVEL PROTEIN LOCALIZATION WITH CONVOLUTIONAL NEURAL NETWORKS.....	3560
<i>Elisabeth Rumetshofer, Markus Hofmarcher, Clemens Röhl, Sepp Hochreiter, Günter Klambauer</i>	
FROM LANGUAGE TO GOALS: INVERSE REINFORCEMENT LEARNING FOR VISION- BASED INSTRUCTION FOLLOWING.....	3578
<i>Justin Fu, Anoop Korattikara, Sergey Levine, Sergio Guadarrama</i>	
CLARINET: PARALLEL WAVE GENERATION IN END-TO-END TEXT-TO-SPEECH.....	3592
<i>Wei Ping, Kainan Peng, Jitong Chen</i>	
VARIATIONAL AUTOENCODER WITH ARBITRARY CONDITIONING .....	3607
<i>Oleg Ivanov, Michael Figurnov, Dmitry Vetrov</i>	
JANOSSY POOLING: LEARNING DEEP PERMUTATION-INVARIANT FUNCTIONS FOR VARIABLE-SIZE INPUTS.....	3632
<i>Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak Rao, Bruno Ribeiro</i>	
THE DEEP WEIGHT PRIOR .....	3654
<i>Andrei Atanov, Arsenii Ashukha, Kirill Struminsky, Dmitriy Vetrov, Max Welling</i>	
A NEW DOG LEARNS OLD TRICKS: RL FINDS CLASSIC OPTIMIZATION ALGORITHMS .....	3671
<i>Weiwei Kong, Christopher Liaw, Aranyak Mehta, D. Sivakumar</i>	
DOM-Q-NET: GROUNDED RL ON STRUCTURED LANGUAGE .....	3696
<i>Sheng Jia, Jamie Ryan Kiros, Jimmy Ba</i>	
INSTAGAN: INSTANCE-AWARE IMAGE-TO-IMAGE TRANSLATION .....	3713
<i>Sangwoo Mo, Minsu Cho, Jinwoo Shin</i>	
LEARNING TO DESCRIBE SCENES WITH PROGRAMS .....	3739
<i>Yunchao Liu, Zheng Wu, Daniel Ritchie, William T. Freeman, Joshua B. Tenenbaum, Jiajun Wu</i>	
DATA-DEPENDENT CORESETS FOR COMPRESSING NEURAL NETWORKS WITH APPLICATIONS TO GENERALIZATION BOUNDS .....	3752
<i>Cenk Baykal, Lucas Liebenwein, Igor Gilitschenski, Dan Feldman, Daniela Rus</i>	
INFOBOT: TRANSFER AND EXPLORATION VIA THE INFORMATION BOTTLENECK .....	3787
<i>Anirudh Goyal, Riashat Islam, DJ Strouse, Zafarali Ahmed, Hugo Larochelle, Matthew Botvinick, Yoshua Bengio, Sergey Levine</i>	
EMERGING DISENTANGLEMENT IN AUTO-ENCODER BASED UNSUPERVISED IMAGE CONTENT TRANSFER .....	3808
<i>Ori Press, Tomer Galanti, Sagie Benaim, Lior Wolf</i>	

LEARNING EMBEDDINGS INTO ENTROPIC WASSERSTEIN SPACES.....	3833
<i>Charlie Frogner, Farzaneh Mirzazadeh, Justin Solomon</i>	
WASSERSTEIN BARYCENTER MODEL ENSEMBLING .....	3845
<i>Pierre Dognin, Igor Melnyk, Youssef Mroueh, Jarret Ross, Cicero Dos Santos, Tom Sercu,</i>	
GENERALIZABLE ADVERSARIAL TRAINING VIA SPECTRAL NORMALIZATION .....	3870
<i>Farzan Farnia, Jesse Zhang, David Tse</i>	
UNSUPERVISED SPEECH RECOGNITION VIA SEGMENTAL EMPIRICAL OUTPUT DISTRIBUTION MATCHING .....	3897
<i>Chih-Kuan Yeh, Jianshu Chen, Chengzhu Yu, Dong Yu</i>	
LEARNING WHEN TO COMMUNICATE AT SCALE IN MULTIAGENT COOPERATIVE AND COMPETITIVE TASKS .....	3911
<i>Amanpreet Singh, Tushar Jain, Sainbayar Sukhbaatar</i>	

## VOLUME 6

PROXQUANT: QUANTIZED NEURAL NETWORKS VIA PROXIMAL OPERATORS.....	3927
<i>Yu Bai, Yu-Xiang Wang, Edo Liberty</i>	
OPTIMAL COMPLETION DISTILLATION FOR SEQUENCE LEARNING .....	3944
<i>Sara Sabour, William Chan, Mohammad Norouzi</i>	
FEATURE INTERTWINER FOR OBJECT DETECTION .....	3960
<i>Hongyang Li, Bo Dai, Shaoshuai Shi, Wanli Ouyang, Xiaogang Wang</i>	
DIVERSITY AND DEPTH IN PER-EXAMPLE ROUTING MODELS .....	3975
<i>Prajit Ramachandran, Quoc V. Le</i>	
LEARNING CONCISE REPRESENTATIONS FOR REGRESSION BY EVOLVING NETWORKS OF TREES.....	3987
<i>William La Cava, Tilak Raj Singh, James Taggart, Srinivas Suri, Jason H. Moore</i>	
EXPLORATION BY RANDOM NETWORK DISTILLATION .....	4003
<i>Yuri Burda, Harrison Edwards, Amos Storkey, Oleg Klimov</i>	
HIERARCHICAL GENERATIVE MODELING FOR CONTROLLABLE SPEECH SYNTHESIS .....	4020
<i>Wei-Ning Hsu, Yu Zhang, Ron J. Weiss, Heiga Zen, Yonghui Wu, Yuxuan Wang, Yuan Cao, Ye Jia, Zhifeng Chen, Jonathan Shen, Patrick Nguyen, Ruoming Pang</i>	
GENERATIVE QUESTION ANSWERING: LEARNING TO ANSWER THE WHOLE QUESTION .....	4047
<i>Mike Lewis, Angela Fan</i>	
DECOUPLED WEIGHT DECAY REGULARIZATION .....	4061
<i>Ilya Loshchilov, Frank Hutter</i>	
PROBABILISTIC RECURSIVE REASONING FOR MULTI-AGENT REINFORCEMENT LEARNING.....	4079
<i>Ying Wen, Yaodong Yang, Rui Luo, Jun Wang, Wei Pan</i>	
LEARNING TO REPRESENT EDITS .....	4098
<i>Pengcheng Yin, Graham Neubig, Miltiadis Allamanis, Marc Brockschmidt, Alexander L. Gaunt</i>	
UNSUPERVISED DOMAIN ADAPTATION FOR DISTANCE METRIC LEARNING .....	4120
<i>Kihyuk Sohn, Wenling Shang, Xiang Yu, Manmohan Chandraker</i>	

BAYESIAN DEEP CONVOLUTIONAL NETWORKS WITH MANY CHANNELS ARE GAUSSIAN PROCESSES .....	4138
<i>Roman Novak, Lechao Xiao, Yasaman Bahri, Jaehoon Lee, Greg Yang, Jiri Hron, Daniel A. Abolafia, Jeffrey Pennington, Jascha Sohl-dickstein</i>	
EFFICIENT TRAINING ON VERY LARGE CORPORA VIA GRAMIAN ESTIMATION.....	4173
<i>Walid Krichene, Nicolas Mayoraz, Steffen Rendle, Li Zhang, Xinyang Yi, Lichan Hong, Ed Chi, John Anderson</i>	
A COMPREHENSIVE, APPLICATION-ORIENTED STUDY OF CATASTROPHIC FORGETTING IN DNNs.....	4193
<i>B. Pfülb, A. Geppert</i>	
VARIATIONAL AUTOENCODERS WITH JOINTLY OPTIMIZED LATENT DEPENDENCY STRUCTURE.....	4207
<i>Jiawei He, Yu Gong, Joseph Marino, Greg Mori, Andreas Lehrmann</i>	
VERIFICATION OF NON-LINEAR SPECIFICATIONS FOR NEURAL NETWORKS .....	4223
<i>Chongli Qin, Krishnamurthy (Dj) Dvijotham, Brendan O'Donoghue, Rudy Bunel, Robert Stanforth, Sven Gowal, Jonathan Uesato, Grzegorz Swirszcz, Pushmeet Kohli</i>	
IMPROVING SEQUENCE-TO-SEQUENCE LEARNING VIA OPTIMAL TRANSPORT .....	4244
<i>Liqun Chen, Yizhe Zhang, Ruiyi Zhang, Chenyang Tao, Zhe Gan, Haichao Zhang, Bai Li, Dinghan Shen, Changyou Chen, Lawrence Carin</i>	
LEARNING TO PROPAGATE LABELS: TRANSDUCTIVE PROPAGATION NETWORK FOR FEW-SHOT LEARNING.....	4260
<i>Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, Eunho Yang, Sung Ju Hwang, Yi Yang</i>	
UNIVERSAL TRANSFORMERS .....	4274
<i>Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, Lukasz Kaiser</i>	
AN EMPIRICAL STUDY OF BINARY NEURAL NETWORKS' OPTIMISATION .....	4297
<i>Milad Alizadeh, Javier Fernández-Marqués, Nicholas D. Lane, Yarin Gal</i>	
A2BCD: ASYNCHRONOUS ACCELERATION WITH OPTIMAL COMPLEXITY .....	4308
<i>Robert Hannah, Fei Feng, Wotao Yin</i>	
BACKPROPAMINE: TRAINING SELF-MODIFYING NEURAL NETWORKS WITH DIFFERENTIABLE NEUROMODULATED PLASTICITY .....	4342
<i>Thomas Miconi, Aditya Rawal, Jeff Clune, Kenneth O. Stanley</i>	
DISCOVERY OF NATURAL LANGUAGE CONCEPTS IN INDIVIDUAL UNITS OF CNNs.....	4357
<i>Seil Na, Yo Joong Choe, Dong-Hyun Lee, Gunhee Kim</i>	
LEARNING MULTI-LEVEL HIERARCHIES WITH HINDSIGHT.....	4377
<i>Andrew Levy, George Konidaris, Robert Platt, Kate Saenko</i>	
PROMP: PROXIMAL META-POLICY SEARCH.....	4393
<i>Jonas Rothfuss, Dennis Lee, Ignasi Clavera, Tamim Asfour, Pieter Abbeel</i>	
COMPLEMENT OBJECTIVE TRAINING .....	4418
<i>Hao-Yun Chen, Pei-Hsin Wang, Chun-Hao Liu, Shih-Chieh Chang, Jia-Yu Pan, Yu-Ting Chen, Wei Wei, Da-Cheng Juan</i>	
BABYAI: A PLATFORM TO STUDY THE SAMPLE EFFICIENCY OF GROUNDED LANGUAGE LEARNING .....	4429
<i>Maxime Chevalier-Boisvert, Dmitry Bahdanau, Salem Lahlou, Lucas Willems, Chitwan Saharia, Thien Huu Nguyen, Yoshua Bengio</i>	
SLIMMABLE NEURAL NETWORKS .....	4448
<i>Jiahui Yu, Linjie Yang, Ning Xu, Jianchao Yang, Thomas Huang</i>	



LEARNING SELF-IMITATING DIVERSE POLICIES.....	4460
<i>Tanmay Gangwani, Qiang Liu, Jian Peng</i>	
GRAPH HYPERNETWORKS FOR NEURAL ARCHITECTURE SEARCH.....	4478
<i>Chris Zhang, Mengye Ren, Raquel Urtasun</i>	
DEEP LAYERS AS STOCHASTIC SOLVERS .....	4495
<i>Adel Bibi, Bernard Ghanem, Vladlen Koltun, Rene Ranfil</i>	
ARM: AUGMENT-REINFORCE-MERGE GRADIENT FOR STOCHASTIC BINARY NETWORKS.....	4514
<i>Mingzhang Yin, Mingyuan Zhou</i>	
SCALABLE UNBALANCED OPTIMAL TRANSPORT USING GENERATIVE ADVERSARIAL NETWORKS.....	4535
<i>Karren D. Yang, Caroline Uhler</i>	
UNSUPERVISED DISCOVERY OF PARTS, STRUCTURE, AND DYNAMICS.....	4555
<i>Zhenjia Xu, Zhijian Liu, Chen Sun, Kevin Murphy, William T. Freeman, Joshua B. Tenenbaum, Jiajun Wu</i>	
LEARNING MULTIMODAL GRAPH-TO-GRAPH TRANSLATION FOR MOLECULE OPTIMIZATION.....	4570
<i>Wengong Jin, Kevin Yang, Regina Barzilay, Tommi Jaakkola</i>	
HARMONIZING MAXIMUM LIKELIHOOD WITH GANS FOR MULTIMODAL CONDITIONAL GENERATION .....	4583
<i>Soochan Lee, Junsoo Ha, Gunhee Kim</i>	
PHASE-AWARE SPEECH ENHANCEMENT WITH DEEP COMPLEX U-NET .....	4608
<i>Hyeong-Seok Choi, Jang-Hyun Kim, Jaesung Huh, Adrian Kim, Jung-Woo Ha, Kyogu Lee</i>	
THE COMPARATIVE POWER OF RELU NETWORKS AND POLYNOMIAL KERNELS IN THE PRESENCE OF SPARSE LATENT STRUCTURE .....	4628
<i>Frederic Koehler, Andrej Risteski</i>	
NEURAL PROBABILISTIC MOTOR PRIMITIVES FOR HUMANOID CONTROL .....	4647
<i>Josh Merel, Leonard Hasenclever, Alexandre Galashov, Arun Ahuja, Vu Pham, Greg Wayne, Yee Whye Teh, Nicolas Heess</i>	
LEARNING TWO-LAYER NEURAL NETWORKS WITH SYMMETRIC INPUTS.....	4661
<i>Rong Ge, Rohith Kudritipudi, Zhize Li, Xiang Wang</i>	

## VOLUME 7

SPECTRAL INFERENCE NETWORKS: UNIFYING DEEP AND SPECTRAL LEARNING.....	4714
<i>David Pfau, Stig Petersen, Ashish Agarwal, David G. T. Barrett, Kimberly L. Stachenfeld</i>	
ON SELF MODULATION FOR GENERATIVE ADVERSARIAL NETWORKS .....	4740
<i>Ting Chen, Mario Lucic, Neil Houlsby, Sylvain Gelly</i>	
AUTOLOSS: LEARNING DISCRETE SCHEDULE FOR ALTERNATE OPTIMIZATION .....	4758
<i>Haowen Xu, Hao Zhang, Zhiting Hu, Xiaodan Liang, Ruslan Salakhutdinov, Eric Xing</i>	
ANALYZING INVERSE PROBLEMS WITH INVERTIBLE NEURAL NETWORKS.....	4773
<i>Lynton Ardizzone, Jakob Kruse, Carsten Rother, Ullrich Köthe</i>	
LEARNING FINITE STATE REPRESENTATIONS OF RECURRENT POLICY NETWORKS.....	4793
<i>Anurag Koul, Alan Fern, Sam Greycanus</i>	
MEASURING AND REGULARIZING NETWORKS IN FUNCTION SPACE .....	4808
<i>Ari Benjamin, David Rolnick, Konrad Kording</i>	

NO TRAINING REQUIRED: EXPLORING RANDOM ENCODERS FOR SENTENCE CLASSIFICATION .....	4826
<i>John Wieting, Douwe Kiela</i>	
DEEP FRANK-WOLFE FOR NEURAL NETWORK OPTIMIZATION .....	4842
<i>Leonard Berrada, Andrew Zisserman, M. Pawan Kumar</i>	
QUASI-HYPERBOLIC MOMENTUM AND ADAM FOR DEEP LEARNING .....	4873
<i>Jerry Ma, Denis Yarats</i>	
ON COMPUTATION AND GENERALIZATION OF GENERATIVE ADVERSARIAL NETWORKS UNDER SPECTRUM CONTROL .....	4911
<i>Haoming Jiang, Zhehui Chen, Minshuo Chen, Feng Liu, Dingding Wang, Tuo Zhao</i>	
BIG-LITTLE NET: AN EFFICIENT MULTI-SCALE FEATURE REPRESENTATION FOR VISUAL AND SPEECH RECOGNITION .....	4937
<i>Chun-Fu (Richard) Chen, Quanfu Fan, Neil Mallinar, Tom Sercu, Rogerio Feris</i>	
DEEP LAGRANGIAN NETWORKS: USING PHYSICS AS MODEL PRIOR FOR DEEP LEARNING.....	4957
<i>Michael Lutter, Christian Ritter, Jan Peters</i>	
ADVERSARIAL AUDIO SYNTHESIS .....	4974
<i>Chris Donahue, Julian McAuley, Miller Puckette</i>	
A DATA-DRIVEN AND DISTRIBUTED APPROACH TO SPARSE SIGNAL REPRESENTATION AND RECOVERY .....	4990
<i>Ali Mousavi, Gautam Dasarathy, Richard G. Baraniuk</i>	
THE LAPLACIAN IN RL: LEARNING REPRESENTATIONS WITH EFFICIENT APPROXIMATIONS .....	5003
<i>Yifan Wu, George Tucker, Ofir Nachum</i>	
ON THE RELATION BETWEEN THE SHARPEST DIRECTIONS OF DNN LOSS AND THE SGD STEP LENGTH.....	5019
<i>Stanislaw Jastrzębski, Zachary Kenton, Nicolas Ballas, Asja Fischer, Yoshua Bengio, Amos Storkey</i>	
DISCRIMINATOR-ACTOR-CRITIC: ADDRESSING SAMPLE INEFFICIENCY AND REWARD BIAS IN ADVERSARIAL IMITATION LEARNING.....	5038
<i>Ilya Kostrikov, Kumar Krishna Agrawal, Debidatta Dwibedi, Sergey Levine, Jonathan Tompson</i>	
EXCESSIVE INVARIANCE CAUSES ADVERSARIAL VULNERABILITY.....	5053
<i>Joern-Henrik Jacobsen, Jens Behrmann, Richard Zemel, Matthias Bethge</i>	
ADAPTIVITY OF DEEP RELU NETWORK FOR LEARNING IN BESOV AND MIXED SMOOTH BESOV SPACES: OPTIMAL RATE AND CURSE OF DIMENSIONALITY .....	5070
<i>Taiji Suzuki</i>	
ANTISYMMETRICRNN: A DYNAMICAL SYSTEM VIEW ON RECURRENT NEURAL NETWORKS.....	5095
<i>Bo Chang, Minmin Chen, Eldad Haber, Ed H. Chi</i>	
DISCRIMINATOR REJECTION SAMPLING .....	5110
<i>Samaneh Azadi, Catherine Olsson, Trevor Darrell, Ian Goodfellow, Augustus Odena</i>	
UNSUPERVISED LEARNING VIA META-LEARNING.....	5126
<i>Kyle Hsu, Sergey Levine, Chelsea Finn</i>	
RECURRENT EXPERIENCE REPLAY IN DISTRIBUTED REINFORCEMENT LEARNING .....	5150
<i>Steven Kapturowski, Georg Ostrovski, John Quan, Remi Munos, Will Dabney</i>	

QUERY-EFFICIENT HARD-LABEL BLACK-BOX ATTACK: AN OPTIMIZATION-BASED APPROACH.....	5169
<i>Minhao Cheng, Thong Le, Pin-Yu Chen, Huan Zhang, JinFeng Yi, Cho-Jui Hsieh</i>	
MULTI-CLASS CLASSIFICATION WITHOUT MULTI-CLASS LABELS .....	5183
<i>Yen-Chang Hsu, Zhaoyang Lv, Joel Schlosser, Phillip Odom, Zsolt Kira</i>	
LARGE-SCALE ANSWERER IN QUESTIONER'S MIND FOR VISUAL DIALOG QUESTION GENERATION.....	5199
<i>Sang-Woo Lee, Tong Gao, Sohee Yang, Jaejun Yoo, Jung-Woo Ha</i>	
UNSUPERVISED HYPER-ALIGNMENT FOR MULTILINGUAL WORD EMBEDDINGS .....	5215
<i>Jean Alaux, Edouard Grave, Marco Cuturi, Armand Joulin</i>	
DIVERSITY IS ALL YOU NEED: LEARNING SKILLS WITHOUT A REWARD FUNCTION.....	5227
<i>Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, Sergey Levine</i>	
SOLVING THE RUBIK'S CUBE WITH APPROXIMATE POLICY ITERATION.....	5249
<i>Stephen McAleer, Forest Agostinelli, Alexander Shmakov, Pierre Baldi</i>	
DYNAMIC CHANNEL PRUNING: FEATURE BOOSTING AND SUPPRESSION .....	5262
<i>Xitong Gao, Yiren Zhao, Lukasz Dudziak, Robert Mullins, Cheng-zhong Xu</i>	
BEYOND PIXEL NORM-BALLS: PARAMETRIC ADVERSARIES USING AN ANALYTICALLY DIFFERENTIABLE RENDERER .....	5276
<i>Hsueh-Ti Derek Liu, Michael Tao, Chun-Liang Li, Derek Nowrouzezahrai, Alec Jacobson</i>	
DETERMINISTIC PAC-BAYESIAN GENERALIZATION BOUNDS FOR DEEP NETWORKS VIA GENERALIZING NOISE-RESILIENCE.....	5297
<i>Vaishnavh Nagarajan, Zico Kolter</i>	
LEARNING-BASED FREQUENCY ESTIMATION ALGORITHMS .....	5333
<i>Chen-Yu Hsu, Piotr Indyk, Dina Katabi, Ali Vakilian</i>	
FROM HARD TO SOFT: UNDERSTANDING DEEP NETWORK NONLINEARITIES VIA VECTOR QUANTIZATION AND STATISTICAL INFERENCE .....	5353
<i>Randall Balestrieri, Richard Baraniuk</i>	
EPISODIC CURIOSITY THROUGH REACHABILITY .....	5367
<i>Nikolay Savinov, Anton Raichuk, Damien Vincent, Raphael Marinier, Marc Pollefeys, Timothy Lillicrap, Sylvain Gelly</i>	
BAYESIAN PREDICTION OF FUTURE STREET SCENES USING SYNTHETIC LIKELIHOODS .....	5387
<i>Apratim Bhattacharyya, Mario Fritz, Bernt Schiele</i>	
GRADIENT DESCENT PROVABLY OPTIMIZES OVER-PARAMETERIZED NEURAL NETWORKS.....	5403
<i>Simon S. Du, Xiyu Zhai, Barnabas Póczos, Aarti Singh</i>	
BAYESIAN POLICY OPTIMIZATION FOR MODEL UNCERTAINTY.....	5422
<i>Gilwoo Lee, Brian Hou, Aditya Mandalika, Jeongseok Lee, Sanjiban Choudhury, Siddhartha S. Srinivasa</i>	
VON MISES-FISHER LOSS FOR TRAINING SEQUENCE TO SEQUENCE MODELS WITH CONTINUOUS OUTPUTS .....	5437
<i>Sachin Kumar, Yulia Tsvetkov</i>	
INFORMATION-DIRECTED EXPLORATION FOR DEEP REINFORCEMENT LEARNING .....	5452
<i>Nikolay Nikolov, Johannes Kirschner, Felix Berkenkamp, Andreas Krause</i>	
GENERALIZED TENSOR MODELS FOR RECURRENT NEURAL NETWORKS .....	5468
<i>Valentin Khrulkov, Oleksii Hrinchuk, Ivan Oseledets</i>	

INVARIANT AND EQUIVARIANT GRAPH NETWORKS .....	5485
<i>Haggai Maron, Heli Ben-Hamu, Nadav Shamir, Yaron Lipman</i>	

## VOLUME 8

WIZARD OF WIKIPEDIA: KNOWLEDGE-POWERED CONVERSATIONAL AGENTS .....	5499
<i>Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, Jason Weston</i>	
RESIDUAL NON-LOCAL ATTENTION NETWORKS FOR IMAGE RESTORATION.....	5517
<i>Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, Yun Fu</i>	
KERNEL RNN LEARNING (KERNL).....	5535
<i>Christopher Roth, Ingmar Kanitscheider, Ila Fiete</i>	
INTEGER NETWORKS FOR DATA COMPRESSION WITH LATENT-VARIABLE MODELS .....	5546
<i>Johannes Ballé, Nick Johnston, David Minnen</i>	
STRUCTURED ADVERSARIAL ATTACK: TOWARDS GENERAL IMPLEMENTATION AND BETTER INTERPRETABILITY .....	5556
<i>Kaidi Xu, Sijia Liu, Pu Zhao, Pin-Yu Chen, Huan Zhang, Quanfu Fan, Deniz Erdogmus, Yanzhi Wang, Xue Lin</i>	
NEURAL NETWORK GRADIENT-BASED LEARNING OF BLACK-BOX FUNCTION INTERFACES .....	5577
<i>Alon Jacovi, Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, Jonathan Berant</i>	
RNNS IMPLICITLY IMPLEMENT TENSOR-PRODUCT REPRESENTATIONS .....	5591
<i>R. Thomas McCoy, Tal Linzen, Ewan Dunbar, Paul Smolensky</i>	
SELF-MONITORING NAVIGATION AGENT VIA AUXILIARY PROGRESS ESTIMATION .....	5613
<i>Chih-Yao Ma, Jiasen Lu, Zuxuan Wu, Ghassan AlRegib, Zsolt Kira, Richard Socher, Caiming Xiong</i>	
POLICY TRANSFER WITH STRATEGY OPTIMIZATION .....	5631
<i>Wenhao Yu, C. Karen Liu, Greg Turk</i>	
DEEPOBS: A DEEP LEARNING OPTIMIZER BENCHMARK SUITE .....	5647
<i>Frank Schneider, Lukas Balles, Philipp Hennig</i>	
SIGNSGD WITH MAJORITY VOTE IS COMMUNICATION EFFICIENT AND FAULT TOLERANT .....	5661
<i>Jeremy Bernstein, Jiawei Zhao, Kamyar Azizzadenesheli, Anima Anandkumar</i>	
MARGINALIZED AVERAGE ATTENTIONAL NETWORK FOR WEAKLY-SUPERVISED LEARNING.....	5681
<i>Yuan Yuan, Yueming Lyu, Xi Shen, Ivor W. Tsang, Dit-Yan Yeung</i>	
STABLE OPPONENT SHAPING IN DIFFERENTIABLE GAMES .....	5700
<i>Alistair Letcher, Jakob Foerster, David Balduzzi, Tim Rocktäschel, Shimon Whiteson</i>	
BOUNCE AND LEARN: MODELING SCENE DYNAMICS WITH REAL-WORLD BOUNCES .....	5720
<i>Senthil Purushwalkam, Abhinav Gupta, Danny Kaufman, Bryan Russell</i>	
UNDERSTANDING COMPOSITION OF WORD EMBEDDINGS VIA TENSOR DECOMPOSITION .....	5744
<i>Abraham Frandsen, Rong Ge</i>	
SNAS: STOCHASTIC NEURAL ARCHITECTURE SEARCH.....	5768
<i>Sirui Xie, Hehui Zheng, Chunxiao Liu, Liang Lin</i>	
LATENT CONVOLUTIONAL MODELS.....	5785
<i>ShahRukh Athar, Evgeny Burnaev, Victor Lempitsky</i>	

NADPEX: AN ON-POLICY TEMPORALLY CONSISTENT EXPLORATION METHOD FOR DEEP REINFORCEMENT LEARNING.....	5805
<i>Sirui Xie, Junning Huang, Lanxin Lei, Chunxiao Liu, Zheng Ma, Wei Zhang, Liang Lin</i>	
REPRESENTATION DEGENERATION PROBLEM IN TRAINING NATURAL LANGUAGE GENERATION MODELS .....	5823
<i>Jun Gao, Di He, Xu Tan, Tao Qin, Liwei Wang, Tiejian Liu</i>	
SOFT Q-LEARNING WITH MUTUAL-INFORMATION REGULARIZATION .....	5837
<i>Jordi Grau-Moya, Felix Leibfried, Peter Vrancx</i>	
LEARNING TO ADAPT IN DYNAMIC, REAL-WORLD ENVIRONMENTS THROUGH META-REINFORCEMENT LEARNING .....	5856
<i>Anusha Nagabandi, Ignasi Clavera, Simin Liu, Ronald S. Fearing, Pieter Abbeel, Sergey Levine, Chelsea Finn</i>	
HOW IMPORTANT IS A NEURON .....	5873
<i>Kedar Dhamdhere, Mukund Sundararajan, Qiqi Yan</i>	
KNOWLEDGE FLOW: IMPROVE UPON YOUR TEACHERS .....	5888
<i>Iou-Jen Liu, Jian Peng, Alexander Schwing</i>	
EMERGENT COORDINATION THROUGH COMPETITION .....	5905
<i>Siqi Liu, Guy Lever, Josh Merel, Saran Tunyasuvunakool, Nicolas Heess, Thore Graepel</i>	
OPTIMISTIC MIRROR DESCENT IN SADDLE-POINT PROBLEMS: GOING THE EXTRA (GRADIENT) MILE .....	5925
<i>Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, Georgios Piliouras</i>	
MULTILINGUAL NEURAL MACHINE TRANSLATION WITH KNOWLEDGE DISTILLATION .....	5948
<i>Xu Tan, Yi Ren, Di He, Tao Qin, Zhou Zhao, Tie-Yan Liu</i>	
STRUCTURED NEURAL SUMMARIZATION.....	5963
<i>Patrick Fernandes, Miltiadis Allamanis, Marc Brockschmidt</i>	
HYPERBOLIC ATTENTION NETWORKS .....	5981
<i>Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, Nando de Freitas</i>	
DEEP LEARNING 3D SHAPES USING ALT-AZ ANISOTROPIC 2-SPHERE CONVOLUTION.....	5996
<i>Min Liu, Fupin Yao, Chiho Choi, Ayan Sinha, Karthik Ramani</i>	
GENERATIVE PREDECESSOR MODELS FOR SAMPLE-EFFICIENT IMITATION LEARNING.....	6010
<i>Yannick Schroecker, Mel Vecerik, Jon Scholz</i>	
RELATIONAL FORWARD MODELS FOR MULTI-AGENT LEARNING .....	6024
<i>Andrea Tacchetti, H. Francis Song, Pedro A. M. Mediano, Vinicius Zambaldi, János Kramár, Neil C. Rabinowitz, Thore Graepel, Matthew Botvinick, Peter W. Battaglia</i>	
VARIATIONAL BAYESIAN PHYLOGENETIC INFERENCE .....	6040
<i>Cheng Zhang, Frederick A. Matsen IV</i>	
ADV-BNN: IMPROVED ADVERSARIAL DEFENSE THROUGH ROBUST BAYESIAN NEURAL NETWORK .....	6055
<i>Xuanqing Liu, Yao Li, Chongruo Wu, Cho-Jui Hsieh</i>	
H-DETACH: MODIFYING THE LSTM GRADIENT TOWARDS BETTER OPTIMIZATION.....	6068
<i>Bhargav Kanuparthi, Devansh Arpit, Giancarlo Kerg, Nan Rosemary Ke, Ioannis Mitliagkas, Yoshua Bengio</i>	

ON THE LOSS LANDSCAPE OF A CLASS OF DEEP NEURAL NETWORKS WITH NO BAD LOCAL VALLEYS .....	6087
<i>Quynh Nguyen, Mahesh Chandra Mukkamala, Matthias Hein</i>	
ACCUMULATION BIT-WIDTH SCALING FOR ULTRA-LOW PRECISION TRAINING OF DEEP NETWORKS .....	6108
<i>Charbel Sakr, Naigang Wang, Chia-Yu Chen, Jungwook Choi, Ankur Agrawal, Naresh Shanbhag, Kailash Gopalakrishnan</i>	
DEEP CONVOLUTIONAL NETWORKS AS SHALLOW GAUSSIAN PROCESSES .....	6123
<i>Adrià Garriga-Alonso, Carl Edward Rasmussen, Laurence Aitchison</i>	
UNIVERSAL SUCCESSOR FEATURES APPROXIMATORS .....	6139
<i>Diana Borsa, Andre Barreto, John Quan, Daniel J. Mankowitz, Hado van Hasselt, Remi Munos, David Silver, Tom Schaul</i>	
ADAPTIVE ESTIMATORS SHOW INFORMATION COMPRESSION IN DEEP NEURAL NETWORKS .....	6163
<i>Ivan Chelombiev, Conor Houghton, Cian O'Donnell</i>	
CBOW IS NOT ALL YOU NEED: COMBINING CBOW WITH THE COMPOSITIONAL MATRIX SPACE MODEL .....	6174
<i>Florian Mai, Lukas Galke, Ansgar Scherp</i>	
FLUCTUATION-DISSIPATION RELATIONS FOR STOCHASTIC GRADIENT DESCENT .....	6187
<i>Sho Yaida</i>	
A UNIVERSAL MUSIC TRANSLATION NETWORK .....	6202
<i>Noam Mor, Lior Wolf, Adam Polyak, Yaniv Taigman</i>	
NEURAL PERSISTENCE: A COMPLEXITY MEASURE FOR DEEP NEURAL NETWORKS USING ALGEBRAIC TOPOLOGY .....	6215
<i>Bastian Rieck, Matteo Togninalli, Christian Bock, Michael Moor, Max Horn, Thomas Gumbsch, Karsten Borgwardt</i>	
DETECTING EGREGIOUS RESPONSES IN NEURAL SEQUENCE-TO-SEQUENCE MODELS .....	6240
<i>Tianxing He, James Glass</i>	
MEASURING COMPOSITIONALITY IN REPRESENTATION LEARNING .....	6262
<i>Jacob Andreas</i>	
LEARNING REPRESENTATIONS OF SETS THROUGH OPTIMIZED PERMUTATIONS .....	6277
<i>Yan Zhang, Jonathon Hare, Adam Prügel-Bennett</i>	

## VOLUME 9

ANALYSING MATHEMATICAL REASONING ABILITIES OF NEURAL MODELS .....	6303
<i>David Saxton, Edward Grefenstette, Felix Hill, Pushmeet Kohli</i>	
FUNCTIONAL VARIATIONAL BAYESIAN NEURAL NETWORKS .....	6320
<i>Shengyang Sun, Guodong Zhang, Jiaxin Shi, Roger Grosse</i>	
IMPROVING MMD-GAN TRAINING WITH REPULSIVE LOSS FUNCTION .....	6342
<i>Wei Wang, Yuan Sun, Saman Halgamuge</i>	
SOM-VAE: INTERPRETABLE DISCRETE REPRESENTATION LEARNING ON TIME SERIES .....	6366
<i>Vincent Fortuin, Matthias Hüser, Francesco Locatello, Heiko Strathmann, Gunnar Rätsch</i>	
RIEMANNIAN ADAPTIVE OPTIMIZATION METHODS .....	6384
<i>Gary Becigneul, Octavian-Eugen Ganea</i>	

NON-VACUOUS GENERALIZATION BOUNDS AT THE IMAGENET SCALE: A PAC-BAYESIAN COMPRESSION APPROACH .....	6400
<i>Wenda Zhou, Victor Veitch, Morgane Austern, Ryan P. Adams, Peter Orbanz</i>	
LEARNING FACTORIZED MULTIMODAL REPRESENTATIONS.....	6416
<i>Yao-Hung Hubert Tsai, Paul Pu Liang, Amir Zadeh, Louis-Philippe Morency, Ruslan Salakhutdinov</i>	
AGGREGATED MOMENTUM: STABILITY THROUGH PASSIVE DAMPING .....	6436
<i>James Lucas, Shengyang Sun, Richard Zemel, Roger Grosse</i>	
LEARNING TO SCHEDULE COMMUNICATION IN MULTI-AGENT REINFORCEMENT LEARNING.....	6458
<i>Daewoo Kim, Sangwoo Moon, David Hostallero, Wan Ju Kang, Taeyoung Lee, Kyunghwan Son, Yung Yi</i>	
MINIMUM DIVERGENCE VS. MAXIMUM MARGIN: AN EMPIRICAL COMPARISON ON SEQ2SEQ MODELS.....	6475
<i>Huan Zhang, Hai Zhao</i>	
OVERCOMING CATASTROPHIC FORGETTING FOR CONTINUAL LEARNING VIA MODEL ADAPTATION .....	6488
<i>Wenpeng Hu, Zhou Lin, Bing Liu, Chongyang Tao, Zhengwei Tao, Jinwen Ma, Dongyan Zhao, Rui Yan</i>	
MULTI-DOMAIN ADVERSARIAL LEARNING.....	6501
<i>Alice Schoenauer-Sebag, Louise Heinrich, Marc Schoenauer, Michele Sebag, Lani F. Wu, Steve J. Altschuler</i>	
LEARNING FROM POSITIVE AND UNLABELED DATA WITH A SELECTION BIAS .....	6525
<i>Masahiro Kato, Takeshi Teshima, Junya Honda</i>	
CEM-RL: COMBINING EVOLUTIONARY AND GRADIENT-BASED METHODS FOR POLICY SEARCH.....	6542
<i>Pourchot, Sigaud</i>	
SGD CONVERGES TO GLOBAL MINIMUM IN DEEP LEARNING VIA STAR-CONVEX PATH .....	6560
<i>Yi Zhou, Junjie Yang, Huishuai Zhang, Yingbin Liang, Vahid Tarokh</i>	
LAYOUTGAN: GENERATING GRAPHIC LAYOUTS WITH WIREFRAME DISCRIMINATORS.....	6574
<i>Jianan Li, Jimei Yang, Aaron Hertzmann, Jianming Zhang, Tingfa Xu</i>	
EQUI-NORMALIZATION OF NEURAL NETWORKS .....	6590
<i>Pierre Stock, Benjamin Graham, Rémi Gribonval, Hervé Jégou</i>	
INFORMATION THEORETIC LOWER BOUNDS ON NEGATIVE LOG LIKELIHOOD .....	6610
<i>Luis A. Lastras-Montaño</i>	
NEURAL SPEED READING WITH STRUCTURAL-JUMP-LSTM .....	6623
<i>Christian Hansen, Casper Hansen, Stephen Alstrup, Jakob Grue Simonsen, Christina Lioma</i>	
DEEP GRAPH INFOMAX .....	6633
<i>Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, R Devon Hjelm</i>	
SNIP: SINGLE-SHOT NETWORK PRUNING BASED ON CONNECTION SENSITIVITY .....	6650
<i>Namhoon Lee, Thalaiyasingam Ajanthan, Philip Torr</i>	
BIAS-REDUCED UNCERTAINTY ESTIMATION FOR DEEP NEURAL CLASSIFIERS.....	6665
<i>Yonatan Geifman, Guy Uziel, Ran El-Yaniv</i>	

APPROXIMABILITY OF DISCRIMINATORS IMPLIES DIVERSITY IN GANS .....	6679
<i>Yu Bai, Tengyu Ma, Andrej Risteski</i>	
ON THE MINIMAL SUPERVISION FOR TRAINING ANY BINARY CLASSIFIER FROM ONLY UNLABELED DATA .....	6717
<i>Nan Lu, Gang Niu, Aditya Krishna Menon, Masashi Sugiyama</i>	
AUXILIARY VARIATIONAL MCMC.....	6735
<i>Raza Habib, David Barber</i>	
PATE-GAN: GENERATING SYNTHETIC DATA WITH DIFFERENTIAL PRIVACY GUARANTEES .....	6748
<i>James Jordon, Jinsung Yoon, Mihaela van der Schaar</i>	
MINIMAL RANDOM CODE LEARNING: GETTING BITS BACK FROM COMPRESSED MODEL PARAMETERS .....	6769
<i>Marton Havasi, Robert Peharz, José Miguel Hernández-Lobato</i>	
GO GRADIENT FOR EXPECTATION-BASED OBJECTIVES .....	6780
<i>Yulai Cong, Miaoyun Zhao, Ke Bai, Lawrence Carin</i>	
BENCHMARKING NEURAL NETWORK ROBUSTNESS TO COMMON CORRUPTIONS AND PERTURBATIONS .....	6810
<i>Dan Hendrycks, Thomas Dietterich</i>	
DEEP REINFORCEMENT LEARNING WITH RELATIONAL INDUCTIVE BIASES .....	6826
<i>Vinicius Zambaldi, David Raposo, Adam Santoro, Victor Bapst, Yujia Li, Igor Babuschkin, Karl Tuyls, David Reichert, Timothy Lillicrap, Edward Lockhart, Murray Shanahan, Victoria Langston, Razvan Pascanu, Matthew Botvinick, Oriol Vinyals, Peter Battaglia</i>	
L-SHAPLEY AND C-SHAPLEY: EFFICIENT MODEL INTERPRETATION FOR STRUCTURED DATA.....	6844
<i>Jianbo Chen, Le Song, Martin J. Wainwright, Michael I. Jordan</i>	
WHITENING AND COLORING BATCH TRANSFORM FOR GANS.....	6877
<i>Aliaksandr Siarohin, Enver Sangineto, Nicu Sebe</i>	
PEERNETS: EXPLOITING PEER WISDOM AGAINST ADVERSARIAL ATTACKS .....	6898
<i>Jan Svoboda, Jonathan Masci, Federico Monti, Michael Bronstein, Leonidas Guibas</i>	
SPARSE DICTIONARY LEARNING BY DYNAMICAL NEURAL NETWORKS.....	6914
<i>Tsung-Han Lin, Ping Tak Peter Tang</i>	
RELAXED QUANTIZATION FOR DISCRETIZED NEURAL NETWORKS.....	6934
<i>Christos Louizos, Matthias Reisser, Tijmen Blankevoort, Efstratios Gavves, Max Welling</i>	
MARGINAL POLICY GRADIENTS: A UNIFIED FAMILY OF ESTIMATORS FOR BOUNDED ACTION SPACES WITH APPLICATIONS.....	6949
<i>Carson Eisenach, Haichuan Yang, Ji Liu, Han Liu</i>	
CODE2SEQ: GENERATING SEQUENCES FROM STRUCTURED REPRESENTATIONS OF CODE.....	6969
<i>Uri Alon, Shaked Brody, Omer Levy, Eran Yahav</i>	
ADEF: AN ITERATIVE ALGORITHM TO CONSTRUCT ADVERSARIAL DEFORMATIONS .....	6991
<i>Rima Alaifari, Giovanni S. Alberti, Tandri Gauksson</i>	
SMALL NONLINEARITIES IN ACTIVATION FUNCTIONS CREATE BAD LOCAL MINIMA IN NEURAL NETWORKS.....	7016
<i>Chulhee Yun, Suvrit Sra, Ali Jadbabaie</i>	



VISCERAL MACHINES: RISK-AVERSION IN REINFORCEMENT LEARNING WITH  
INTRINSIC PHYSIOLOGICAL REWARDS..... 7049  
*Daniel McDuff, Ashish Kapoor*

ACCELERATING NONCONVEX LEARNING VIA REPLICA EXCHANGE LANGEVIN  
DIFFUSION..... 7060  
*Yi Chen, Jinglin Chen, Jing Dong, Jian Peng, Zhaoran Wang*

## VOLUME 10

THE SINGULAR VALUES OF CONVOLUTIONAL LAYERS..... 7083  
*Hanie Sedghi, Vineet Gupta, Philip M. Long*

IMPROVING GENERALIZATION AND STABILITY OF GENERATIVE ADVERSARIAL  
NETWORKS..... 7095  
*Hoang Thanh-Tung, Truyen Tran, Svetha Venkatesh*

GAMEPAD: A LEARNING ENVIRONMENT FOR THEOREM PROVING .....7113  
*Daniel Huang, Prafulla Dhariwal, Dawn Song, Ilya Sutskever*

TOWARDS UNDERSTANDING REGULARIZATION IN BATCH NORMALIZATION..... 7125  
*Ping Luo, Xinjiang Wang, Wenqi Shao, Zhanglin Peng*

LEARNING TO MAKE ANALOGIES BY CONTRASTING ABSTRACT RELATIONAL  
STRUCTURE..... 7148  
*Felix Hill, Adam Santoro, David Barrett, Ari Morcos, Timothy Lillicrap*

ATTENTION, LEARN TO SOLVE ROUTING PROBLEMS!..... 7166  
*Wouter Kool, Herke van Hoof, Max Welling*

CRITICAL LEARNING PERIODS IN DEEP NETWORKS..... 7191  
*Alessandro Achille, Matteo Rovere, Stefano Soatto*

META-LEARNING PROBABILISTIC INFERENCE FOR PREDICTION..... 7205  
*Jonathan Gordon, John Bronskill, Matthias Bauer, Sebastian Nowozin, Richard Turner*

TRELLIS NETWORKS FOR SEQUENCE MODELING ..... 7226  
*Shaojie Bai, J. Zico Kolter, Vladlen Koltun*

GENERATIVE CODE MODELING WITH GRAPHS ..... 7244  
*Marc Brockschmidt, Miltiadis Allamanis, Alexander L. Gaunt, Oleksandr Polozov*

LEARNING RECURRENT BINARY/TERNARY WEIGHTS..... 7268  
*Arash Ardakani, Zhengyun Ji, Sean C. Smithson, Brett H. Meyer, Warren J. Gross*

DYNAMICALLY UNFOLDING RECURRENT RESTORER: A MOVING ENDPOINT  
CONTROL METHOD FOR IMAGE RESTORATION..... 7287  
*Xiaoshuai Zhang, Yiping Lu, Jiaying Liu, Bin Dong*

LEARNING A SAT SOLVER FROM SINGLE-BIT SUPERVISION..... 7299  
*Daniel Selsam, Matthew Lamm, Benedikt Bunz, Percy Liang, Leonardo de Moura, David L.  
Dill*

OPTIMAL TRANSPORT MAPS FOR DISTRIBUTION PRESERVING OPERATIONS ON  
LATENT SPACES OF GENERATIVE MODELS..... 7310  
*Eirikur Agustsson, Alexander Sage, Radu Timofte, Luc Van Gool*

EFFICIENT LIFELONG LEARNING WITH A-GEM ..... 7327  
*Arslan Chaudhry, Marc Aurelio Ranzato, Marcus Rohrbach, Mohamed Elhoseiny*

META-LEARNING FOR STOCHASTIC GRADIENT MCMC..... 7347  
*Wenbo Gong, Yingzhen Li, José Miguel Hernández-Lobato*

AUGMENTED CYCLIC ADVERSARIAL LEARNING FOR LOW RESOURCE DOMAIN ADAPTATION .....	7369
<i>Ehsan Hosseini-Asl, Yingbo Zhou, Caiming Xiong, Richard Socher</i>	
OFF-POLICY EVALUATION AND LEARNING FROM LOGGED BANDIT FEEDBACK: ERROR REDUCTION VIA SURROGATE POLICY .....	7383
<i>Yuan Xie, Boyi Liu, Qiang Liu, Zhaoran Wang, Yuan Zhou, Jian Peng</i>	
DOUBLE VITERBI: WEIGHT ENCODING FOR HIGH COMPRESSION RATIO AND FAST ON-CHIP RECONSTRUCTION FOR DEEP NEURAL NETWORK.....	7405
<i>Daehyun Ahn, Dongsoo Lee, Taesu Kim, Jae-Joon Kim</i>	
GRAPH WAVELET NEURAL NETWORK.....	7420
<i>Bingbing Xu, Huawei Shen, Qi Cao, Yunqi Qiu, Xueqi Cheng</i>	
THE UNUSUAL EFFECTIVENESS OF AVERAGING IN GAN TRAINING.....	7433
<i>Yasin Yazici, Chuan-Sheng Foo, Stefan Winkler, Kim-Hui Yap, Georgios Piliouras, Vijay Chandrasekhar</i>	
EVALUATING ROBUSTNESS OF NEURAL NETWORKS WITH MIXED INTEGER PROGRAMMING.....	7455
<i>Vincent Tjeng, Kai Y. Xiao, Russ Tedrake</i>	
TOWARDS THE FIRST ADVERSARIALY ROBUST NEURAL NETWORK MODEL ON MNIST .....	7476
<i>Lukas Schott, Jonas Rauber, Matthias Bethge, Wieland Brendel</i>	
ON THE TURING COMPLETENESS OF MODERN NEURAL NETWORK ARCHITECTURES.....	7493
<i>Jorge Pérez, Javier Marinković, Pablo Barceló</i>	
ADAPTIVE POSTERIOR LEARNING: FEW-SHOT LEARNING WITH A SURPRISE-BASED MEMORY MODULE .....	7529
<i>Tiago Ramalho, Marta Garnelo</i>	
A CLOSER LOOK AT DEEP LEARNING HEURISTICS: LEARNING RATE RESTARTS, WARMUP AND DISTILLATION .....	7543
<i>Akhilesh Gotmare, Nitish Shirish Keskar, Caiming Xiong, Richard Socher</i>	
COARSE-GRAIN FINE-GRAIN COATTENTION NETWORK FOR MULTI-EVIDENCE QUESTION ANSWERING .....	7559
<i>Victor Zhong, Caiming Xiong, Nitish Shirish Keskar, Richard Socher</i>	
NEAR-OPTIMAL REPRESENTATION LEARNING FOR HIERARCHICAL REINFORCEMENT LEARNING .....	7586
<i>Ofir Nachum, Shixiang Gu, Honglak Lee, Sergey Levine</i>	
EXECUTION-GUIDED NEURAL PROGRAM SYNTHESIS.....	7604
<i>Xinyun Chen, Chang Liu, Dawn Song</i>	
IMPOSING CATEGORY TREES ONTO WORD-EMBEDDINGS USING A GEOMETRIC CONSTRUCTION .....	7619
<i>Tiansi Dong, Chrisitan Bauckhage, Hailong Jin, Juanzi Li, Olaf Cremers, Daniel Speicher, Armin B. Cremers, Joerg Zimmermann</i>	
ROBUST ESTIMATION VIA GENERATIVE ADVERSARIAL NETWORKS .....	7629
<i>Chao GAO, jiyi LIU, Yuan YAO, Weizhi ZHU</i>	
WOULDA, COULDA, SHOULDA: COUNTERFACTUALLY-GUIDED POLICY SEARCH .....	7656
<i>Lars Buesing, Theophane Weber, Yori Zwols, Nicolas Heess, Sebastien Racaniere, Arthur Guez, Jean-Baptiste Lespiau</i>	
ROBUST CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS.....	7671
<i>Grigorios G. Chrysos, Jean Kossaiifi, Stefanos Zafeiriou</i>	

ATTENTIVE NEURAL PROCESSES .....	7698
<i>Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, Yee Whye Teh</i>	
VISUAL REASONING BY PROGRESSIVE MODULE NETWORKS .....	7716
<i>Seung Wook Kim, Makarand Tapaswi, Sanja Fidler</i>	
HINDSIGHT POLICY GRADIENTS.....	7734
<i>Paulo Rauber, Avinash Ummadisingu, Filipe Mutz, Jürgen Schmidhuber</i>	
LEMONADE: LEARNED MOTIF AND NEURONAL ASSEMBLY DETECTION IN CALCIUM IMAGING VIDEOS.....	7772
<i>Elke Kirschbaum, Manuel Hausmann, Steffen Wolf, Hannah Sonntag, Justus Schneider, Shehabeldin Elzoheiry, Oliver Kann, Daniel Durstewitz, Fred A Hamprecht</i>	
VISUAL EXPLANATION BY INTERPRETATION: IMPROVING VISUAL FEEDBACK CAPABILITIES OF DEEP NEURAL NETWORKS.....	7800
<i>Jose Oramas, Kaili Wang, Tinne Tuytelaars</i>	
MAX-MIG: AN INFORMATION THEORETIC APPROACH FOR JOINT LEARNING FROM CROWDS.....	7829
<i>Peng Cao, Yilun Xu, Yuqing Kong, Yizhou Wang</i>	
HIERARCHICAL VISUOMOTOR CONTROL OF HUMANOIDS .....	7847
<i>Josh Merel, Arun Ahuja, Vu Pham, Saran Tunyasuvunakool, Siqi Liu, Dhruva Tirumala, Nicolas Heess, Greg Wayne</i>	

## VOLUME 11

FUNCTION SPACE PARTICLE OPTIMIZATION FOR BAYESIAN NEURAL NETWORKS .....	7866
<i>Ziyu Wang, Tongzheng Ren, Jun Zhu, Bo Zhang</i>	
FEED-FORWARD PROPAGATION IN PROBABILISTIC NEURAL NETWORKS WITH CATEGORICAL AND MAX LAYERS.....	7891
<i>Alexander Shekhovtsov, Boris Flach</i>	
HIERARCHICAL REINFORCEMENT LEARNING VIA ADVANTAGE-WEIGHTED INFORMATION MAXIMIZATION.....	7912
<i>Takayuki Osa, Voot Tangkaratt, Masashi Sugiyama</i>	
LARGE-SCALE STUDY OF CURIOSITY-DRIVEN LEARNING .....	7928
<i>Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, Alexei A. Efros</i>	
STROKENET: A NEURAL PAINTING ENVIRONMENT .....	7945
<i>Ningyuan Zheng, Yifan Jiang, Dingjiang Huang</i>	
DIALOGWAE: MULTIMODAL RESPONSE GENERATION WITH CONDITIONAL WASSERSTEIN AUTO-ENCODER .....	7957
<i>Xiaodong Gu, Kyunghyun Cho, Jung-Woo Ha, Sunghun Kim</i>	
QUATERNION RECURRENT NEURAL NETWORKS.....	7968
<i>Titouan Parcollet, Mirco Ravanelli, Mohamed Morchid, Georges Linarès, Chiheb Trabelsi, Renato De Mori, Yoshua Bengio</i>	
REWARD CONSTRAINED POLICY OPTIMIZATION.....	7987
<i>Chen Tessler, Daniel J. Mankowitz, Shie Mannor</i>	
LEARNING LATENT SUPERSTRUCTURES IN VARIATIONAL AUTOENCODERS FOR DEEP MULTIDIMENSIONAL CLUSTERING.....	8002
<i>Xiaopeng Li, Zhoung Chen, Leonard K. M. Poon, Nevin L. Zhang</i>	

VARIATIONAL SMOOTHING IN RECURRENT NEURAL NETWORK LANGUAGE MODELS .....	8017
<i>Lingpeng Kong, Gabor Melis, Wang Ling, Lei Yu, Dani Yogatama</i>	
INITIALIZED EQUILIBRIUM PROPAGATION FOR BACKPROP-FREE TRAINING .....	8027
<i>Peter O'Connor, Efstratios Gavves, Max Welling</i>	
IDENTIFYING AND CONTROLLING IMPORTANT NEURONS IN NEURAL MACHINE TRANSLATION .....	8042
<i>Anthony Bau, Yonatan Belinkov, Hassan Sajjad, Nadir Durrani, Fahim Dalvi, James Glass</i>	
SIGNSGD VIA ZERO-ORDER ORACLE.....	8061
<i>Sijia Liu, Pin-Yu Chen, Xiangyi Chen, Mingyi Hong</i>	
DELTA: DEEP LEARNING TRANSFER USING FEATURE MAP WITH ATTENTION FOR CONVOLUTIONAL NETWORKS.....	8085
<i>Xingjian Li, Haoyi Xiong, Hanchao Wang, Yuxuan Rao, Liping Liu, Jun Huan</i>	
VARIANCE NETWORKS: WHEN EXPECTATION DOES NOT MEET YOUR EXPECTATIONS .....	8098
<i>Kirill Neklyudov, Dmitry Molchanov, Arsenii Ashukha, Dmitry Vetrov</i>	
CONDITIONAL NETWORK EMBEDDINGS.....	8114
<i>Bo Kang, Jefrey Lijffijt, Tijl De Bie</i>	
DISTRIBUTION-INTERPOLATION TRADE OFF IN GENERATIVE MODELS .....	8130
<i>Damian Leśniak, Igor Sieradzki, Igor Podolak</i>	
SAMPLE EFFICIENT ADAPTIVE TEXT-TO-SPEECH .....	8149
<i>Yutian Chen, Yannis Assael, Brendan Shillingford, David Budden, Scott Reed, Heiga Zen, Quan Wang, Luis C. Cobo, Andrew Trask, Ben Laurie, Caglar Gulcehre, Aäron van den Oord, Oriol Vinyals, Nando de Freitas</i>	
MAXIMAL DIVERGENCE SEQUENTIAL AUTOENCODER FOR BINARY SOFTWARE VULNERABILITY DETECTION.....	8164
<i>Tue Le, Tuan Nguyen, Trung Le, Dinh Phung, Paul Montague, Olivier De Vel, Lizhen Qu</i>	
SAMPLE EFFICIENT IMITATION LEARNING FOR CONTINUOUS CONTROL.....	8179
<i>Fumihiko Sasaki, Tetsuya Yohira, Atsuo Kawaguchi</i>	
PRACTICAL LOSSLESS COMPRESSION WITH LATENT VARIABLES USING BITS BACK CODING .....	8191
<i>James Townsend, Thomas Bird, David Barber</i>	
CONTEXT-ADAPTIVE ENTROPY MODEL FOR END-TO-END OPTIMIZED IMAGE COMPRESSION .....	8204
<i>Jooyoung Lee, Seunghyun Cho, Seung-Kwon Beack</i>	
ANALYSIS OF QUANTIZED MODELS .....	8224
<i>Lu Hou, Ruiliang Zhang, James T. Kwok</i>	
GENERATING MULTIPLE OBJECTS AT SPATIALLY DISTINCT LOCATIONS.....	8242
<i>Tobias Hinz, Stefan Heinrich, Stefan Wermter</i>	
ANYTIME MINIBATCH: EXPLOITING STRAGGLERS IN ONLINE DISTRIBUTED OPTIMIZATION.....	8265
<i>Nuwan Ferdinand, Haider Al-Lawati, Stark Draper, Matthew Nokleby</i>	
A ROTATION-EQUIVARIANT CONVOLUTIONAL NEURAL NETWORK MODEL OF PRIMARY VISUAL CORTEX .....	8292
<i>Alexander S. Ecker, Fabian H. Sinz, Emmanouil Froudarakis, Paul G. Fahey, Santiago A. Cadena, Edgar Y. Walker, Erick Cobos, Jacob Reimer, Andreas S. Tolias, Matthias Bethge</i>	

AN ANALYTIC THEORY OF GENERALIZATION DYNAMICS AND TRANSFER LEARNING IN DEEP LINEAR NETWORKS .....	8304
<i>Andrew K. Lampinen, Surya Ganguli</i>	
ARE ADVERSARIAL EXAMPLES INEVITABLE? .....	8324
<i>Ali Shafahi, W. Ronny Huang, Christoph Studer, Soheil Feizi, Tom Goldstein</i>	
DIRECTED-INFO GAIL: LEARNING HIERARCHICAL POLICIES FROM UNSEGMENTED DEMONSTRATIONS USING DIRECTED INFORMATION .....	8341
<i>Mohit Sharma, Arjun Sharma, Nicholas Rhinehart, Kris M. Kitani</i>	
M <sup>3</sup> RL: MIND-AWARE MULTI-AGENT MANAGEMENT REINFORCEMENT LEARNING.....	8356
<i>Tianmin Shu, Yuandong Tian</i>	
DIFFERENTIABLE LEARNING-TO-NORMALIZE VIA SWITCHABLE NORMALIZATION .....	8374
<i>Ping Luo, Jiamin Ren, Zhanglin Peng, Ruimao Zhang, Jingyu Li</i>	
SUPERVISED POLICY UPDATE FOR DEEP REINFORCEMENT LEARNING.....	8392
<i>Quan Vuong, Yiming Zhang, Keith W. Ross</i>	
ADVERSARIAL DOMAIN ADAPTATION FOR STABLE BRAIN-MACHINE INTERFACES.....	8413
<i>Ali Farshchian, Juan A. Gallego, Joseph P. Cohen, Yoshua Bengio, Lee E. Miller, Sara A. Solla</i>	
LEARNING FACTORIZED REPRESENTATIONS FOR OPEN-SET DOMAIN ADAPTATION.....	8427
<i>Mahsa Baktashmotlagh, Masoud Faraki, Tom Drummond, Mathieu Salzmann</i>	
LEARNING TO UNDERSTAND GOAL SPECIFICATIONS BY MODELLING REWARD .....	8438
<i>Dzmitry Bahdanau, Felix Hill, Jan Leike, Edward Hughes, Arian Hosseini, Pushmeet Kohli, Edward Grefenstette</i>	
DYNAMIC SPARSE GRAPH FOR EFFICIENT DEEP LEARNING.....	8457
<i>Liu Liu, Lei Deng, Xing Hu, Maohua Zhu, Guoqi Li, Yufei Ding, Yuan Xie</i>	
HIERARCHICAL INTERPRETATIONS FOR NEURAL NETWORK PREDICTIONS .....	8475
<i>Chandan Singh, W. James Murdoch, Bin Yu</i>	
POST SELECTION INFERENCE WITH INCOMPLETE MAXIMUM MEAN DISCREPANCY ESTIMATOR.....	8501
<i>Makoto Yamada, Denny Wu, Yao-Hung Hubert Tsai, Hirofumi Ohta, Ruslan Salakhutdinov, Ichiro Takeuchi, Kenji Fukumizu</i>	
DIFFUSION SCATTERING TRANSFORMS ON GRAPHS .....	8517
<i>Fernando Gama, Alejandro Ribeiro, Joan Bruna</i>	
PRECONDITIONER ON MATRIX LIE GROUP FOR SGD .....	8534
<i>Xi-Lin Li</i>	
DPSNET: END-TO-END DEEP PLANE SWEEP STEREO .....	8546
<i>Sunghoon Im, Hae-Gon Jeon, Stephen Lin, In So Kweon</i>	
DARTS: DIFFERENTIABLE ARCHITECTURE SEARCH .....	8558
<i>Hanxiao Liu, Karen Simonyan, Yiming Yang</i>	
ADVERSARIAL REPROGRAMMING OF NEURAL NETWORKS.....	8571
<i>Gamaleldin F. Elsayed, Ian Goodfellow, Jascha Sohl-Dickstein</i>	
OPPORTUNISTIC LEARNING: BUDGETED COST-SENSITIVE LEARNING FROM DATA STREAMS .....	8586
<i>Mohammad Kachuee, Orpaz Goldstein, Kimmo Kärkkäinen, Sajad Darabi, Majid Sarrafzadeh</i>	
INVASE: INSTANCE-WISE VARIABLE SELECTION USING NEURAL NETWORKS .....	8598
<i>Jinsung Yoon, James Jordon, Mihaela van der Schaar</i>	

THE RELATIVISTIC DISCRIMINATOR: A KEY ELEMENT MISSING FROM STANDARD GAN .....	8622
<i>Alexia Jolicoeur-Martineau</i>	

## VOLUME 12

A KERNEL RANDOM MATRIX-BASED APPROACH FOR SPARSE PCA .....	8648
<i>Mohamed El Amine Seddik, Mohamed Tamaazousti, Romain Couillet</i>	
CAVEATS FOR INFORMATION BOTTLENECK IN DETERMINISTIC SCENARIOS .....	8664
<i>Artemy Kolchinsky, Brendan D. Tracey, Steven Van Kuyk</i>	
LEARNING LOCALIZED GENERATIVE MODELS FOR 3D POINT CLOUDS VIA GRAPH CONVOLUTION .....	8687
<i>Diego Valsesia, Giulia Fracastoro, Enrico Magli</i>	
UNDERSTANDING AND IMPROVING INTERPOLATION IN AUTOENCODERS VIA AN ADVERSARIAL REGULARIZER .....	8702
<i>David Berthelot, Colin Raffel, Aurko Roy, Ian Goodfellow</i>	
ADVERSARIAL IMITATION VIA VARIATIONAL INVERSE REINFORCEMENT LEARNING.....	8722
<i>Ahmed H. Qureshi, Byron Boots, Michael C. Yip</i>	
GENERATING LIQUID SIMULATIONS WITH DEFORMATION-AWARE NEURAL NETWORKS.....	8736
<i>Lukas Prantl, Boris Bonev, Nils Thuerey</i>	
L2-NONEXPANSIVE NEURAL NETWORKS .....	8757
<i>Haifeng Qian, Mark N. Wegman</i>	
DEEP, SKINNY NEURAL NETWORKS ARE NOT UNIVERSAL APPROXIMATORS.....	8775
<i>Jesse Johnson</i>	
LARGE SCALE GRAPH LEARNING FROM SMOOTH SIGNALS.....	8785
<i>Vassilis Kalofolias, Nathanaël Perraudin</i>	
ROTDCF: DECOMPOSITION OF CONVOLUTIONAL FILTERS FOR ROTATION-EQUIVARIANT DEEP NETWORKS .....	8804
<i>Xiuyuan Cheng, Qiang Qiu, Robert Calderbank, Guillermo Sapiro</i>	
PER-TENSOR FIXED-POINT QUANTIZATION OF THE BACK-PROPAGATION ALGORITHM.....	8822
<i>Charbel Sakr, Naresh Shanbhag</i>	
UNDERSTANDING STRAIGHT-THROUGH ESTIMATOR IN TRAINING ACTIVATION QUANTIZED NEURAL NETS .....	8848
<i>Penghang Yin, Jiancheng Lyu, Shuai Zhang, Stanley Osher, Yingyong Qi, Jack Xin</i>	
CONVOLUTIONAL NEURAL NETWORKS ON NON-UNIFORM GEOMETRICAL SIGNALS USING EUCLIDEAN SPECTRAL TRANSFORMATION .....	8878
<i>Chiyu Max Jiang, Dequan Wang, Jingwei Huang, Philip Marcus, Matthias Niessner</i>	
TRAINING FOR FASTER ADVERSARIAL ROBUSTNESS VERIFICATION VIA INDUCING RELU STABILITY .....	8894
<i>Kai Y. Xiao, Vincent Tjeng, Nur Muhammad (Mahi) Shafiullah, Aleksander Madry</i>	

### **ORAL PRESENTATIONS**

BA-NET: DENSE BUNDLE ADJUSTMENT NETWORKS.....	8914
<i>Chengzhou Tang, Ping Tan</i>	

TRANSFERRING KNOWLEDGE ACROSS LEARNING PROCESSES .....	8931
<i>Sebastian Flennerhag, Pablo G. Moreno, Neil D. Lawrence, Andreas Damianou</i>	
THE LOTTERY TICKET HYPOTHESIS: FINDING SPARSE, TRAINABLE NEURAL NETWORKS.....	8954
<i>Jonathan Frankle, Michael Carbin</i>	
PAY LESS ATTENTION WITH LIGHTWEIGHT AND DYNAMIC CONVOLUTIONS.....	8996
<i>Felix Wu, Angela Fan, Alexei Baevski, Yann Dauphin, Michael Auli</i>	
TEMPORAL DIFFERENCE VARIATIONAL AUTO-ENCODER.....	9010
<i>Karol Gregor, George Papamakarios, Frederic Besse, Lars Buesing, Theophane Weber</i>	
THE NEURO-SYMBOLIC CONCEPT LEARNER: INTERPRETING SCENES, WORDS, AND SENTENCES FROM NATURAL SUPERVISION .....	9027
<i>Jiayuan Mao, Chuang Gan, Pushmeet Kohli, Joshua B. Tenenbaum, Jiajun Wu</i>	
LEARNING PROTEIN STRUCTURE WITH A DIFFERENTIABLE SIMULATOR .....	9055
<i>John Ingraham, Adam Riesselman, Chris Sander, Debora Marks</i>	
FFJORD: FREE-FORM CONTINUOUS DYNAMICS FOR SCALABLE REVERSIBLE GENERATIVE MODELS.....	9079
<i>Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, Ilya Sutskever, David Duvenaud</i>	
ENABLING FACTORIZED PIANO MUSIC MODELING AND GENERATION WITH THE MAESTRO DATASET.....	9092
<i>Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, Douglas Eck</i>	
HOW POWERFUL ARE GRAPH NEURAL NETWORKS? .....	9104
<i>Keyulu Xu, Weihua Hu, Jure Leskovec, Stefanie Jegelka</i>	
GENERATING HIGH FIDELITY IMAGES WITH SUBSCALE PIXEL NETWORKS AND MULTIDIMENSIONAL UPSCALING.....	9121
<i>Jacob Menick, Nal Kalchbrenner</i>	
A UNIFIED THEORY OF EARLY VISUAL REPRESENTATIONS FROM RETINA TO CORTEX THROUGH ANATOMICALLY CONSTRAINED DEEP CNNs.....	9136
<i>Jack Lindsey, Samuel A. Ocko, Surya Ganguli, Stephane Deny</i>	
LEARNING DEEP REPRESENTATIONS BY MUTUAL INFORMATION ESTIMATION AND MAXIMIZATION .....	9153
<i>R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, Yoshua Bengio</i>	
LEARNING ROBUST REPRESENTATIONS BY PROJECTING SUPERFICIAL STATISTICS OUT.....	9177
<i>Haohan Wang, Zexue He, Zachary C. Lipton, Eric P. Xing</i>	
META-LEARNING UPDATE RULES FOR UNSUPERVISED REPRESENTATION LEARNING .....	9193
<i>Luke Metz, Niru Maheswaranathan, Brian Cheung, Jascha Sohl-Dickstein</i>	
ORDERED NEURONS: INTEGRATING TREE STRUCTURES INTO RECURRENT NEURAL NETWORKS.....	9220
<i>Yikang Shen, Shawn Tan, Alessandro Sordani, Aaron Courville</i>	
IMAGENET-TRAINED CNNs ARE BIASED TOWARDS TEXTURE; INCREASING SHAPE BIAS IMPROVES ACCURACY AND ROBUSTNESS .....	9234
<i>Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, Wieland Brendel</i>	
LARGE SCALE GAN TRAINING FOR HIGH FIDELITY NATURAL IMAGE SYNTHESIS .....	9256
<i>Andrew Brock, Jeff Donahue, Karen Simonyan</i>	

KNOCKOFFGAN: GENERATING KNOCKOFFS FOR FEATURE SELECTION USING GENERATIVE ADVERSARIAL NETWORKS .....	9291
<i>James Jordon, Jinsung Yoon, Mihaela van der Schaar</i>	
LEARNING TO REMEMBER MORE WITH LESS MEMORIZATION .....	9316
<i>Hung Le, Truyen Tran, Svetha Venkatesh</i>	
DETERMINISTIC VARIATIONAL INFERENCE FOR ROBUST BAYESIAN NEURAL NETWORKS .....	9336
<i>Anqi Wu, Sebastian Nowozin, Edward Meeds, Richard E. Turner, José Miguel Hernández-Lobato, Alexander L. Gaunt</i>	
SLALOM: FAST, VERIFIABLE AND PRIVATE EXECUTION OF NEURAL NETWORKS IN TRUSTED HARDWARE .....	9360
<i>Florian Tramer, Dan Boneh</i>	
ON RANDOM DEEP WEIGHT-TIED AUTOENCODERS: EXACT ASYMPTOTIC ANALYSIS, PHASE TRANSITIONS, AND IMPLICATIONS TO TRAINING .....	9379
<i>Ping Li, Phan-Minh Nguyen</i>	
SMOOTHING THE GEOMETRY OF PROBABILISTIC BOX EMBEDDINGS.....	9411
<i>Xiang Li, Luke Vilnis, Dongxu Zhang, Michael Boratko, Andrew McCallum</i>	

**Author Index**