5th International Conference on Learning Representations (ICLR 2017)

Toulon, France
24 – 26 April 2017

Volume 1 of 4

ISBN: 978-1-7138-7271-9

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© (2017) 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

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
Web: www.proceedings.com

TABLE OF CONTENTS

VOLUME 1

POSTER PRESENTATIONS

FRACTALNET: ULTRA-DEEP NEURAL NETWORKS WITHOUT RESIDUALS 1
Gustav Larsson, Michael Maire, Gregory Shakhnarovich

DEEP INFORMATION PROPAGATION..... 12
Samuel S. Schoenholz, Justin Gilmer, Surya Ganguli, Jascha Sohl-Dickstein

PRUNING CONVOLUTIONAL NEURAL NETWORKS FOR RESOURCE EFFICIENT INFERENCE..... 30
Pavlo Molchanov, Stephen Tyree, Tero Karras, Timo Aila, Jan Kautz

RECURRENT BATCH NORMALIZATION 47
Tim Cooijmans, Nicolas Ballas, César Laurent, Çağlar Gülçehre, Aaron Courville

BETA-VAE: LEARNING BASIC VISUAL CONCEPTS WITH A CONSTRAINED VARIATIONAL FRAMEWORK..... 60
Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, Alexander Lerchner

WORDS OR CHARACTERS? FINE-GRAINED GATING FOR READING COMPREHENSION..... 82
Zhilin Yang, Bhuwan Dhingra, Ye Yuan, Junjie Hu, William W. Cohen, Ruslan Salakhutdinov

DEEPDSL: A COMPILATION-BASED DOMAIN-SPECIFIC LANGUAGE FOR DEEP LEARNING..... 92
Tian Zhao, Xiao Bing Huang, Yu Cao

HYPERNETWORKS..... 103
David Ha, Andrew M. Dai, Quoc V. Le

CAPACITY AND TRAINABILITY IN RECURRENT NEURAL NETWORKS..... 121
Jasmine Collins, Jascha Sohl-Dickstein, David Sussillo

RECURRENT HIDDEN SEMI-MARKOV MODEL..... 138
Hanjun Dai, Bo Dai, Yan-Ming Zhang, Shuang Li, Le Song

LEARNING CURVE PREDICTION WITH BAYESIAN NEURAL NETWORKS..... 155
Aaron Klein, Stefan Falkner, Jost Tobias Springenberg, Frank Hutter

A SIMPLE BUT TOUGH-TO-BEAT BASELINE FOR SENTENCE EMBEDDINGS..... 171
Sanjeev Arora, Yingyu Liang, Tengyu Ma

LEARNING TO REPEAT: FINE GRAINED ACTION REPETITION FOR DEEP REINFORCEMENT LEARNING 187
Sahil Sharma, Aravind S. Lakshminarayanan, Balaraman Ravindran

IMPROVING NEURAL LANGUAGE MODELS WITH A CONTINUOUS CACHE 211
Edouard Grave, Armand Joulin, Nicolas Usunier

SNAPSHOT ENSEMBLES: TRAIN 1, GET M FOR FREE 220
Gao Huang, Yixuan Li, Geoff Pleiss, Zhuang Liu, John E. Hopcroft, Kilian Q. Weinberger

SOFT WEIGHT-SHARING FOR NEURAL NETWORK COMPRESSION 234
Karen Ullrich, Edward Meeds, Max Welling

LEARNING TO PERFORM PHYSICS EXPERIMENTS VIA DEEP REINFORCEMENT LEARNING 250
Misha Denil, Pulkit Agrawal, Tejas D Kulkarni, Tom Erez, Peter Battaglia, Nando de Freitas

QUERY-REDUCTION NETWORKS FOR QUESTION ANSWERING 265
Minjoon Seo, Sewon Min, Ali Farhadi, Hannaneh Hajishirzi

ADVERSARIAL MACHINE LEARNING AT SCALE 278
Alexey Kurakin, Ian J. Goodfellow, Samy Bengio

VARIATIONAL RECURRENT ADVERSARIAL DEEP DOMAIN ADAPTATION 295
Sanjay Purushotham, Wilka Carvalho, Tanachat Nilanon, Yan Liu

DISCRETE VARIATIONAL AUTOENCODERS 310
Jason Tyler Rolfe

SPARSELY-CONNECTED NEURAL NETWORKS: TOWARDS EFFICIENT VLSI IMPLEMENTATION OF DEEP NEURAL NETWORKS 343
Arash Ardakani, Carlo Condo, Warren J. Gross

RECURRENT MIXTURE DENSITY NETWORK FOR SPATIOTEMPORAL VISUAL ATTENTION 357
Loris Bazzani, Hugo Larochelle, Lorenzo Torresani

EFFICIENT REPRESENTATION OF LOW-DIMENSIONAL MANIFOLDS USING DEEP NETWORKS 372
Ronen Basri, David W. Jacobs

REINFORCEMENT LEARNING THROUGH ASYNCHRONOUS ADVANTAGE ACTOR-CRITIC ON A GPU 385
Mohammad Babaeizadeh, Iuri Frosio, Stephen Tyree, Jason Clemons, Jan Kautz

TOPOLOGY AND GEOMETRY OF HALF-RECTIFIED NETWORK OPTIMIZATION 397
C. Daniel Freeman, Joan Bruna

TIGHTER BOUNDS LEAD TO IMPROVED CLASSIFIERS 416
Nicolas Le Roux

ZONEOUT: REGULARIZING RNNS BY RANDOMLY PRESERVING HIDDEN ACTIVATIONS 427
David Krueger, Tegan Maharaj, Janos Kramar, Mohammad Pezeshki, Nicolas Ballas, Nan Rosemary Ke, Anirudh Goyal, Yoshua Bengio, Aaron Courville, Christopher Pal

AN INFORMATION-THEORETIC FRAMEWORK FOR FAST AND ROBUST UNSUPERVISED LEARNING VIA NEURAL POPULATION INFOMAX 438
Wentao Huang, Kechen Zhang

THIRD PERSON IMITATION LEARNING 463
Bradly C Stadie, Pieter Abbeel, Ilya Sutskever

DEEP VARIATIONAL BAYES FILTERS: UNSUPERVISED LEARNING OF STATE SPACE MODELS FROM RAW DATA 479
Maximilian Karl, Maximilian Soelch, Justin Bayer, Patrick van der Smagt

LEARNING INVARIANT REPRESENTATIONS OF PLANAR CURVES 492
Gautam Pai, Aaron Wetzler, Ron Kimmel

DENSITY ESTIMATION USING REAL NVP 503
Laurent Dinh, Jascha Sohl-Dickstein, Samy Bengio

METACONTROL FOR ADAPTIVE IMAGINATION-BASED OPTIMIZATION 535
Jessica B. Hamrick, Andrew J. Ballard, Razvan Pascanu, Oriol Vinyals, Nicolas Heess, Peter W. Battaglia

THE NEURAL NOISY CHANNEL 556
Lei Yu, Phil Blunsom, Chris Dyer, Edward Grefenstette, Tomas Kocisky

A COMPARE-AGGREGATE MODEL FOR MATCHING TEXT SEQUENCES 569
Shuohang Wang, Jing Jiang

CALIBRATING ENERGY-BASED GENERATIVE ADVERSARIAL NETWORKS 580
Zihang Dai, Amjad Almahairi, Philip Bachman, Eduard Hovy, Aaron Courville

LEARNING A NATURAL LANGUAGE INTERFACE WITH NEURAL PROGRAMMER 597
Arvind Neelakantan, Quoc V. Le, Martin Abadi, Andrew McCallum, Dario Amodei

LEARNING TO SUPEROPTIMIZE PROGRAMS 610
Rudy Bunel, Alban Desmaison, M. Pawan Kumar, Philip H.S. Torr, Pushmeet Kohli

OPTIMAL BINARY AUTOENCODING WITH PAIRWISE CORRELATIONS 624
Akshay Balsubramani

NEURAL PROGRAM LATTICES 644
Chengtao Li, Daniel Tarlow, Alexander L. Gaunt, Marc Brockschmidt, Nate Kushman

ADVERSARIALLY LEARNED INFERENCE 661
Vincent Dumoulin, Ishmael Belghazi, Ben Poole, Alex Lamb, Martin Arjovsky, Olivier Mastropietro, Aaron Courville

FRUSTRATINGLY SHORT ATTENTION SPANS IN NEURAL LANGUAGE MODELING 679
Michał Daniluk, Tim Rocktäschel, Johannes Welbl, Sebastian Riedel

STEERABLE CNNS 689
Taco S. Cohen, Max Welling

MULTI-VIEW RECURRENT NEURAL ACOUSTIC WORD EMBEDDINGS 703
Wanjia He, Weiran Wang, Karen Livescu

TRANSFER LEARNING FOR SEQUENCE TAGGING WITH HIERARCHICAL RECURRENT NETWORKS 716
Zhilin Yang, Ruslan Salakhutdinov, William W. Cohen

TRAINING AGENT FOR FIRST-PERSON SHOOTER GAME WITH ACTOR-CRITIC CURRICULUM LEARNING 726
Yuxin Wu, Yuandong Tian

LOSSY IMAGE COMPRESSION WITH COMPRESSIVE AUTOENCODERS 736
Lucas Theis, Wenzhe Shi, Andrew Cunningham, Ferenc Huszár

LEARNING AND POLICY SEARCH IN STOCHASTIC DYNAMICAL SYSTEMS WITH BAYESIAN NEURAL NETWORKS 755
Stefan Depeweg, José Miguel Hernández-Lobato, Finale Doshi-Velez, Steffen Udluft

VOLUME 2

NEURAL PHOTO EDITING WITH INTROSPECTIVE ADVERSARIAL NETWORKS 769
Andrew Brock, Theodore Lim, J.M. Ritchie, Nick Weston

BATCH POLICY GRADIENT METHODS FOR IMPROVING NEURAL CONVERSATION MODELS 784
Kirthevasan Kandasamy, Yoram Bachrach, Ryota Tomioka, Daniel Tarlow, David Carter

QUASI-RECURRENT NEURAL NETWORKS 798
James Bradbury, Stephen Merity, Caiming Xiong, Richard Socher

SIGMA DELTA QUANTIZED NETWORKS 810
Peter O'Connor, Max Welling

HIGHWAY AND RESIDUAL NETWORKS LEARN UNROLLED ITERATIVE ESTIMATION 825
Klaus Greff, Rupesh K. Srivastava, Jürgen Schmidhuber

MODE REGULARIZED GENERATIVE ADVERSARIAL NETWORKS 839
Tong Che, Yanran Li, Athul Jacob, Yoshua Bengio, Wenjie Li

INTROSPECTION:ACCELERATING NEURAL NETWORK TRAINING BY LEARNING WEIGHT EVOLUTION 852
Abhishek Sinha, Aahitagni Mukherjee, Mausoom Sarkar, Balaji Krishnamurthy

IMPROVING GENERATIVE ADVERSARIAL NETWORKS WITH DENOISING FEATURE MATCHING 867
David Warde-Farley, Yoshua Bengio

OUTRAGEOUSLY LARGE NEURAL NETWORKS: THE SPARSELY-GATED MIXTURE-OF-EXPERTS LAYER 878
Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, Jeff Dean

NORMALIZING THE NORMALIZERS: COMPARING AND EXTENDING NETWORK NORMALIZATION SCHEMES 897
Mengye Ren, Renjie Liao, Raquel Urtasun, Fabian H. Sinz, Richard S. Zemel

TOWARDS DEEP INTERPRETABILITY (MUS-ROVER II): LEARNING HIERARCHICAL REPRESENTATIONS OF TONAL MUSIC 913
Haizi Yu, Lav R. Varshney

INCORPORATING LONG-RANGE CONSISTENCY IN CNN-BASED TEXTURE GENERATION 927
Guillaume Berger, Roland Memisevic

SUPPORT REGULARIZED SPARSE CODING AND ITS FAST ENCODER 946
Yingzhen Yang, Jiahui Yu, Pushmeet Kohli, Jianchao Yang, Thomas S. Huang

TRANSFER OF VIEW-MANIFOLD LEARNING TO SIMILARITY PERCEPTION OF NOVEL OBJECTS 965
Xingyu Lin, Hao Wang, Zhihao Li, Yimeng Zhang, Alan Yuille, Tai Sing Lee

VARIABLE COMPUTATION IN RECURRENT NEURAL NETWORKS 978
Yacine Jernite, Edouard Grave, Armand Joulin, Tomas Mikolov

LEARNING VISUAL SERVOING WITH DEEP FEATURES AND FITTED Q-ITERATION 990
Alex X. Lee, Sergey Levine, Pieter Abbeel

LR-GAN: LAYERED RECURSIVE GENERATIVE ADVERSARIAL NETWORKS FOR IMAGE GENERATION 1010
Jianwei Yang, Anitha Kannan, Dhruv Batra, Devi Parikh

LEARNING RECURRENT REPRESENTATIONS FOR HIERARCHICAL BEHAVIOR MODELING 1031
Eyrun Eyjolfsdottir, Kristin Branson, Yisong Yue, Pietro Perona

EPISODIC EXPLORATION FOR DEEP DETERMINISTIC POLICIES FOR STARCRAFT MICROMANAGEMENT 1043
Nicolas Usunier, Gabriel Synnaeve, Zeming Lin, Soumith Chintala

LEARNING TO NAVIGATE IN COMPLEX ENVIRONMENTS 1059
Piotr Mirowski, Razvan Pascanu, Fabio Viola, Hubert Soyer, Andy Ballard, Andrea Banino, Misha Denil, Ross Goroshin, Laurent Sifre, Koray Kavukcuoglu, Dharshan Kumaran, Raia Hadsell

A RECURRENT NEURAL NETWORK WITHOUT CHAOS 1075
Thomas Laurent, James von Brecht

ENERGY-BASED GENERATIVE ADVERSARIAL NETWORKS 1086
Junbo Zhao, Michael Mathieu, Yann LeCun

DESIGNING NEURAL NETWORK ARCHITECTURES USING REINFORCEMENT LEARNING 1103
Bowen Baker, Otkrist Gupta, Nikhil Naik, Ramesh Raskar

STRUCTURED ATTENTION NETWORKS 1121
Yoon Kim, Carl Denton, Luong Hoang, Alexander M. Rush

A STRUCTURED SELF-ATTENTIVE SENTENCE EMBEDDING 1142
Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, Yoshua Bengio

LEARNING TO PLAY IN A DAY: FASTER DEEP REINFORCEMENT LEARNING BY OPTIMALITY TIGHTENING 1157
Frank S.He, Yang Liu, Alexander G. Schwing, Jian Peng

DEEP LEARNING WITH DYNAMIC COMPUTATION GRAPHS 1170
Moshe Looks, Marcello Herreshoff, DeLesley Hutchins, Peter Norvig

DEEP VARIATIONAL INFORMATION BOTTLENECK 1182
Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, Kevin Murphy

GENERALIZING SKILLS WITH SEMI-SUPERVISED REINFORCEMENT LEARNING 1201
Chelsea Finn, Tianhe Yu, Justin Fu, Pieter Abbeel, Sergey Levine

LOSS-AWARE BINARIZATION OF DEEP NETWORKS 1212
Lu Hou, Quanming Yao, James T. Kwok

MACHINE COMPREHENSION USING MATCH-LSTM AND ANSWER POINTER 1223
Shuohang Wang, Jing Jiang

RECURRENT ENVIRONMENT SIMULATORS 1238
Silvia Chiappa, Sébastien Racaniere, Daan Wierstra, Shakir Mohamed

HYPERBAND: BANDIT-BASED CONFIGURATION EVALUATION FOR HYPERPARAMETER OPTIMIZATION 1299
Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, Ameet Talwalkar

HIERARCHICAL MULTISCALE RECURRENT NEURAL NETWORKS 1314
Junyoung Chung, Sungjin Ahn, Yoshua Bengio

TREE-STRUCTURED DECODING WITH DOUBLY-RECURRENT NEURAL NETWORKS 1327
David Alvarez-Melis, Tommi S. Jaakkola

TRAINING COMPRESSED FULLY-CONNECTED NETWORKS WITH A DENSITY-DIVERSITY PENALTY 1344
Shengjie Wang, Haoran Cai, Jeff Bilmes, William Noble

DIET NETWORKS: THIN PARAMETERS FOR FAT GENOMICS 1354
Adriana Romero, Pierre Luc Carrier, Akram Erraqabi, Tristan Sylvain, Alex Auvolat, Etienne Dejoie, Marc-André Legault, Marie-Pierre Dubé, Julie G. Hussin, Yoshua Bengio

WHAT DOES IT TAKE TO GENERATE NATURAL TEXTURES? 1365
Ivan Ustyuzhaninov , Wieland Brendel , Leon Gatys, Matthias Bethge

LEARNING FEATURES OF MUSIC FROM SCRATCH 1378
John Thickstun, Zaid Harchaoui, Sham Kakade

AUTOMATIC RULE EXTRACTION FROM LONG SHORT TERM MEMORY NETWORKS 1392
W. James Murdoch, Arthur Szlam

REASONING WITH MEMORY AUGMENTED NEURAL NETWORKS FOR LANGUAGE COMPREHENSION 1404
Tsendsuren Munkhdalai, Hong Yu

GEOMETRY OF POLYSEMY 1417
Jiaqi Mu, Suma Bhat, Pramod Viswanath

UNSUPERVISED CROSS-DOMAIN IMAGE GENERATION 1441
Yaniv Taigman, Adam Polyak, Lior Wolf

LEARNING INVARIANT FEATURE SPACES TO TRANSFER SKILLS WITH REINFORCEMENT LEARNING 1456
Abhishek Gupta, Coline Devin, YuXuan Liu, Pieter Abbeel, Sergey Levine

ENTROPY-SGD: BIASING GRADIENT DESCENT INTO WIDE VALLEYS 1470
Pratik Chaudhari, Anna Choromanska, Stefano Soatto, Yann LeCun, Carlo Baldassi, Christian Borgs, Jennifer Chayes, Levent Sagun, Riccardo Zecchina

PAYING MORE ATTENTION TO ATTENTION: IMPROVING THE PERFORMANCE OF CONVOLUTIONAL NEURAL NETWORKS VIA ATTENTION TRANSFER 1489
Sergey Zagoruyko, Nikos Komodakis

FILTER SHAPING FOR CONVOLUTIONAL NEURAL NETWORKS 1502
Xingyi Li, Fuxin Li, Xiaoli Fern, Raviv Raich

VOLUME 3

UNROLLED GENERATIVE ADVERSARIAL NETWORKS 1516
Luke Metz, Ben Poole, David Pfau, Jascha Sohl-Dickstein

AN ACTOR-CRITIC ALGORITHM FOR SEQUENCE PREDICTION 1541
Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron Courville, Yoshua Bengio

LEARNING TO GENERATE SAMPLES FROM NOISE THROUGH INFUSION TRAINING 1558
Florian Bordes, Sina Honari, Pascal Vincent

DEEP PROBABILISTIC PROGRAMMING 1577
Dustin Tran, Matthew D. Hoffman, Rif A. Saurous, Eugene Brevdo, Kevin Murphy, David M. Blei

MOLLIFYING NETWORKS 1595
Caglar Gulcehre, Marcin Moczulski, Francesco Visin, Yoshua Bengio

PROGRAM SYNTHESIS FOR CHARACTER LEVEL LANGUAGE MODELING 1610
Pavol Bielik, Veselin Raychev, Martin Vechev

IDENTITY MATTERS IN DEEP LEARNING 1627
Moritz Hardt, Tengyu Ma

HADAMARD PRODUCT FOR LOW-RANK BILINEAR POOLING 1641
Jin-Hwa Kim, Kyoung-Woon On, Woosang Lim, Jeonghee Kim, Jung-Woo Ha, Byoung-Tak Zhang

INDUCTIVE BIAS OF DEEP CONVOLUTIONAL NETWORKS THROUGH POOLING GEOMETRY 1655
Nadav Cohen, Amnon Shashua

PRUNING FILTERS FOR EFFICIENT CONVNETS 1683
Hao Li, Asim Kadav, Igor Durdanovic, Hanan Samet, Hans Peter Graf

SAMPLERNN: AN UNCONDITIONAL END-TO-END NEURAL AUDIO GENERATION MODEL 1696
Soroush Mehri, Kundan Kumar, Ishaan Gulrajani, Rithesh Kumar, Shubham Jain, Jose Sotelo, Aaron Courville, Yoshua Bengio

DEEP MULTI-TASK REPRESENTATION LEARNING: A TENSOR FACTORISATION APPROACH 1707
Yongxin Yang, Timothy M. Hospedales

EPOPT: LEARNING ROBUST NEURAL NETWORK POLICIES USING MODEL ENSEMBLES 1719
Aravind Rajeswaran, Sarvjeet Ghotra, Balaraman Ravindran, Sergey Levine

HOLSTEP: A MACHINE LEARNING DATASET FOR HIGHER-ORDER LOGIC THEOREM PROVING 1734
Cezary Kaliszyk, François Chollet, Christian Szegedy

TOWARDS A NEURAL STATISTICIAN 1746
Harrison Edwards, Amos Storkey

EXPLORING SPARSITY IN RECURRENT NEURAL NETWORKS 1759
Sharan Narang, Greg Diamos, Shubho Sengupta, Erich Elsen

SGDR: STOCHASTIC GRADIENT DESCENT WITH WARM RESTARTS 1769
Ilya Loshchilov, Frank Hutter

PREDICTING MEDICATIONS FROM DIAGNOSTIC CODES WITH RECURRENT NEURAL NETWORKS 1785
Jacek M. Bajor, Thomas A. Lasko

ONLINE BAYESIAN TRANSFER LEARNING FOR SEQUENTIAL DATA MODELING 1804
Priyank Jaini, Zhitang Chen, Pablo Carbajal, Edith Law, Laura Middleton, Kayla Regan, Mike Schaekermann, George Trimponias, James Tung, Pascal Poupart

FINE-GRAINED ANALYSIS OF SENTENCE EMBEDDINGS USING AUXILIARY PREDICTION TASKS 1824
Yossi Adi, Einat Kermany, Yonatan Belinkov, Ofer Lavi, Yoav Goldberg

TRACKING THE WORLD STATE WITH RECURRENT ENTITY NETWORKS 1837
Mikael Henaff, Jason Weston, Arthur Szlam, Antoine Bordes, Yann LeCun

POINTER SENTINEL MIXTURE MODELS 1851
Stephen Merity, Caiming Xiong, James Bradbury, Richard Socher

TRAINING DEEP NEURAL-NETWORKS USING A NOISE ADAPTATION LAYER 1866
Jacob Goldberger, Ehud Ben-Reuven

SAMPLE EFFICIENT ACTOR-CRITIC WITH EXPERIENCE REPLAY 1875
Ziyu Wang, Victor Bapst, Nicolas Heess, Volodymyr Mnih, Remi Munos, Koray Kavukcuoglu, Nando de Freitas

REVISITING CLASSIFIER TWO-SAMPLE TESTS 1895
David Lopez-Paz, Maxime Oquab

PIXELCNN++: IMPROVING THE PIXELCNN WITH DISCRETIZED LOGISTIC MIXTURE LIKELIHOOD AND OTHER MODIFICATIONS 1910
Tim Salimans, Andrej Karpathy, Xi Chen, Diederik P. Kingma

CATEGORICAL REPARAMETERIZATION WITH GUMBEL-SOFTMAX 1920
Eric Jang, Shixiang Gu, Ben Poole

CENTRAL MOMENT DISCREPANCY (CMD) FOR DOMAIN-INVARIANT REPRESENTATION LEARNING 1932
Werner Zellinger, Thomas Grubinger, Edwin Lughofer, Thomas Natschläger, Susanne Saminger-Platz

REGULARIZING CNNS WITH LOCALLY CONSTRAINED DECORRELATIONS 1945
Pau Rodríguez, Jordi Gonzàlez, Guillem Cucurull, Josep M. Gonfaus, Xavier Roca

DEEP PREDICTIVE CODING NETWORKS FOR VIDEO PREDICTION AND UNSUPERVISED LEARNING 1956
William Lotter, Gabriel Kreiman, David Cox

STICK-BREAKING VARIATIONAL AUTOENCODERS 1974
Eric Nalisnick, Padhraic Smyth

WHY DEEP NEURAL NETWORKS FOR FUNCTION APPROXIMATION? 1986
Shiyu Liang, R. Srikant

MAXIMUM ENTROPY FLOW NETWORKS 2003
Gabriel Loaiza-Ganem , Yuanjun Gao , John P. Cunningham

LATENT SEQUENCE DECOMPOSITIONS 2016
William Chan, Yu Zhang, Quoc Le, Navdeep Jaitly

GENERATIVE MODELS AND MODEL CRITICISM VIA OPTIMIZED MAXIMUM MEAN DISCREPANCY 2028
Danica J. Sutherland, Hsiao-Yu Tung, Heiko Strathmann, Soumyajit De, Aaditya Ramdas, Alex Smola, Arthur Gretton

EFFICIENT VECTOR REPRESENTATION FOR DOCUMENTS THROUGH CORRUPTION 2039
Minmin Chen

DEEPCODER: LEARNING TO WRITE PROGRAMS 2052
Matej Balog, Alexander L. Gaunt, Marc Brockschmidt, Sebastian Nowozin, Daniel Tarlow

ON DETECTING ADVERSARIAL PERTURBATIONS 2072
Jan Hendrik Metzen, Tim Genewein, Volker Fischer, Bastian Bischoff

LEARNING TO REMEMBER RARE EVENTS 2084
Lukasz Kaiser, Ofir Nachum, Aurko Roy, Samy Bengio

DATA NOISING AS SMOOTHING IN NEURAL NETWORK LANGUAGE MODELS 2094
Ziang Xie, Sida I. Wang, Jiwei Li, Daniel Lévy, Aiming Nie, Dan Jurafsky, Andrew Y. Ng

TRAINED TERNARY QUANTIZATION 2106
Chenzhuo Zhu, Song Han, Huizi Mao, William J. Dally

DO DEEP CONVOLUTIONAL NETS REALLY NEED TO BE DEEP AND CONVOLUTIONAL? 2116
Gregor Urban, Krzysztof J. Geras, Samira Ebrahimi Kahou, Ozlem Aslan, Shengjie Wang, Abdelrahman Mohamed, Matthai Philipose, Matt Richardson, Rich Caruana

VARIATIONAL LOSSY AUTOENCODER 2129
Xi Chen, Diederik P. Kingma, Tim Salimans, Yan Duan, Prafulla Dhariwal, John Schulman, Ilya Sutskever, Pieter Abbeel

GENERATIVE MULTI-ADVERSARIAL NETWORKS 2146
Ishan Durugkar, Ian Gemp, Sridhar Mahadevan

INCREMENTAL NETWORK QUANTIZATION: TOWARDS LOSSLESS CNNS WITH LOW-PRECISION WEIGHTS 2160
Aojun Zhou, Anbang Yao, Yiwen Guo, Lin Xu, Yurong Chen

MULTILAYER RECURRENT NETWORK MODELS OF PRIMATE RETINAL GANGLION CELL RESPONSES 2174
Eleanor Batty, Josh Merel, Nora Brackbill, Alexander Heitman, Alexander Sher, Alan Litke, E.J. Chichilnisky, Liam Paninski

A COMPOSITIONAL OBJECT-BASED APPROACH TO LEARNING PHYSICAL DYNAMICS 2186
Michael Chang, Tomer Ullman, Antonio Torralba, Joshua Tenenbaum

DECOMPOSING MOTION AND CONTENT FOR NATURAL VIDEO SEQUENCE PREDICTION 2201
Ruben Villegas, Jimei Yang, Seunghoon Hong, Xunyu Lin, Honglak Lee

AUTOENCODING VARIATIONAL INFERENCE FOR TOPIC MODELS 2223
Akash Srivastava, Charles Sutton

DELVING INTO TRANSFERABLE ADVERSARIAL EXAMPLES AND BLACK-BOX ATTACKS 2235
Yanpei Liu, Xinyun Chen, Chang Liu, Dawn Song

LEARNING TO OPTIMIZE 2249
Ke Li, Jitendra Malik

UNDERSTANDING TRAINABLE SPARSE CODING WITH MATRIX FACTORIZATION 2262
Thomas Moreau, Joan Bruna

VOLUME 4

EMERGENCE OF FOVEAL IMAGE SAMPLING FROM LEARNING TO ATTEND IN VISUAL SCENES 2275
Brian Cheung, Eric Weiss, Bruno Olshausen

IMPROVING POLICY GRADIENT BY EXPLORING UNDER-APPRECIATED REWARDS 2284
Ofir Nachum, Mohammad Norouzi, Dale Schuurmans

LEARNING TO QUERY, REASON, AND ANSWER QUESTIONS ON AMBIGUOUS TEXTS 2301
Xiaoxiao Guo, Tim Klinger, Clemens Rosenbaum, Joseph P. Bigus, Murray Campbell, Ban Kawas, Kartik Talamadupula, Gerry Tesauro, Satinder Singh

NEURO-SYMBOLIC PROGRAM SYNTHESIS 2313
Emilio Parisotto, Abdel-rahman Mohamed, Rishabh Singh, Lihong Li, Dengyong Zhou, Pushmeet Kohli

DSD: DENSE-SPARSE-DENSE TRAINING FOR DEEP NEURAL NETWORKS 2328
Song Han, Jeff Pool, Sharan Narang, Huizi Mao, Enhao Gong, Shijian Tang, Erich Elsen, Peter Vajda, Manohar Paluri, John Tran, Bryan Catanzaro, William J. Dally

PIXELVAE: A LATENT VARIABLE MODEL FOR NATURAL IMAGES 2341
Ishaan Gulrajani, Kundan Kumar, Faruk Ahmed, Adrien Ali Taiga, Francesco Visin, David Vazquez, Aaron Courville

DEEP BIAFFINE ATTENTION FOR NEURAL DEPENDENCY PARSING 2357
Timothy Dozat, Christopher D. Manning

TOWARDS THE LIMIT OF NETWORK QUANTIZATION 2365
Yoojin Choi, Mostafa El-Khamy, Jungwon Lee

STOCHASTIC NEURAL NETWORKS FOR HIERARCHICAL REINFORCEMENT LEARNING 2379
Carlos Florensa, Yan Duan, Pieter Abbeel

DYNAMIC COATTENTION NETWORKS FOR QUESTION ANSWERING 2396
Caiming Xiong, Victor Zhong, Richard Socher

A BASELINE FOR DETECTING MISCLASSIFIED AND OUT-OF-DISTRIBUTION EXAMPLES IN NEURAL NETWORKS 2410
Dan Hendrycks, Kevin Gimpel

ADVERSARIAL FEATURE LEARNING 2422
Jeff Donahue, Philipp Krähenbühl, Trevor Darrell

DROPOUT WITH EXPECTATION-LINEAR REGULARIZATION 2440
Xuezhe Ma, Yingkai Gao, Zhiting Hu, Yaoliang Yu, Yuntian Deng, Eduard Hovy

FASTER CNNS WITH DIRECT SPARSE CONVOLUTIONS AND GUIDED PRUNING 2463
Jongsoo Park, Sheng Li, Wei Wen, Ping Tak Peter Tang, Hai Li, Yiran Chen, Pradeep Dubey

DISTRIBUTED SECOND-ORDER OPTIMIZATION USING KRONECKER-FACTORED APPROXIMATIONS 2474
Jimmy Ba, Roger Grosse, James Martens

TOPICRNN: A RECURRENT NEURAL NETWORK WITH LONG-RANGE SEMANTIC DEPENDENCY 2491
Adji B. Dieng, Chong Wang, Jianfeng Gao, John Paisley

ON THE QUANTITATIVE ANALYSIS OF DECODER-BASED GENERATIVE MODELS 2504
Yuhuai Wu, Yuri Burda, Ruslan Salakhutdinov, Roger Grosse

OFFLINE BILINGUAL WORD VECTORS, ORTHOGONAL TRANSFORMATIONS AND THE INVERTED SOFTMAX 2521
Samuel L. Smith, David H. P. Turban, Steven Hamblin, Nils Y. Hammerla

ATTEND, ADAPT AND TRANSFER: ATTENTIVE DEEP ARCHITECTURE FOR ADAPTIVE TRANSFER FROM MULTIPLE SOURCES IN THE SAME DOMAIN 2531
Janarthanan Rajendran, Aravind Lakshminarayanan, Mitesh M. Khapra, Prasanna P, Balaraman Ravindran

LEARNING THROUGH DIALOGUE INTERACTIONS BY ASKING QUESTIONS 2549
Jiwei Li, Alexander H. Miller, Sumit Chopra, Marc'Aurelio Ranzato, Jason Weston

PALEO: A PERFORMANCE MODEL FOR DEEP NEURAL NETWORKS 2565
Hang Qi, Evan R. Sparks, Ameet Talwalkar

A LEARNED REPRESENTATION FOR ARTISTIC STYLE 2575
Vincent Dumoulin, Jonathon Shlens, Manjunath Kudlur

VISUALIZING DEEP NEURAL NETWORK DECISIONS: PREDICTION DIFFERENCE ANALYSIS 2601
Luisa M Zintgraf, Taco S Cohen, Tameem Adel, Max Welling

TYING WORD VECTORS AND WORD CLASSIFIERS: A LOSS FRAMEWORK FOR LANGUAGE MODELING 2613
Hakan Inan, Khashayar Khosravi, Richard Socher

DIALOGUE LEARNING WITH HUMAN-IN-THE-LOOP 2626
Jiwei Li, Alexander H. Miller, Sumit Chopra, Marc'Aurelio Ranzato, Jason Weston

BIDIRECTIONAL ATTENTION FLOW FOR MACHINE COMPREHENSION 2649
Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, Hannaneh Hajishirzi

LEARNING TO COMPOSE WORDS INTO SENTENCES WITH REINFORCEMENT LEARNING..... 2662
Dani Yogatama, Phil Blunsom, Chris Dyer, Edward Grefenstette, Wang Ling

NONPARAMETRIC NEURAL NETWORKS 2672
George Philipp, Jaime G. Carbonell

TEMPORAL ENSEMBLING FOR SEMI-SUPERVISED LEARNING 2700
Samuli Laine, Timo Aila

SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS..... 2713
Thomas N. Kipf, Max Welling

TRUSTING SVM FOR PIECEWISE LINEAR CNNS 2727
Leonard Berrada, Andrew Zisserman, M. Pawan Kumar

LIE-ACCESS NEURAL TURING MACHINES..... 2750
Greg Yang, Alexander Rush

ADVERSARIAL TRAINING METHODS FOR SEMI-SUPERVISED TEXT CLASSIFICATION..... 2763
Takeru Miyato, Andrew M. Dai, Ian Goodfellow

COMBINING POLICY GRADIENT AND Q-LEARNING..... 2774
Brendan O'Donoghue, Remi Munos, Koray Kavukcuoglu, Volodymyr Mnih

ORAL PRESENTATIONS

MAKING NEURAL PROGRAMMING ARCHITECTURES GENERALIZE VIA RECURSION 2789
Jonathon Cai, Richard Shin, Dawn Song

TOWARDS PRINCIPLED METHODS FOR TRAINING GENERATIVE ADVERSARIAL NETWORKS..... 2809
Martin Arjovsky, Leon Bottou

LEARNING END-TO-END GOAL-ORIENTED DIALOG 2826
Antoine Bordes, Y-Lan Boureau, Jason Weston

NEURAL ARCHITECTURE SEARCH WITH REINFORCEMENT LEARNING 2841
Barret Zoph, Quoc Le

AMORTISED MAP INFERENCE FOR IMAGE SUPER-RESOLUTION 2857
Casper Kaae Sønderby, Jose Caballero, Lucas Theis, Wenzhe Shi, Ferenc Huszár

ON LARGE-BATCH TRAINING FOR DEEP LEARNING: GENERALIZATION GAP AND SHARP MINIMA..... 2874
Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, Ping Tak Peter Tang

SEMI-SUPERVISED KNOWLEDGE TRANSFER FOR DEEP LEARNING FROM PRIVATE TRAINING DATA 2890
Nicolas Papernot, Martín Abadi, Úlfar Erlingsson, Ian Goodfellow, Kunal Talwar

LEARNING GRAPHICAL STATE TRANSITIONS 2906
Daniel D. Johnson

MULTI-AGENT COOPERATION AND THE EMERGENCE OF (NATURAL) LANGUAGE 2925
Angeliki Lazaridou, Alexander Peysakhovich, Marco Baroni

OPTIMIZATION AS A MODEL FOR FEW-SHOT LEARNING 2936
Sachin Ravi, Hugo Larochelle

LEARNING TO ACT BY PREDICTING THE FUTURE 2947
Alexey Dosovitskiy, Vladlen Koltun

END-TO-END OPTIMIZED IMAGE COMPRESSION 2961
Johannes Ballé, Valero Laparra, Eero P. Simoncelli

Q-PROP: SAMPLE-EFFICIENT POLICY GRADIENT WITH AN OFF-POLICY CRITIC 2988
Shixiang Gu, Timothy Lillicrap, Zoubin Ghahramani, Richard E. Turner, Sergey Levine

UNDERSTANDING DEEP LEARNING REQUIRES RETHINKING GENERALIZATION 3001
Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, Oriol Vinyals

REINFORCEMENT LEARNING WITH UNSUPERVISED AUXILIARY TASKS 3016
Max Jaderberg, Volodymyr Mnih, Wojciech Marian Czarnecki, Tom Schaul, Joel Z Leibo, David Silver, Koray Kavukcuoglu

Author Index