

1, TITLE: Convolutional Neural Networks on Non-uniform Geometrical Signals Using Euclidean Spectral Transformation  
<https://openreview.net/forum?id=B1G5ViAqFm>  
AUTHORS: Chiyu Max Jiang, Dequan Wang, Jingwei Huang, Philip Marcus, Matthias Niessner  
TL;DR: We use non-Euclidean Fourier Transformation of shapes defined by a simplicial complex for deep learning, achieving significantly better results than point-based sampling techniques used in current 3D learning literature.  
HIGHLIGHT: To this end, we develop mathematical formulations for Non-Uniform Fourier Transforms (NUFT) to directly, and optimally, sample nonuniform data signals of different topologies defined on a simplex mesh into the spectral domain with no spatial sampling error.

2, TITLE: Augmented Cyclic Adversarial Learning for Low Resource Domain Adaptation  
<https://openreview.net/forum?id=B1G9doA9F7>  
AUTHORS: Ehsan Hosseini-Asl, Yingbo Zhou, Caiming Xiong, Richard Socher  
TL;DR: A new cyclic adversarial learning augmented with auxiliary task model which improves domain adaptation performance in low resource supervised and unsupervised situations  
HIGHLIGHT: In this paper, we propose an augmented cyclic adversarial learning model that enforces the cycle-consistency constraint via an external task specific model, which encourages the preservation of task-relevant content as opposed to exact reconstruction.

3, TITLE: Variance Networks: When Expectation Does Not Meet Your Expectations  
<https://openreview.net/forum?id=B1GAUs0cKQ>  
AUTHORS: Kirill Neklyudov, Dmitry Molchanov, Arsenii Ashukha, Dmitry Vetrov  
TL;DR: It is possible to learn a zero-centered Gaussian distribution over the weights of a neural network by learning only variances, and it works surprisingly well.  
HIGHLIGHT: In this paper, we introduce variance layers, a different kind of stochastic layers.

4, TITLE: Initialized Equilibrium Propagation for Backprop-Free Training  
<https://openreview.net/forum?id=B1GMDsR5tm>  
AUTHORS: Peter O'Connor, Efstratios Gavves, Max Welling  
TL;DR: We train a feedforward network without backprop by using an energy-based model to provide local targets  
HIGHLIGHT: In response to this problem, we propose Initialized Equilibrium Propagation, which trains a feedforward network to initialize the iterative inference procedure for Equilibrium propagation.

5, TITLE: Explaining Image Classifiers by Counterfactual Generation  
<https://openreview.net/forum?id=B1MXz20cYQ>  
AUTHORS: Chun-Hao Chang, Elliot Creager, Anna Goldenberg, David Duvenaud  
TL;DR: We compute saliency by using a strong generative model to efficiently marginalize over plausible alternative inputs, revealing concentrated pixel areas that preserve label information.  
HIGHLIGHT: Explaining Image Classifiers by Counterfactual Generation.

6, TITLE: SNIP: SINGLE-SHOT NETWORK PRUNING BASED ON CONNECTION SENSITIVITY  
<https://openreview.net/forum?id=B1VZqjAcYX>  
AUTHORS: Namhoon Lee, Thalaiyasingam Ajanthan, Philip Torr  
TL;DR: We present a new approach, SNIP, that is simple, versatile and interpretable; it prunes irrelevant connections for a given task at single-shot prior to training and is applicable to a variety of neural network models without modifications.  
HIGHLIGHT: In this work, we present a new approach that prunes a given network once at initialization prior to training.

7, TITLE: Diagnosing and Enhancing VAE Models  
<https://openreview.net/forum?id=B1e0X3C9tQ>  
AUTHORS: Bin Dai, David Wipf  
TL;DR: We closely analyze the VAE objective function and draw novel conclusions that lead to simple enhancements.  
HIGHLIGHT: In this regard, we rigorously analyze the VAE objective, differentiating situations where this belief is and is not actually true.

8, TITLE: Disjoint Mapping Network for Cross-modal Matching of Voices and Faces  
<https://openreview.net/forum?id=B1exrnCcF7>  
AUTHORS: Yandong Wen, Mahmoud Al Ismail, Weiyang Liu, Bhiksha Raj, Rita Singh  
TL;DR: We closely analyze the VAE objective function and draw novel conclusions that lead to simple enhancements.  
HIGHLIGHT: We propose a novel framework, called Disjoint Mapping Network (DIMNet), for cross-modal biometric matching, in particular of voices and faces.

9, TITLE: Automatically Composing Representation Transformations as a Means for Generalization  
<https://openreview.net/forum?id=B1ffQnRcKX>  
AUTHORS: Michael Chang, Abhishek Gupta, Sergey Levine, Thomas L. Griffiths  
TL;DR: We explore the problem of compositional generalization and propose a means for endowing neural network architectures with the ability to compose themselves to solve these problems.  
HIGHLIGHT: As a first step for tackling compositional generalization, we introduce the compositional recursive learner, a domain-general framework for learning algorithmic procedures for composing representation transformations, producing a learner that reasons about what computation to execute by making analogies to previously seen problems.  
We propose the compositional generalization problem for measuring how readily old knowledge can be reused and hence built upon.

10, TITLE: Visual Reasoning by Progressive Module Networks  
<https://openreview.net/forum?id=B1fpDsAqt7>  
AUTHORS: Seung Wook Kim, Makarand Tapaswi, Sanja Fidler  
TL;DR: We explore the problem of compositional generalization and propose a means for endowing neural network architectures with the ability to compose themselves to solve these problems.  
HIGHLIGHT: We propose to represent a solver for each task as a neural module that calls existing modules (solvers for simpler tasks) in a functional program-like manner.

11, TITLE: Bayesian Deep Convolutional Networks with Many Channels are Gaussian Processes  
<https://openreview.net/forum?id=B1g30j0qF7>  
AUTHORS: Roman Novak, Lechao Xiao, Yasaman Bahri, Jaehoon Lee, Greg Yang, Jiri Hron, Daniel A. Abolafia, Jeffrey Pennington, Jascha Sohl-dickstein  
TL;DR: Finite-width SGD trained CNNs vs. infinitely wide fully Bayesian CNNs. Who wins?  
HIGHLIGHT: In this work, we derive an analogous equivalence for multi-layer convolutional neural networks (CNNs) both with and without pooling layers, and achieve state of the art results on CIFAR10 for GPs without trainable kernels.

12, TITLE: Learning to Learn without Forgetting by Maximizing Transfer and Minimizing Interference  
<https://openreview.net/forum?id=B1gTShAct7>

AUTHORS: Matthew Riemer, Ignacio Cases, Robert Ajemian, Miao Liu, Irina Rish, Yuhai Tu, and Gerald Tesauro

TL;DR: Finite-width SGD trained CNNs vs. infinitely wide fully Bayesian CNNs. Who wins?

HIGHLIGHT: In this work we propose a new conceptualization of the continual learning problem in terms of a temporally symmetric trade-off between transfer and interference that can be optimized by enforcing gradient alignment across examples.

13, TITLE: Sparse Dictionary Learning by Dynamical Neural Networks

<https://openreview.net/forum?id=B1gstsCqt7>

AUTHORS: Tsung-Han Lin, Ping Tak Peter Tang

TL;DR: Finite-width SGD trained CNNs vs. infinitely wide fully Bayesian CNNs. Who wins?

HIGHLIGHT: Using spiking neurons to construct our dynamical network, we present a learning process, its rigorous mathematical analysis, and numerical results on several dictionary learning problems.

14, TITLE: Eidetic 3D LSTM: A Model for Video Prediction and Beyond

<https://openreview.net/forum?id=B1IKS2AqtX>

AUTHORS: Yunbo Wang, Lu Jiang, Ming-Hsuan Yang, Li-Jia Li, Mingsheng Long, Li Fei-Fei

TL;DR: Finite-width SGD trained CNNs vs. infinitely wide fully Bayesian CNNs. Who wins?

HIGHLIGHT: We present a new model, Eidetic 3D LSTM (E3D-LSTM), that integrates 3D convolutions into RNNs.

15, TITLE: ALISTA: Analytic Weights Are As Good As Learned Weights in LISTA

<https://openreview.net/forum?id=B1lnzn0ctQ>

AUTHORS: Jialin Liu, Xiaohan Chen, Zhangyang Wang, Wotao Yin

TL;DR: Finite-width SGD trained CNNs vs. infinitely wide fully Bayesian CNNs. Who wins?

HIGHLIGHT: In this work, we propose Analytic LISTA (ALISTA), where the weight matrix in LISTA is computed as the solution to a data-free optimization problem, leaving only the stepsize and threshold parameters to data-driven learning.

16, TITLE: Three Mechanisms of Weight Decay Regularization

<https://openreview.net/forum?id=B1lz-3Rct7>

AUTHORS: Guodong Zhang, Chaoqi Wang, Bowen Xu, Roger Grosse

TL;DR: We investigate weight decay regularization for different optimizers and identify three distinct mechanisms by which weight decay improves generalization.

HIGHLIGHT: We identify three distinct mechanisms by which weight decay exerts a regularization effect, depending on the particular optimization algorithm and architecture: (1) increasing the effective learning rate, (2) approximately regularizing the input-output Jacobian norm, and (3) reducing the effective damping coefficient for second-order optimization. Our results provide insight into how to improve the regularization of neural networks.

17, TITLE: Learning Multimodal Graph-to-Graph Translation for Molecule Optimization

<https://openreview.net/forum?id=B1xJAsA5F7>

AUTHORS: Wengong Jin, Kevin Yang, Regina Barzilay, Tommi Jaakkola

TL;DR: We introduce a graph-to-graph encoder-decoder framework for learning diverse graph translations.

HIGHLIGHT: Our primary contributions include a junction tree encoder-decoder for learning diverse graph translations along with a novel adversarial training method for aligning distributions of molecules.

18, TITLE: A Data-Driven and Distributed Approach to Sparse Signal Representation and Recovery

<https://openreview.net/forum?id=B1xVTjCqKQ>

AUTHORS: Ali Mousavi, Gautam Dasarathy, Richard G. Baraniuk  
TL;DR: We use deep learning techniques to solve the sparse signal representation and recovery problem.  
HIGHLIGHT: In this paper, we focus on two challenges which offset the promise of sparse signal representation, sensing, and recovery.

19, TITLE: On the Minimal Supervision for Training Any Binary Classifier from Only Unlabeled Data

<https://openreview.net/forum?id=B1xWcj0qYm>

AUTHORS: Nan Lu, Gang Niu, Aditya Krishna Menon, Masashi Sugiyama

TL;DR: Three class priors are all you need to train deep models from only U data, while any two should not be enough.

HIGHLIGHT: In this paper, we study training arbitrary (from linear to deep) binary classifier from only unlabeled (U) data by ERM.

20, TITLE: Neural Logic Machines

<https://openreview.net/forum?id=B1xY-hRctX>

AUTHORS: Honghua Dong, Jiayuan Mao, Tian Lin, Chong Wang, Lihong Li, Denny Zhou

TL;DR: We propose the Neural Logic Machine (NLM), a neural-symbolic architecture for both inductive learning and logic reasoning.

HIGHLIGHT: We propose the Neural Logic Machine (NLM), a neural-symbolic architecture for both inductive learning and logic reasoning.

21, TITLE: Neural Speed Reading with Structural-Jump-LSTM

<https://openreview.net/forum?id=B1xf9jAqFQ>

AUTHORS: Christian Hansen, Casper Hansen, Stephen Alstrup, Jakob Grue Simonsen, Christina Lioma

TL;DR: We propose a new model for neural speed reading that utilizes the inherent punctuation structure of a text to define effective jumping and skipping behavior.

HIGHLIGHT: We present Structural-Jump-LSTM: the first neural speed reading model to both skip and jump text during inference.

22, TITLE: Rigorous Agent Evaluation: An Adversarial Approach to Uncover Catastrophic Failures

<https://openreview.net/forum?id=B1xhQhRcK7>

AUTHORS: Jonathan Uesato\*, Ananya Kumar\*, Csaba Szepesvari\*, Tom Erez, Avraham Ruderman, Keith Anderson, Krishnamurthy (Dj) Dvijotham, Nicolas Heess, Pushmeet Kohli

TL;DR: We show that rare but catastrophic failures may be missed entirely by random testing, which poses issues for safe deployment. Our proposed approach for adversarial testing fixes this.

HIGHLIGHT: To solve this we propose a continuation approach that learns failure modes in related but less robust agents.

23, TITLE: Woulda, Coulda, Shoulda: Counterfactually-Guided Policy Search

<https://openreview.net/forum?id=BJG0voC9YQ>

AUTHORS: Lars Buesing, Theophane Weber, Yori Zwols, Nicolas Heess, Sebastien Racaniere, Arthur Guez, Jean-Baptiste Lespiau

TL;DR: We show that rare but catastrophic failures may be missed entirely by random testing, which poses issues for safe deployment. Our proposed approach for adversarial testing fixes this.

HIGHLIGHT: Based on this, we propose the Counterfactually-Guided Policy Search (CF-GPS) algorithm for learning policies in POMDPs from off-policy experience.

24, TITLE: signSGD via Zeroth-Order Oracle

<https://openreview.net/forum?id=BJe-DsC5Fm>

AUTHORS: Sijia Liu, Pin-Yu Chen, Xiangyi Chen, Mingyi Hong

TL;DR: We design and analyze a new zeroth-order stochastic optimization algorithm, ZO-signSGD, and demonstrate its connection and application to black-box adversarial attacks in robust deep learning

HIGHLIGHT: In this paper, we design and analyze a new zeroth-order (ZO) stochastic optimization algorithm, ZO-signSGD, which enjoys dual advantages of gradient-free operations and signSGD.

25, TITLE: Preventing Posterior Collapse with delta-VAEs

<https://openreview.net/forum?id=BJe0Gn0cY7>

AUTHORS: Ali Razavi, Aaron van den Oord, Ben Poole, Oriol Vinyals

TL;DR: Avoid posterior collapse by lower bounding the rate.

HIGHLIGHT: Due to the phenomenon of ?posterior collapse,? current latent variable generative models pose a challenging design choice that either weakens the capacity of the decoder or requires altering the training objective.

26, TITLE: Algorithmic Framework for Model-based Deep Reinforcement Learning with Theoretical Guarantees

<https://openreview.net/forum?id=BJe1E2R5KX>

AUTHORS: Yuping Luo, Huazhe Xu, Yuanzhi Li, Yuandong Tian, Trevor Darrell, Tengyu Ma

TL;DR: We design model-based reinforcement learning algorithms with theoretical guarantees and achieve state-of-the-art results on Mujoco benchmark tasks when one million or fewer samples are permitted.

HIGHLIGHT: This paper introduces a novel algorithmic framework for designing and analyzing model-based RL algorithms with theoretical guarantees.

27, TITLE: Knowledge Flow: Improve Upon Your Teachers

<https://openreview.net/forum?id=BJeOioA9Y7>

AUTHORS: Iou-Jen Liu, Jian Peng, Alexander Schwing

TL;DR: ?Knowledge Flow? trains a deep net (student) by injecting information from multiple nets (teachers). The student is independent upon training and performs very well on learned tasks irrespective of the setting (reinforcement or supervised learning).

HIGHLIGHT: To address this issue, in this paper, we develop knowledge flow which moves ?knowledge? from multiple deep nets, referred to as teachers, to a new deep net model, called the student.

28, TITLE: Directed-Info GAIL: Learning Hierarchical Policies from Unsegmented Demonstrations using Directed Information

<https://openreview.net/forum?id=BJeWUs05KQ>

AUTHORS: Mohit Sharma, Arjun Sharma, Nicholas Rhinehart, Kris M. Kitani

TL;DR: Learning Hierarchical Policies from Unsegmented Demonstrations using Directed Information

HIGHLIGHT: We propose a new algorithm based on the generative adversarial imitation learning framework which automatically learns sub-task policies from unsegmented demonstrations.

29, TITLE: A Max-Affine Spline Perspective of Recurrent Neural Networks

<https://openreview.net/forum?id=BJej72AqF7>

AUTHORS: Zichao Wang, Randall Balestriero, Richard Baraniuk

TL;DR: We provide new insights and interpretations of RNNs from a max-affine spline operators perspective.

HIGHLIGHT: The resulting representation provides several new perspectives for analyzing RNNs, three of which we study in this paper.

- 30, TITLE: Learning to Navigate the Web  
<https://openreview.net/forum?id=BJemQ209FQ>  
AUTHORS: Izzeddin Gur, Ulrich Rueckert, Aleksandra Faust, Dilek Hakkani-Tur  
TL;DR: We train reinforcement learning policies using reward augmentation, curriculum learning, and meta-learning to successfully navigate web pages.  
HIGHLIGHT: We approach the aforementioned problems from a different perspective and propose guided RL approaches that can generate unbounded amount of experience for an agent to learn from.
- 31, TITLE: Training for Faster Adversarial Robustness Verification via Inducing ReLU Stability  
<https://openreview.net/forum?id=BJfIVjAcKm>  
AUTHORS: Kai Y. Xiao, Vincent Tjeng, Nur Muhammad (Mahi) Shafiullah, Aleksander Madry  
TL;DR: We develop methods to train deep neural models that are both robust to adversarial perturbations and whose robustness is significantly easier to verify.  
HIGHLIGHT: Specifically, we aim to train deep neural networks that not only are robust to adversarial perturbations but also whose robustness can be verified more easily.
- 32, TITLE: Learning to Learn with Conditional Class Dependencies  
<https://openreview.net/forum?id=BJfOXnActQ>  
AUTHORS: Xiang Jiang, Mohammad Havaei, Farshid Varno, Gabriel Chartrand, Nicolas Chapados, Stan Matwin  
TL;DR: CAML is an instance of MAML with conditional class dependencies.  
HIGHLIGHT: We propose a meta-learning framework, Conditional class-Aware Meta-Learning (CAML), that conditionally transforms feature representations based on a metric space that is trained to capture inter-class dependencies.
- 33, TITLE: Hierarchical Visuomotor Control of Humanoids  
<https://openreview.net/forum?id=BJfYvo09Y7>  
AUTHORS: Josh Merel, Arun Ahuja, Vu Pham, Saran Tunyasuvunakool, Siqi Liu, Dhruva Tirumala, Nicolas Heess, Greg Wayne  
TL;DR: Solve tasks involving vision-guided humanoid locomotion, reusing locomotion behavior from motion capture data.  
HIGHLIGHT: In this work, we partly factor this problem into low-level motor control from proprioception and high-level coordination of the low-level skills informed by vision.
- 34, TITLE: Unsupervised Adversarial Image Reconstruction  
<https://openreview.net/forum?id=BJg4Z3RqF7>  
AUTHORS: Arthur Pajot, Emmanuel de Bezenac, Patrick Gallinari  
TL;DR: Solve tasks involving vision-guided humanoid locomotion, reusing locomotion behavior from motion capture data.  
HIGHLIGHT: We cast the problem as finding the  $\text{maximum a posteriori}$  estimate of the signal given each measurement, and propose a general framework for the reconstruction problem.
- 35, TITLE: Max-MIG: an Information Theoretic Approach for Joint Learning from Crowds  
<https://openreview.net/forum?id=BJg9DoR9t7>  
AUTHORS: Peng Cao, Yilun Xu, Yuqing Kong, Yizhou Wang  
TL;DR: Solve tasks involving vision-guided humanoid locomotion, reusing locomotion behavior from motion capture data.  
HIGHLIGHT: We propose an information theoretic approach, Max-MIG, for joint learning from crowds, with a common assumption: the crowdsourced labels and the data are independent conditioning on the ground truth.

- 36, TITLE: AutoLoss: Learning Discrete Schedule for Alternate Optimization  
<https://openreview.net/forum?id=BJgK6iA5KX>  
AUTHORS: Haowen Xu, Hao Zhang, Zhiting Hu, Xiaodan Liang, Ruslan Salakhutdinov, Eric Xing  
TL;DR: We propose a unified formulation for iterative alternate optimization and develop AutoLoss, a framework to automatically learn and generate optimization schedules.  
HIGHLIGHT: In this paper, we present AutoLoss, a meta-learning framework that automatically learns and determines the optimization schedule.
- 37, TITLE: Learning what and where to attend  
<https://openreview.net/forum?id=BJgLg3R9KQ>  
AUTHORS: Drew Linsley, Dan Shiebler, Sven Eberhardt, Thomas Serre  
TL;DR: A large-scale dataset for training attention models for object recognition leads to more accurate, interpretable, and human-like object recognition.  
HIGHLIGHT: Here, we demonstrate the benefit of using stronger supervisory signals by teaching DCNs to attend to image regions that humans deem important for object recognition.
- 38, TITLE: ROBUST ESTIMATION VIA GENERATIVE ADVERSARIAL NETWORKS  
<https://openreview.net/forum?id=BJgRDjR9tQ>  
AUTHORS: Chao GAO, jiyi LIU, Yuan YAO, Weizhi ZHU  
TL;DR: GANs are shown to provide us a new effective robust mean estimate against agnostic contaminations with both statistical optimality and practical tractability.  
HIGHLIGHT: In this paper, we establish an intriguing connection between f-GANs and various depth functions through the lens of f-Learning.
- 39, TITLE: INVASE: Instance-wise Variable Selection using Neural Networks  
[https://openreview.net/forum?id=BJg\\_roAcK7](https://openreview.net/forum?id=BJg_roAcK7)  
AUTHORS: Jinsung Yoon, James Jordon, Mihaela van der Schaar  
TL;DR: GANs are shown to provide us a new effective robust mean estimate against agnostic contaminations with both statistical optimality and practical tractability.  
HIGHLIGHT: In this paper, we propose a new instance-wise feature selection method, which we term INVASE.
- 40, TITLE: Meta-Learning with Latent Embedding Optimization  
<https://openreview.net/forum?id=BJgklhAcK7>  
AUTHORS: Andrei A. Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, Raia Hadsell  
TL;DR: Latent Embedding Optimization (LEO) is a novel gradient-based meta-learner with state-of-the-art performance on the challenging 5-way 1-shot and 5-shot miniImageNet and tieredImageNet classification tasks.  
HIGHLIGHT: We show that it is possible to bypass these limitations by learning a data-dependent latent generative representation of model parameters, and performing gradient-based meta-learning in this low-dimensional latent space.
- 41, TITLE: Non-vacuous Generalization Bounds at the ImageNet Scale: a PAC-Bayesian Compression Approach  
<https://openreview.net/forum?id=BJgqqsAct7>  
AUTHORS: Wenda Zhou, Victor Veitch, Morgane Austern, Ryan P. Adams, Peter Orbanz  
TL;DR: We obtain non-vacuous generalization bounds on ImageNet-scale deep neural networks by combining an original PAC-Bayes bound and an off-the-shelf neural network compression method.  
HIGHLIGHT: The purpose of this paper is to connect these two empirical observations.

42, TITLE: Learning to Represent Edits  
<https://openreview.net/forum?id=BJl6AjC5F7>  
AUTHORS: Pengcheng Yin, Graham Neubig, Miltiadis Allamanis, Marc Brockschmidt, Alexander L. Gaunt  
TL;DR: We obtain non-vacuous generalization bounds on ImageNet-scale deep neural networks by combining an original PAC-Bayes bound and an off-the-shelf neural network compression method.  
HIGHLIGHT: We introduce the problem of learning distributed representations of edits.

43, TITLE: Neural Probabilistic Motor Primitives for Humanoid Control  
<https://openreview.net/forum?id=BJl6TjRcY7>  
AUTHORS: Josh Merel, Leonard Hasenclever, Alexandre Galashov, Arun Ahuja, Vu Pham, Greg Wayne, Yee Whye Teh, Nicolas Heess  
TL;DR: Neural Probabilistic Motor Primitives compress motion capture tracking policies into one flexible model capable of one-shot imitation and reuse as a low-level controller.  
HIGHLIGHT: To do this, we propose a motor architecture that has the general structure of an inverse model with a latent-variable bottleneck.

44, TITLE: Differentiable Perturb-and-Parse: Semi-Supervised Parsing with a Structured Variational Autoencoder  
<https://openreview.net/forum?id=BJlgNh0qKQ>  
AUTHORS: Caio Corro, Ivan Titov  
TL;DR: Differentiable dynamic programming over perturbed input weights with application to semi-supervised VAE  
HIGHLIGHT: To this end, we propose a novel latent-variable generative model for semi-supervised syntactic dependency parsing.

45, TITLE: Janossy Pooling: Learning Deep Permutation-Invariant Functions for Variable-Size Inputs  
<https://openreview.net/forum?id=BJluy2RcFm>  
AUTHORS: Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak Rao, Bruno Ribeiro  
TL;DR: We propose Janossy pooling, a method for learning deep permutation invariant functions designed to exploit relationships within the input sequence and tractable inference strategies such as a stochastic optimization procedure we call piSGD  
HIGHLIGHT: We consider a simple and overarching representation for permutation-invariant functions of sequences (or set functions).

46, TITLE: An Empirical Study of Example Forgetting during Deep Neural Network Learning  
<https://openreview.net/forum?id=BJlxm30cKm>  
AUTHORS: Mariya Toneva\*, Alessandro Sordani\*, Remi Tachet des Combes\*, Adam Trischler, Yoshua Bengio, Geoffrey J. Gordon  
TL;DR: We show that catastrophic forgetting occurs within what is considered to be a single task and find that examples that are not prone to forgetting can be removed from the training set without loss of generalization.  
HIGHLIGHT: Our goal is to understand whether a related phenomenon occurs when data does not undergo a clear distributional shift.

47, TITLE: RNNs implicitly implement tensor-product representations  
<https://openreview.net/forum?id=BJx0sjC5FX>  
AUTHORS: R. Thomas McCoy, Tal Linzen, Ewan Dunbar, Paul Smolensky

TL;DR: RNNs implicitly implement tensor-product representations, a principled and interpretable method for representing symbolic structures in continuous space.

HIGHLIGHT: To test this hypothesis, we introduce Tensor Product Decomposition Networks (TPDNs), which use TPRs to approximate existing vector representations.

48, TITLE: Learning To Solve Circuit-SAT: An Unsupervised Differentiable Approach

<https://openreview.net/forum?id=BJxgz2R9t7>

AUTHORS: Saeed Amizadeh, Sergiy Matussevych, Markus Weimer

TL;DR: We propose a neural framework that can learn to solve the Circuit Satisfiability problem from (unlabeled) circuit instances.

HIGHLIGHT: In this paper, we propose a neural framework that can learn to solve the Circuit Satisfiability problem.

49, TITLE: Dynamic Channel Pruning: Feature Boosting and Suppression

<https://openreview.net/forum?id=BJxh2j0qYm>

AUTHORS: Xitong Gao, Yiren Zhao, Lukasz Dudziak, Robert Mullins, Cheng-zhong Xu

TL;DR: We make convolutional layers run faster by dynamically boosting and suppressing channels in feature computation.

HIGHLIGHT: In this paper, we reduce this cost by exploiting the fact that the importance of features computed by convolutional layers is highly input-dependent, and propose feature boosting and suppression (FBS), a new method to predictively amplify salient convolutional channels and skip unimportant ones at run-time.

50, TITLE: signSGD with Majority Vote is Communication Efficient and Fault Tolerant

<https://openreview.net/forum?id=BJxhijAcY7>

AUTHORS: Jeremy Bernstein, Jiawei Zhao, Kamyar Azizzadenesheli, Anima Anandkumar

TL;DR: Workers send gradient signs to the server, and the update is decided by majority vote. We show that this algorithm is convergent, communication efficient and fault tolerant, both in theory and in practice.

HIGHLIGHT: We explore a particularly simple algorithm for robust, communication-efficient learning--signSGD.

51, TITLE: Bounce and Learn: Modeling Scene Dynamics with Real-World Bounces

<https://openreview.net/forum?id=BJxss0A5KX>

AUTHORS: Senthil Purushwalkam, Abhinav Gupta, Danny Kaufman, Bryan Russell

TL;DR: Workers send gradient signs to the server, and the update is decided by majority vote. We show that this algorithm is convergent, communication efficient and fault tolerant, both in theory and in practice.

HIGHLIGHT: We introduce an approach to model surface properties governing bounces in everyday scenes.

VIM learns to infer physical parameters for locations in a scene given a single still image, while PIM learns to model physical interactions for the prediction task given physical parameters and observed pre-collision 3D trajectories. To achieve our results, we introduce the Bounce Dataset comprising 5K RGB-D videos of bouncing trajectories of a foam ball to probe surfaces of varying shapes and materials in everyday scenes including homes and offices. Our proposed model learns from our collected dataset of real-world bounces and is bootstrapped with additional information from simple physics simulations.

52, TITLE: K for the Price of 1: Parameter-efficient Multi-task and Transfer Learning

<https://openreview.net/forum?id=BJxvEh0cFQ>

AUTHORS: Pramod Kaushik Mudrakarta, Mark Sandler, Andrey Zhmoginov, Andrew Howard

TL;DR: A novel and practically effective method to adapt pretrained neural networks to new tasks by retraining a minimal (e.g., less than 2%) number of parameters

**HIGHLIGHT:** We introduce a novel method that enables parameter-efficient transfer and multi-task learning with deep neural networks.

53, **TITLE:** Towards Metamerism via Foveated Style Transfer  
<https://openreview.net/forum?id=BJzbG20cFQ>  
**AUTHORS:** Arturo Deza, Aditya Jonnalagadda, Miguel P. Eckstein  
**TL;DR:** We introduce a novel feed-forward framework to generate visual metamers  
**HIGHLIGHT:** In this paper, we propose our NeuroFovea metamer model, a foveated generative model that is based on a mixture of peripheral representations and style transfer forward-pass algorithms.

54, **TITLE:** Post Selection Inference with Incomplete Maximum Mean Discrepancy Estimator  
<https://openreview.net/forum?id=BkG5SjR5YQ>  
**AUTHORS:** Makoto Yamada, Denny Wu, Yao-Hung Hubert Tsai, Hirofumi Ohta, Ruslan Salakhutdinov, Ichiro Takeuchi, Kenji Fukumizu  
**TL;DR:** We introduce a novel feed-forward framework to generate visual metamers  
**HIGHLIGHT:** In this paper, we propose a post selection inference (PSI) framework for divergence measure, which can select a set of statistically significant features that discriminate two distributions.

55, **TITLE:** Emergent Coordination Through Competition  
<https://openreview.net/forum?id=BkG8sjR5Km>  
**AUTHORS:** Siqi Liu, Guy Lever, Josh Merel, Saran Tunyasuvunakool, Nicolas Heess, Thore Graepel  
**TL;DR:** We introduce a new MuJoCo soccer environment for continuous multi-agent reinforcement learning research, and show that population-based training of independent reinforcement learners can learn cooperative behaviors  
**HIGHLIGHT:** We study the emergence of cooperative behaviors in reinforcement learning agents by introducing a challenging competitive multi-agent soccer environment with continuous simulated physics.

56, **TITLE:** Prior Convictions: Black-box Adversarial Attacks with Bandits and Priors  
<https://openreview.net/forum?id=BkMiWhR5K7>  
**AUTHORS:** Andrew Ilyas, Logan Engstrom, Aleksander Madry  
**TL;DR:** We present a unifying view on black-box adversarial attacks as a gradient estimation problem, and then present a framework (based on bandits optimization) to integrate priors into gradient estimation, leading to significantly increased performance.  
**HIGHLIGHT:** We introduce a framework that conceptually unifies much of the existing work on black-box attacks, and demonstrate that the current state-of-the-art methods are optimal in a natural sense.

57, **TITLE:** Sample Efficient Imitation Learning for Continuous Control  
<https://openreview.net/forum?id=BkN5UoAqF7>  
**AUTHORS:** Fumihiko Sasaki  
**TL;DR:** In this paper, we proposed a model-free, off-policy IL algorithm for continuous control. Experimental results showed that our algorithm achieves competitive results with GAIL while significantly reducing the environment interactions.  
**HIGHLIGHT:** We believe that IL algorithms could be more applicable to real-world problems if the number of interactions could be reduced. In this paper, we propose a model-free IL algorithm for continuous control.

58, **TITLE:** Generative Code Modeling with Graphs  
<https://openreview.net/forum?id=Bke4KsA5FX>  
**AUTHORS:** Marc Brockschmidt, Miltiadis Allamanis, Alexander L. Gaunt, Oleksandr Polozov  
**TL;DR:** Representing programs as graphs including semantics helps when generating programs

**HIGHLIGHT:** We present a novel model for this problem that uses a graph to represent the intermediate state of the generated output.

59, **TITLE:** Critical Learning Periods in Deep Networks

<https://openreview.net/forum?id=BkeStsCckQ>

**AUTHORS:** Alessandro Achille, Matteo Rovere, Stefano Soatto

**TL;DR:** Sensory deficits in early training phases can lead to irreversible performance loss in both artificial and neuronal networks, suggesting information phenomena as the common cause, and point to the importance of the initial transient and forgetting.

**HIGHLIGHT:** To better understand this phenomenon, we use the Fisher Information of the weights to measure the effective connectivity between layers of a network during training.

60, **TITLE:** CEM-RL: Combining evolutionary and gradient-based methods for policy search

<https://openreview.net/forum?id=BkeU5j0ctQ>

**AUTHORS:** Pourchot, Sigaud

**TL;DR:** We propose a new combination of evolution strategy and deep reinforcement learning which takes the best of both worlds

**HIGHLIGHT:** In this paper, we propose a different combination scheme using the simple cross-entropy method (CEM) and Twin Delayed Deep Deterministic policy gradient (TD3), another off-policy deep RL algorithm which improves over DDPG.

61, **TITLE:** LanczosNet: Multi-Scale Deep Graph Convolutional Networks

<https://openreview.net/forum?id=BkedznAqKQ>

**AUTHORS:** Renjie Liao, Zhizhen Zhao, Raquel Urtasun, Richard Zemel

**TL;DR:** We propose a new combination of evolution strategy and deep reinforcement learning which takes the best of both worlds

**HIGHLIGHT:** We propose Lanczos network (LanczosNet) which uses the Lanczos algorithm to construct low rank approximations of the graph Laplacian for graph convolution. Relying on the tridiagonal decomposition of the Lanczos algorithm, we not only efficiently exploit multi-scale information via fast approximated computation of matrix power but also design learnable spectral filters. Being fully differentiable, LanczosNet facilitates both graph kernel learning as well as learning node embeddings. We show the connection between our LanczosNet and graph based manifold learning, especially diffusion maps. We benchmark our model against the recent deep graph networks on citation datasets and QM8 quantum chemistry dataset. Experimental results show that our model achieves the state-of-the-art performance in most tasks.

62, **TITLE:** Excessive Invariance Causes Adversarial Vulnerability

<https://openreview.net/forum?id=BkfbpsAcF7>

**AUTHORS:** Joern-Henrik Jacobsen, Jens Behrmann, Richard Zemel, Matthias Bethge

**TL;DR:** We show deep networks are not only too sensitive to task-irrelevant changes of their input, but also too invariant to a wide range of task-relevant changes, thus making vast regions in input space vulnerable to adversarial attacks.

**HIGHLIGHT:** One core idea of adversarial example research is to reveal neural network errors under such distribution shifts.

63, **TITLE:** Hindsight policy gradients

<https://openreview.net/forum?id=Bkg2viA5FQ>

**AUTHORS:** Paulo Rauber, Avinash Ummadisingu, Filipe Mutz, Jürgen Schmidhuber

**TL;DR:** We introduce the capacity to exploit information about the degree to which an arbitrary goal has been achieved while another goal was intended to policy gradient methods.

**HIGHLIGHT:** In this paper, we demonstrate how hindsight can be introduced to policy gradient methods, generalizing this idea to a broad class of successful algorithms.

64, **TITLE:** Adaptive Gradient Methods with Dynamic Bound of Learning Rate  
<https://openreview.net/forum?id=Bkg3g2R9FX>  
**AUTHORS:** Liangchen Luo, Yuanhao Xiong, Yan Liu, Xu Sun  
**TL;DR:** Novel variants of optimization methods that combine the benefits of both adaptive and non-adaptive methods.  
**HIGHLIGHT:** Adaptive optimization methods such as AdaGrad, RMSprop and Adam have been proposed to achieve a rapid training process with an element-wise scaling term on learning rates.

65, **TITLE:** Decoupled Weight Decay Regularization  
<https://openreview.net/forum?id=Bkg6RiCqY7>  
**AUTHORS:** Ilya Loshchilov, Frank Hutter  
**TL;DR:** Novel variants of optimization methods that combine the benefits of both adaptive and non-adaptive methods.  
**HIGHLIGHT:**  $L_2$  regularization and weight decay regularization are equivalent for standard stochastic gradient descent (when rescaled by the learning rate), but as we demonstrate this is *not* the case for adaptive gradient algorithms, such as Adam.

66, **TITLE:** Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile  
<https://openreview.net/forum?id=Bkg8jjC9KQ>  
**AUTHORS:** Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, Georgios Piliouras  
**TL;DR:** We show how the inclusion of an extra-gradient step in first-order GAN training methods can improve stability and lead to improved convergence results.  
**HIGHLIGHT:** To make piecemeal progress along these lines, we analyze the behavior of mirror descent (MD) in a class of non-monotone problems whose solutions coincide with those of a naturally associated variational inequality ? a property which we call coherence.

67, **TITLE:** DialogWAE: Multimodal Response Generation with Conditional Wasserstein Auto-Encoder  
<https://openreview.net/forum?id=BkgBvsC9FQ>  
**AUTHORS:** Xiaodong Gu, Kyunghyun Cho, Jung-Woo Ha, Sunghun Kim  
**TL;DR:** We show how the inclusion of an extra-gradient step in first-order GAN training methods can improve stability and lead to improved convergence results.  
**HIGHLIGHT:** In this paper, we propose DialogWAE, a conditional Wasserstein autoencoder (WAE) specially designed for dialogue modeling.

68, **TITLE:** No Training Required: Exploring Random Encoders for Sentence Classification  
<https://openreview.net/forum?id=BkgPajAcY7>  
**AUTHORS:** John Wieting, Douwe Kiela  
**TL;DR:** We show how the inclusion of an extra-gradient step in first-order GAN training methods can improve stability and lead to improved convergence results.  
**HIGHLIGHT:** Our aim is to put sentence embeddings on more solid footing by 1) looking at how much modern sentence embeddings gain over random methods--as it turns out, surprisingly little; and by 2) providing the field with more appropriate baselines going forward--which are, as it turns out, quite strong.

69, **TITLE:** Neural Graph Evolution: Towards Efficient Automatic Robot Design  
<https://openreview.net/forum?id=BkgWHnR5tm>

AUTHORS: Tingwu Wang, Yuhao Zhou, Sanja Fidler, Jimmy Ba  
TL;DR: Automatic robotic design search with graph neural networks  
HIGHLIGHT: We propose Neural Graph Evolution (NGE), which performs selection on current candidates and evolves new ones iteratively.

70, TITLE: Function Space Particle Optimization for Bayesian Neural Networks  
<https://openreview.net/forum?id=BkgtDsCcKQ>  
AUTHORS: Ziyu Wang, Tongzheng Ren, Jun Zhu, Bo Zhang  
TL;DR: Automatic robotic design search with graph neural networks  
HIGHLIGHT: In this paper, we propose to solve this issue by performing particle optimization directly in the space of regression functions.

71, TITLE: Structured Adversarial Attack: Towards General Implementation and Better Interpretability  
<https://openreview.net/forum?id=BkgzniCqY7>  
AUTHORS: Kaidi Xu, Sijia Liu, Pu Zhao, Pin-Yu Chen, Huan Zhang, Quanfu Fan, Deniz Erdogmus, Yanzhi Wang, Xue Lin  
TL;DR: Automatic robotic design search with graph neural networks  
HIGHLIGHT: This work develops a more general attack model, i.e., the structured attack (StrAttack), which explores group sparsity in adversarial perturbation by sliding a mask through images aiming for extracting key spatial structures.

72, TITLE: Spherical CNNs on Unstructured Grids  
<https://openreview.net/forum?id=Bkl-43C9FQ>  
AUTHORS: Chiyu Max Jiang, Jingwei Huang, Karthik Kashinath, Prabhat, Philip Marcus, Matthias Niessner  
TL;DR: We present a new CNN kernel for unstructured grids for spherical signals, and show significant accuracy and parameter efficiency gain on tasks such as 3D classification and omnidirectional image segmentation.  
HIGHLIGHT: We present an efficient convolution kernel for Convolutional Neural Networks (CNNs) on unstructured grids using parameterized differential operators while focusing on spherical signals such as panorama images or planetary signals. To this end, we replace conventional convolution kernels with linear combinations of differential operators that are weighted by learnable parameters.

73, TITLE: Optimal Transport Maps For Distribution Preserving Operations on Latent Spaces of Generative Models  
<https://openreview.net/forum?id=BklCusRct7>  
AUTHORS: Eirikur Agustsson, Alexander Sage, Radu Timofte, Luc Van Gool  
TL;DR: We propose a framework for modifying the latent space operations such that the distribution mismatch between the resulting outputs and the prior distribution the generative model was trained on is fully eliminated.  
HIGHLIGHT: In this paper, we propose a framework for modifying the latent space operations such that the distribution mismatch is fully eliminated.

74, TITLE: Deep Lagrangian Networks: Using Physics as Model Prior for Deep Learning  
<https://openreview.net/forum?id=BklHpjCqKm>  
AUTHORS: Michael Lutter, Christian Ritter, Jan Peters  
TL;DR: This paper introduces a physics prior for Deep Learning and applies the resulting network topology for model-based control.  
HIGHLIGHT: As a first example, we propose Deep Lagrangian Networks (DeLaN) as a deep network structure upon which Lagrangian Mechanics have been imposed.

75, TITLE: Accumulation Bit-Width Scaling For Ultra-Low Precision Training Of Deep Networks  
<https://openreview.net/forum?id=BklMjsRqY7>  
AUTHORS: Charbel Sakr, Naigang Wang, Chia-Yu Chen, Jungwook Choi, Ankur Agrawal, Naresh Shanbhag, Kailash Gopalakrishnan  
TL;DR: We present an analytical framework to determine accumulation bit-width requirements in all three deep learning training GEMMs and verify the validity and tightness of our method via benchmarking experiments.  
HIGHLIGHT: We present a statistical approach to analyze the impact of reduced accumulation precision on deep learning training.

76, TITLE: Deep Convolutional Networks as shallow Gaussian Processes  
<https://openreview.net/forum?id=Bklfsi0cKk>  
AUTHORS: Adrià Garriga-Alonso, Carl Edward Rasmussen, Laurence Aitchison  
TL;DR: We show that CNNs and ResNets with appropriate priors on the parameters are Gaussian processes in the limit of infinitely many convolutional filters.  
HIGHLIGHT: We show that the output of a (residual) CNN with an appropriate prior over the weights and biases is a GP in the limit of infinitely many convolutional filters, extending similar results for dense networks.

77, TITLE: Unsupervised Domain Adaptation for Distance Metric Learning  
<https://openreview.net/forum?id=BklhAj09K7>  
AUTHORS: Kihyuk Sohn, Wenling Shang, Xiang Yu, Manmohan Chandraker  
TL;DR: A new theory of unsupervised domain adaptation for distance metric learning and its application to face recognition across diverse ethnicity variations.  
HIGHLIGHT: To handle both within and cross domain verifications, we propose a Feature Transfer Network (FTN) to separate the target feature space from the original source space while aligned with a transformed source space.

78, TITLE: A comprehensive, application-oriented study of catastrophic forgetting in DNNs  
<https://openreview.net/forum?id=BkloRs0qK7>  
AUTHORS: B. Pfister, A. Gepperth  
TL;DR: We check DNN models for catastrophic forgetting using a new evaluation scheme that reflects typical application conditions, with surprising results.  
HIGHLIGHT: We present a large-scale empirical study of catastrophic forgetting (CF) in modern Deep Neural Network (DNN) models that perform sequential (or: incremental) learning. A new experimental protocol is proposed that takes into account typical constraints encountered in application scenarios. As the investigation is empirical, we evaluate CF behavior on the hitherto largest number of visual classification datasets, from each of which we construct a representative number of Sequential Learning Tasks (SLTs) in close alignment to previous works on CF. Our results clearly indicate that there is no model that avoids CF for all investigated datasets and SLTs under application conditions.

79, TITLE: Posterior Attention Models for Sequence to Sequence Learning  
<https://openreview.net/forum?id=BkltNhC9FX>  
AUTHORS: Shiv Shankar, Sunita Sarawagi  
TL;DR: Computing attention based on posterior distribution leads to more meaningful attention and better performance  
HIGHLIGHT: In this paper we show that prevalent attention architectures do not adequately model the dependence among the attention and output tokens across a predicted sequence. We present an alternative architecture called Posterior Attention Models that after a principled factorization of the full joint distribution of the attention and output variables, proposes two major changes.

80, TITLE: Generative Question Answering: Learning to Answer the Whole Question  
<https://openreview.net/forum?id=Bkx0RjA9tX>  
AUTHORS: Mike Lewis, Angela Fan  
TL;DR: Question answering models that model the joint distribution of questions and answers can learn more than discriminative models  
HIGHLIGHT: We introduce generative models of the joint distribution of questions and answers, which are trained to explain the whole question, not just to answer it. Our question answering (QA) model is implemented by learning a prior over answers, and a conditional language model to generate the question given the answer?allowing scalable and interpretable many-hop reasoning as the question is generated word-by-word.

81, TITLE: Diversity and Depth in Per-Example Routing Models  
<https://openreview.net/forum?id=BkxWJnC9tX>  
AUTHORS: Prajit Ramachandran, Quoc V. Le  
TL;DR: Per-example routing models benefit from architectural diversity, but still struggle to scale to a large number of routing decisions.  
HIGHLIGHT: In this work, we address both of these deficiencies.

82, TITLE: Selfless Sequential Learning  
<https://openreview.net/forum?id=Bkxbrn0cYX>  
AUTHORS: Rahaf Aljundi, Marcus Rohrbach, Tinne Tuytelaars  
TL;DR: A regularization strategy for improving the performance of sequential learning  
HIGHLIGHT: In this paper we look at a scenario with fixed model capacity, and postulate that the learning process should not be selfish, i.e. it should account for future tasks to be added and thus leave enough capacity for them.

83, TITLE: M<sup>3</sup>RL: Mind-aware Multi-agent Management Reinforcement Learning  
<https://openreview.net/forum?id=BkzeUiRcY7>  
AUTHORS: Tianmin Shu, Yuandong Tian  
TL;DR: We propose Mind-aware Multi-agent Management Reinforcement Learning (M<sup>3</sup>RL) for training a manager to motivate self-interested workers to achieve optimal collaboration by assigning suitable contracts to them.  
HIGHLIGHT: In this paper, we aim to address this from a different angle.

84, TITLE: The Deep Weight Prior  
<https://openreview.net/forum?id=ByGuynAct7>  
AUTHORS: Andrei Atanov, Arsenii Ashukha, Kirill Struminsky, Dmitriy Vetrov, Max Welling  
TL;DR: The generative model for kernels of convolutional neural networks, that acts as a prior distribution while training on new datasets.  
HIGHLIGHT: In this work, we propose a new type of prior distributions for convolutional neural networks, deep weight prior (DWP), that exploit generative models to encourage a specific structure of trained convolutional filters e.g., spatial correlations of weights.

85, TITLE: Efficient Multi-Objective Neural Architecture Search via Lamarckian Evolution  
<https://openreview.net/forum?id=ByME42AqK7>  
AUTHORS: Thomas Elsken, Jan Hendrik Metzen, Frank Hutter  
TL;DR: We propose a method for efficient Multi-Objective Neural Architecture Search based on Lamarckian inheritance and evolutionary algorithms.

**HIGHLIGHT:** We address the first shortcoming by proposing LEMONADE, an evolutionary algorithm for multi-objective architecture search that allows approximating the Pareto-front of architectures under multiple objectives, such as predictive performance and number of parameters, in a single run of the method.

86, **TITLE:** Quaternion Recurrent Neural Networks

<https://openreview.net/forum?id=ByMHvs0cFQ>

**AUTHORS:** Titouan Parcollet, Mirco Ravanelli, Mohamed Morchid, Georges Linar's, Chiheb Trabelsi, Renato De Mori, Yoshua Bengio

**TL;DR:** We propose a method for efficient Multi-Objective Neural Architecture Search based on Lamarckian inheritance and evolutionary algorithms.

**HIGHLIGHT:** We propose a novel quaternion recurrent neural network (QRNN), alongside with a quaternion long-short term memory neural network (QLSTM), that take into account both the external relations and these internal structural dependencies with the quaternion algebra.

87, **TITLE:** Adversarial Audio Synthesis

<https://openreview.net/forum?id=ByMVTsR5KQ>

**AUTHORS:** Chris Donahue, Julian McAuley, Miller Puckette

**TL;DR:** Learning to synthesize raw waveform audio with GANs

**HIGHLIGHT:** In this paper we introduce WaveGAN, a first attempt at applying GANs to unsupervised synthesis of raw-waveform audio.

88, **TITLE:** Preconditioner on Matrix Lie Group for SGD

<https://openreview.net/forum?id=Bye5SiAqKX>

**AUTHORS:** Xi-Lin Li

**TL;DR:** We propose a new framework for preconditioner learning, derive new forms of preconditioners and learning methods, and reveal the relationship to methods like RMSProp, Adam, Adagrad, ESGD, KFAC, batch normalization, etc.

**HIGHLIGHT:** We study two types of preconditioners and preconditioned stochastic gradient descent (SGD) methods in a unified framework.

89, **TITLE:** Learning to Screen for Fast Softmax Inference on Large Vocabulary Neural Networks

<https://openreview.net/forum?id=ByeMB3Act7>

**AUTHORS:** Patrick Chen, Si Si, Sanjiv Kumar, Yang Li, Cho-Jui Hsieh

**TL;DR:** We propose a new framework for preconditioner learning, derive new forms of preconditioners and learning methods, and reveal the relationship to methods like RMSProp, Adam, Adagrad, ESGD, KFAC, batch normalization, etc.

**HIGHLIGHT:** In this paper, we introduce a novel softmax layer approximation algorithm by exploiting the clustering structure of context vectors.

90, **TITLE:** Adaptive Posterior Learning: few-shot learning with a surprise-based memory module

<https://openreview.net/forum?id=ByeSdsC9Km>

**AUTHORS:** Tiago Ramalho, Marta Garnelo

**TL;DR:** We introduce a model which generalizes quickly from few observations by storing surprising information and attending over the most relevant data at each time point.

**HIGHLIGHT:** In this paper we introduce APL, an algorithm that approximates probability distributions by remembering the most surprising observations it has encountered.

91, **TITLE:** Probabilistic Planning with Sequential Monte Carlo methods

<https://openreview.net/forum?id=ByetGn0cYX>

AUTHORS: Alexandre Piche, Valentin Thomas, Cyril Ibrahim, Yoshua Bengio, Chris Pal  
TL;DR: Leveraging control as inference and Sequential Monte Carlo methods, we proposed a probabilistic planning algorithm.  
HIGHLIGHT: In this work, we propose a novel formulation of planning which views it as a probabilistic inference problem over future optimal trajectories.

92, TITLE: Plan Online, Learn Offline: Efficient Learning and Exploration via Model-Based Control  
<https://openreview.net/forum?id=Byey7n05FQ>  
AUTHORS: Kendall Lowrey, Aravind Rajeswaran, Sham Kakade, Emanuel Todorov, Igor Mordatch  
TL;DR: We propose a framework that incorporates planning for efficient exploration and learning in complex environments.  
HIGHLIGHT: We propose a "plan online and learn offline" framework for the setting where an agent, with an internal model, needs to continually act and learn in the world.

93, TITLE: DHER: Hindsight Experience Replay for Dynamic Goals  
<https://openreview.net/forum?id=Byf5-30qFX>  
AUTHORS: Meng Fang, Cheng Zhou, Bei Shi, Boqing Gong, Jia Xu, Tong Zhang  
TL;DR: We propose a framework that incorporates planning for efficient exploration and learning in complex environments.  
HIGHLIGHT: DHER automatically assembles successful experiences from two relevant failures and can be used to enhance an arbitrary off-policy RL algorithm when the tasks' goals are dynamic.

94, TITLE: FlowQA: Grasping Flow in History for Conversational Machine Comprehension  
<https://openreview.net/forum?id=ByftGnR9KX>  
AUTHORS: Hsin-Yuan Huang, Eunsol Choi, Wen-tau Yih  
TL;DR: We propose the Flow mechanism and an end-to-end architecture, FlowQA, that achieves SotA on two conversational QA datasets and a sequential instruction understanding task.  
HIGHLIGHT: To enable traditional, single-turn models to encode the history comprehensively, we introduce Flow, a mechanism that can incorporate intermediate representations generated during the process of answering previous questions, through an alternating parallel processing structure.

95, TITLE: Learning to Design RNA  
<https://openreview.net/forum?id=ByfyHh05tQ>  
AUTHORS: Frederic Runge, Danny Stoll, Stefan Falkner, Frank Hutter  
TL;DR: We learn to solve the RNA Design problem with reinforcement learning using meta learning and autoML approaches.  
HIGHLIGHT: Here, we propose a new algorithm for the RNA Design problem, dubbed LEARNA.

96, TITLE: Robust Conditional Generative Adversarial Networks  
<https://openreview.net/forum?id=Byg0DsCqYQ>  
AUTHORS: Grigorios G. Chrysos, Jean Kossaifi, Stefanos Zafeiriou  
TL;DR: We introduce a new type of conditional GAN, which aims to leverage structure in the target space of the generator. We augment the generator with a new, unsupervised pathway to learn the target structure.  
HIGHLIGHT: In this work, we introduce a novel conditional GAN model, called RoCGAN, which leverages structure in the target space of the model to address the issue.

97, TITLE: Top-Down Neural Model For Formulae  
<https://openreview.net/forum?id=Byg5QhR5FQ>  
AUTHORS: Karel Chvalovsk?

TL;DR: A top-down approach how to recursively represent propositional formulae by neural networks is presented.

HIGHLIGHT: We present a simple neural model that given a formula and a property tries to answer the question whether the formula has the given property, for example whether a propositional formula is always true.

98, TITLE: Cost-Sensitive Robustness against Adversarial Examples

<https://openreview.net/forum?id=BygANhA9tQ>

AUTHORS: Xiao Zhang, David Evans

TL;DR: A general method for training certified cost-sensitive robust classifier against adversarial perturbations

HIGHLIGHT: We encode the potential harm of each adversarial transformation in a cost matrix, and propose a general objective function to adapt the robust training method of Wong & Kolter (2018) to optimize for cost-sensitive robustness.

99, TITLE: The role of over-parametrization in generalization of neural networks

<https://openreview.net/forum?id=BygfhgAcYX>

AUTHORS: Behnam Neyshabur, Zhiyuan Li, Srinadh Bhojanapalli, Yann LeCun, Nathan Srebro

TL;DR: We suggest a generalization bound that could partly explain the improvement in generalization with over-parametrization.

HIGHLIGHT: In this work we suggest a novel complexity measure based on unit-wise capacities resulting in a tighter generalization bound for two layer ReLU networks.

100, TITLE: Diffusion Scattering Transforms on Graphs

<https://openreview.net/forum?id=BygqBiRcFQ>

AUTHORS: Fernando Gama, Alejandro Ribeiro, Joan Bruna

TL;DR: Stability of scattering transform representations of graph data to deformations of the underlying graph support.

HIGHLIGHT: This stability to deformations can be interpreted as stability with respect to changes in the metric structure of the domain. In this work, we show that scattering transforms can be generalized to non-Euclidean domains using diffusion wavelets, while preserving a notion of stability with respect to metric changes in the domain, measured with diffusion maps.

101, TITLE: Capsule Graph Neural Network

<https://openreview.net/forum?id=Byl8BnRcYm>

AUTHORS: Zhang Xinyi, Lihui Chen

TL;DR: Inspired by CapsNet, we propose a novel architecture for graph embeddings on the basis of node features extracted from GNN.

HIGHLIGHT: Capsule Graph Neural Network.

102, TITLE: Energy-Constrained Compression for Deep Neural Networks via Weighted Sparse Projection and Layer Input Masking

<https://openreview.net/forum?id=BylBr3C9K7>

AUTHORS: Haichuan Yang, Yuhao Zhu, Ji Liu

TL;DR: Inspired by CapsNet, we propose a novel architecture for graph embeddings on the basis of node features extracted from GNN.

HIGHLIGHT: This paper proposes the first end-to-end DNN training framework that provides quantitative energy consumption guarantees via weighted sparse projection and input masking.

- 103, TITLE: Emerging Disentanglement in Auto-Encoder Based Unsupervised Image Content Transfer  
https://openreview.net/forum?id=ByIE1205Fm  
AUTHORS: Ori Press, Tomer Galanti, Sagie Benaim, Lior Wolf  
TL;DR: An image to image translation method which adds to one image the content of another thereby creating a new image.  
HIGHLIGHT: We study the problem of learning to map, in an unsupervised way, between domains  $A$  and  $B$ , such that the samples  $x$  in  $B$  contain all the information that exists in samples  $x$  in  $A$  and some additional information.
- 104, TITLE: SGD Converges to Global Minimum in Deep Learning via Star-convex Path  
https://openreview.net/forum?id=BylIciRcYQ  
AUTHORS: Yi Zhou, Junjie Yang, Huishuai Zhang, Yingbin Liang, Vahid Tarokh  
TL;DR: An image to image translation method which adds to one image the content of another thereby creating a new image.  
HIGHLIGHT: In this study, we establish the convergence of SGD to a global minimum for nonconvex optimization problems that are commonly encountered in neural network training.
- 105, TITLE: Toward Understanding the Impact of Staleness in Distributed Machine Learning  
https://openreview.net/forum?id=BylQV305YQ  
AUTHORS: Wei Dai, Yi Zhou, Nanqing Dong, Hao Zhang, Eric Xing  
TL;DR: Empirical and theoretical study of the effects of staleness in non-synchronous execution on machine learning algorithms.  
HIGHLIGHT: In this work, we study the convergence behaviors of a wide array of ML models and algorithms under delayed updates.
- 106, TITLE: Transfer Learning for Sequences via Learning to Collocate  
https://openreview.net/forum?id=ByldlhAqYQ  
AUTHORS: Wanyun Cui, Guangyu Zheng, Zhiqiang Shen, Sihang Jiang, Wei Wang  
TL;DR: Transfer learning for sequence via learning to align cell-level information across domains.  
HIGHLIGHT: We conducted extensive experiments on both sequence labeling tasks (POS tagging, NER) and sentence classification (sentiment analysis).
- 107, TITLE: Learning Procedural Abstractions and Evaluating Discrete Latent Temporal Structure  
https://openreview.net/forum?id=ByleB2CcKm  
AUTHORS: Karan Goel, Emma Brunskill  
TL;DR: Transfer learning for sequence via learning to align cell-level information across domains.  
HIGHLIGHT: In this work, we consider the problem of learning procedural abstractions from possibly high-dimensional observational sequences, such as video demonstrations.
- 108, TITLE: Unsupervised Speech Recognition via Segmental Empirical Output Distribution Matching  
https://openreview.net/forum?id=Bylmkh05KX  
AUTHORS: Chih-Kuan Yeh, Jianshu Chen, Chengzhu Yu, Dong Yu  
TL;DR: Transfer learning for sequence via learning to align cell-level information across domains.  
HIGHLIGHT: We propose a fully unsupervised learning algorithm that alternates between solving two sub-problems: (i) learn a phoneme classifier for a given set of phoneme segmentation boundaries, and (ii) refining the phoneme boundaries based on a given classifier.
- 109, TITLE: Adversarial Attacks on Graph Neural Networks via Meta Learning

<https://openreview.net/forum?id=Bylnx209YX>

AUTHORS: Daniel Z?gner, Stephan G?nnemann

TL;DR: We use meta-gradients to attack the training procedure of deep neural networks for graphs.

HIGHLIGHT: Deep learning models for graphs have advanced the state of the art on many tasks.

110, TITLE: Maximal Divergence Sequential Autoencoder for Binary Software Vulnerability Detection

<https://openreview.net/forum?id=ByloIiCqYQ>

AUTHORS: Tue Le, Tuan Nguyen, Trung Le, Dinh Phung, Paul Montague, Olivier De Vel, Lizhen Qu

TL;DR: We propose a novel method named Maximal Divergence Sequential Auto-Encoder that leverages Variational AutoEncoder representation for binary code vulnerability detection.

HIGHLIGHT: In this paper, we attempt to alleviate this severe binary vulnerability detection bottleneck by leveraging recent advances in deep learning representations and propose the Maximal Divergence Sequential Auto-Encoder.

111, TITLE: Neural Program Repair by Jointly Learning to Localize and Repair

<https://openreview.net/forum?id=ByloJ20qtm>

AUTHORS: Marko Vasic, Aditya Kanade, Petros Maniatis, David Bieber, Rishabh Singh

TL;DR: Multi-headed Pointer Networks for jointly learning to localize and repair Variable Misuse bugs

HIGHLIGHT: In this work, we consider a recently identified class of bugs called variable-misuse bugs.

112, TITLE: Information-Directed Exploration for Deep Reinforcement Learning

<https://openreview.net/forum?id=Byx83s09Km>

AUTHORS: Nikolay Nikolov, Johannes Kirschner, Felix Berkenkamp, Andreas Krause

TL;DR: We develop a practical extension of Information-Directed Sampling for Reinforcement Learning, which accounts for parametric uncertainty and heteroscedasticity in the return distribution for exploration.

HIGHLIGHT: Motivated by recent findings that address this issue in bandits, we propose to use Information-Directed Sampling (IDS) for exploration in reinforcement learning.

113, TITLE: Attention, Learn to Solve Routing Problems!

<https://openreview.net/forum?id=ByxBFsRqYm>

AUTHORS: Wouter Kool, Herke van Hoof, Max Welling

TL;DR: Attention based model trained with REINFORCE with greedy rollout baseline to learn heuristics with competitive results on TSP and other routing problems

HIGHLIGHT: We contribute in both directions: we propose a model based on attention layers with benefits over the Pointer Network and we show how to train this model using REINFORCE with a simple baseline based on a deterministic greedy rollout, which we find is more efficient than using a value function.

114, TITLE: L2-Nonexpansive Neural Networks

<https://openreview.net/forum?id=ByxGSsR9FQ>

AUTHORS: Haifeng Qian, Mark N. Wegman

TL;DR: Attention based model trained with REINFORCE with greedy rollout baseline to learn heuristics with competitive results on TSP and other routing problems

HIGHLIGHT: This paper proposes a class of well-conditioned neural networks in which a unit amount of change in the inputs causes at most a unit amount of change in the outputs or any of the internal layers.

115, TITLE: Improving Generalization and Stability of Generative Adversarial Networks

<https://openreview.net/forum?id=ByxPYjC5KQ>

AUTHORS: Hoang Thanh-Tung, Truyen Tran, Svetha Venkatesh  
TL;DR: We propose a zero-centered gradient penalty for improving generalization and stability of GANs  
HIGHLIGHT: In this paper, we analyze the generalization of GANs in practical settings.

116, TITLE: Adaptive Input Representations for Neural Language Modeling  
<https://openreview.net/forum?id=ByxZX20qFQ>  
AUTHORS: Alexei Baevski, Michael Auli  
TL;DR: Variable capacity input word embeddings and SOTA on WikiText-103, Billion Word benchmarks.  
HIGHLIGHT: We introduce adaptive input representations for neural language modeling which extend the adaptive softmax of Grave et al. (2017) to input representations of variable capacity.

117, TITLE: Neural Persistence: A Complexity Measure for Deep Neural Networks Using Algebraic Topology  
<https://openreview.net/forum?id=ByxkijC5FQ>  
AUTHORS: Bastian Rieck, Matteo Togninalli, Christian Bock, Michael Moor, Max Horn, Thomas Gumbsch, Karsten Borgwardt  
TL;DR: We develop a new topological complexity measure for deep neural networks and demonstrate that it captures their salient properties.  
HIGHLIGHT: In this work, we propose neural persistence, a complexity measure for neural network architectures based on topological data analysis on weighted stratified graphs.

118, TITLE: Efficient Augmentation via Data Subsampling  
<https://openreview.net/forum?id=Byxpfh0cFm>  
AUTHORS: Michael Kuchnik, Virginia Smith  
TL;DR: Selectively augmenting difficult to classify points results in efficient training.  
HIGHLIGHT: In this work, we demonstrate that it is possible to significantly reduce the number of data points included in data augmentation while realizing the same accuracy and invariance benefits of augmenting the entire dataset.  
We propose a novel set of subsampling policies, based on model influence and loss, that can achieve a 90% reduction in augmentation set size while maintaining the accuracy gains of standard data augmentation.

119, TITLE: Neural TTS Stylization with Adversarial and Collaborative Games  
<https://openreview.net/forum?id=ByzcS3AcYX>  
AUTHORS: Shuang Ma, Daniel McDuff, Yale Song  
TL;DR: a generative adversarial network for style modeling in a text-to-speech system  
HIGHLIGHT: In this work, we introduce an end-to-end TTS model that offers enhanced content-style disentanglement ability and controllability.

120, TITLE: Optimal Control Via Neural Networks: A Convex Approach  
<https://openreview.net/forum?id=H1MW72AcK7>  
AUTHORS: Yize Chen, Yuanyuan Shi, Baosen Zhang  
TL;DR: a generative adversarial network for style modeling in a text-to-speech system  
HIGHLIGHT: Therefore many systems are still identified and controlled based on simple linear models despite their poor representation capability. In this paper we bridge the gap between model accuracy and control tractability faced by neural networks, by explicitly constructing networks that are convex with respect to their inputs.

121, TITLE: CBOW Is Not All You Need: Combining CBOW with the Compositional Matrix Space Model  
<https://openreview.net/forum?id=H1MgjoR9tQ>

AUTHORS: Florian Mai, Lukas Galke, Ansgar Scherp  
TL;DR: We present a novel training scheme for efficiently obtaining order-aware sentence representations.  
HIGHLIGHT: Motivated by these findings, we propose a hybrid model that combines the strengths of CBOW and CMOW.

122, TITLE: Stochastic Optimization of Sorting Networks via Continuous Relaxations  
<https://openreview.net/forum?id=H1eSS3CcKX>  
AUTHORS: Aditya Grover, Eric Wang, Aaron Zweig, Stefano Ermon  
TL;DR: We provide a continuous relaxation to the sorting operator, enabling end-to-end, gradient-based stochastic optimization.  
HIGHLIGHT: In this work, we propose NeuralSort, a general-purpose continuous relaxation of the output of the sorting operator from permutation matrices to the set of unimodal row-stochastic matrices, where every row sums to one and has a distinct argmax.

123, TITLE: Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: optimal rate and curse of dimensionality  
<https://openreview.net/forum?id=H1ebTsActm>  
AUTHORS: Taiji Suzuki  
TL;DR: We provide a continuous relaxation to the sorting operator, enabling end-to-end, gradient-based stochastic optimization.  
HIGHLIGHT: Deep learning has shown high performances in various types of tasks from visual recognition to natural language processing, which indicates superior flexibility and adaptivity of deep learning. To understand this phenomenon theoretically, we develop a new approximation and estimation error analysis of deep learning with the ReLU activation for functions in a Besov space and its variant with mixed smoothness. The Besov space is a considerably general function space including the Holder space and Sobolev space, and especially can capture spatial inhomogeneity of smoothness.

124, TITLE: Generating Multiple Objects at Spatially Distinct Locations  
<https://openreview.net/forum?id=H1edLiA9KQ>  
AUTHORS: Tobias Hinz, Stefan Heinrich, Stefan Wermter  
TL;DR: Extend GAN architecture to obtain control over locations and identities of multiple objects within generated images.  
HIGHLIGHT: We introduce a new approach which allows us to control the location of arbitrarily many objects within an image by adding an object pathway to both the generator and the discriminator.

125, TITLE: Near-Optimal Representation Learning for Hierarchical Reinforcement Learning  
<https://openreview.net/forum?id=H1emus0qF7>  
AUTHORS: Ofir Nachum, Shixiang Gu, Honglak Lee, Sergey Levine  
TL;DR: We translate a bound on sub-optimality of representations to a practical training objective in the context of hierarchical reinforcement learning.  
HIGHLIGHT: We study the problem of representation learning in goal-conditioned hierarchical reinforcement learning.

126, TITLE: Understanding Composition of Word Embeddings via Tensor Decomposition  
<https://openreview.net/forum?id=H1eqjiCctX>  
AUTHORS: Abraham Frandsen, Rong Ge  
TL;DR: We present a generative model for compositional word embeddings that captures syntactic relations, and provide empirical verification and evaluation.  
HIGHLIGHT: In this paper we consider the problem of word embedding composition  $\vec{w} + \vec{v}$  given vector representations of two words, compute a vector for the entire phrase.

127, TITLE: Structured Neural Summarization  
<https://openreview.net/forum?id=H1ersoRqtm>  
AUTHORS: Patrick Fernandes, Miltiadis Allamanis, Marc Brockschmidt  
TL;DR: One simple trick to improve sequence models: Compose them with a graph model  
HIGHLIGHT: Based on the promising results of graph neural networks on highly structured data, we develop a framework to extend existing sequence encoders with a graph component that can reason about long-distance relationships in weakly structured data such as text.

128, TITLE: Graph Wavelet Neural Network  
<https://openreview.net/forum?id=H1ewdiR5tQ>  
AUTHORS: Bingbing Xu, Huawei Shen, Qi Cao, Yunqi Qiu, Xueqi Cheng  
TL;DR: We present graph wavelet neural network (GWNN), a novel graph convolutional neural network (CNN), leveraging graph wavelet transform to address the shortcoming of previous spectral graph CNN methods that depend on graph Fourier transform.  
HIGHLIGHT: We present graph wavelet neural network (GWNN), a novel graph convolutional neural network (CNN), leveraging graph wavelet transform to address the shortcomings of previous spectral graph CNN methods that depend on graph Fourier transform.

129, TITLE: A rotation-equivariant convolutional neural network model of primary visual cortex  
<https://openreview.net/forum?id=H1fU8iAqKX>  
AUTHORS: 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  
TL;DR: A rotation-equivariant CNN model of V1 that outperforms previous models and suggest functional groupings of V1 neurons.  
HIGHLIGHT: We present a framework for identifying common features independent of individual neurons' orientation selectivity by using a rotation-equivariant convolutional neural network, which automatically extracts every feature at multiple different orientations.

130, TITLE: Supervised Community Detection with Line Graph Neural Networks  
<https://openreview.net/forum?id=H1g0Z3A9Fm>  
AUTHORS: Zhengdao Chen, Lisha Li, Joan Bruna  
TL;DR: We propose a novel graph neural network architecture based on the non-backtracking matrix defined over the edge adjacencies and demonstrate its effectiveness in community detection tasks on graphs.  
HIGHLIGHT: We present a novel family of Graph Neural Networks (GNNs) for solving community detection problems in a supervised learning setting.

131, TITLE: Multiple-Attribute Text Rewriting  
<https://openreview.net/forum?id=H1g2NhC5KQ>  
AUTHORS: Guillaume Lample, Sandeep Subramanian, Eric Smith, Ludovic Denoyer, Marc'Aurelio Ranzato, Y-Lan Boureau  
TL;DR: A system for rewriting text conditioned on multiple controllable attributes  
HIGHLIGHT: We thus propose a new model that controls several factors of variation in textual data where this condition on disentanglement is replaced with a simpler mechanism based on back-translation.

132, TITLE: Wasserstein Barycenter Model Ensembling  
<https://openreview.net/forum?id=H1g4k309F7>  
AUTHORS: Pierre Dognin\*, Igor Melnyk\*, Youssef Mroueh\*, Jarret Ross\*, Cicero Dos Santos\*, Tom Sercu\*  
TL;DR: we propose to use Wasserstein barycenters for semantic model ensembling

**HIGHLIGHT:** In this paper we propose to perform model ensembling in a multiclass or a multilabel learning setting using Wasserstein (W.) barycenters.

133, **TITLE:** Policy Transfer with Strategy Optimization  
<https://openreview.net/forum?id=H1g6osRcFQ>  
**AUTHORS:** Wenhao Yu, C. Karen Liu, Greg Turk  
**TL;DR:** We propose a policy transfer algorithm that can overcome large and challenging discrepancies in the system dynamics such as latency, actuator modeling error, etc.  
**HIGHLIGHT:** In this paper, we present a different approach that leverages domain randomization for transferring control policies to unknown environments.

134, **TITLE:** code2seq: Generating Sequences from Structured Representations of Code  
<https://openreview.net/forum?id=H1gKYo09tX>  
**AUTHORS:** Uri Alon, Shaked Brody, Omer Levy, Eran Yahav  
**TL;DR:** We leverage the syntactic structure of source code to generate natural language sequences.  
**HIGHLIGHT:** We present code2seq: an alternative approach that leverages the syntactic structure of programming languages to better encode source code.

135, **TITLE:** Predict then Propagate: Graph Neural Networks meet Personalized PageRank  
<https://openreview.net/forum?id=H1gL-2A9Ym>  
**AUTHORS:** Johannes Klicpera, Aleksandar Bojchevski, Stephan G?nnemann  
**TL;DR:** Personalized propagation of neural predictions (PPNP) improves graph neural networks by separating them into prediction and propagation via personalized PageRank.  
**HIGHLIGHT:** In this paper, we use the relationship between graph convolutional networks (GCN) and PageRank to derive an improved propagation scheme based on personalized PageRank.

136, **TITLE:** Slimmable Neural Networks  
<https://openreview.net/forum?id=H1gMCsAqY7>  
**AUTHORS:** Jiahui Yu, Linjie Yang, Ning Xu, Jianchao Yang, Thomas Huang  
**TL;DR:** We present a simple and general method to train a single neural network executable at different widths (number of channels in a layer), permitting instant and adaptive accuracy-efficiency trade-offs at runtime.  
**HIGHLIGHT:** We present a simple and general method to train a single neural network executable at different widths (number of channels in a layer), permitting instant and adaptive accuracy-efficiency trade-offs at runtime.

137, **TITLE:** Analysing Mathematical Reasoning Abilities of Neural Models  
<https://openreview.net/forum?id=H1gR5iR5FX>  
**AUTHORS:** David Saxton, Edward Grefenstette, Felix Hill, Pushmeet Kohli  
**TL;DR:** A dataset for testing mathematical reasoning (and algebraic generalization), and results on current sequence-to-sequence models.  
**HIGHLIGHT:** In this paper, we present a new challenge for the evaluation (and eventually the design) of neural architectures and similar system, developing a task suite of mathematics problems involving sequential questions and answers in a free-form textual input/output format.

138, **TITLE:** RotDCF: Decomposition of Convolutional Filters for Rotation-Equivariant Deep Networks  
<https://openreview.net/forum?id=H1gTEj09FX>  
**AUTHORS:** Xiuyuan Cheng, Qiang Qiu, Robert Calderbank, Guillermo Sapiro

TL;DR: A dataset for testing mathematical reasoning (and algebraic generalization), and results on current sequence-to-sequence models.

HIGHLIGHT: This paper proposes to decompose the convolutional filters over joint steerable bases across the space and the group geometry simultaneously, namely a rotation-equivariant CNN with decomposed convolutional filters (RotDCF).

139, TITLE: Execution-Guided Neural Program Synthesis

<https://openreview.net/forum?id=H1gfOiAqYm>

AUTHORS: Xinyun Chen, Chang Liu, Dawn Song

TL;DR: A dataset for testing mathematical reasoning (and algebraic generalization), and results on current sequence-to-sequence models.

HIGHLIGHT: In this work, we propose two simple yet principled techniques to better leverage the semantic information, which are execution-guided synthesis and synthesizer ensemble.

140, TITLE: Dynamic Sparse Graph for Efficient Deep Learning

<https://openreview.net/forum?id=H1goBoR9F7>

AUTHORS: Liu Liu, Lei Deng, Xing Hu, Maohua Zhu, Guoqi Li, Yufei Ding, Yuan Xie

TL;DR: We construct dynamic sparse graph via dimension-reduction search to reduce compute and memory cost in both DNN training and inference.

HIGHLIGHT: We propose to execute deep neural networks (DNNs) with dynamic and sparse graph (DSG) structure for compressive memory and accelerative execution during both training and inference.

141, TITLE: Fixup Initialization: Residual Learning Without Normalization

<https://openreview.net/forum?id=H1gsz30cKX>

AUTHORS: Hongyi Zhang, Yann N. Dauphin, Tengyu Ma

TL;DR: All you need to train deep residual networks is a good initialization; normalization layers are not necessary.

HIGHLIGHT: In this work, we challenge the commonly-held beliefs by showing that none of the perceived benefits is unique to normalization.

142, TITLE: ProbGAN: Towards Probabilistic GAN with Theoretical Guarantees

<https://openreview.net/forum?id=H117bnR5Ym>

AUTHORS: Hao He, Hao Wang, Guang-He Lee, Yonglong Tian

TL;DR: A novel probabilistic treatment for GAN with theoretical guarantee.

HIGHLIGHT: In this paper, we propose a novel probabilistic framework for GANs, ProbGAN, which iteratively learns a distribution over generators with a carefully crafted prior.

143, TITLE: Exploration by random network distillation

<https://openreview.net/forum?id=H11JJnR5Ym>

AUTHORS: Yuri Burda, Harrison Edwards, Amos Storkey, Oleg Klimov

TL;DR: A simple exploration bonus is introduced and achieves state of the art performance in 3 hard exploration Atari games.

HIGHLIGHT: We introduce an exploration bonus for deep reinforcement learning methods that is easy to implement and adds minimal overhead to the computation performed.

144, TITLE: Unsupervised Learning of the Set of Local Maxima

<https://openreview.net/forum?id=H11qZhRcFm>

AUTHORS: Lior Wolf, Sagie Benaim, Tomer Galanti

TL;DR: A simple exploration bonus is introduced and achieves state of the art performance in 3 hard exploration Atari games.

HIGHLIGHT: We present an algorithm, show an example where it is more efficient to use local maxima as an indicator function than to employ conventional classification, and derive a suitable generalization bound.

145, TITLE: On the Convergence of A Class of Adam-Type Algorithms for Non-Convex Optimization

<https://openreview.net/forum?id=H1x-x309tm>

AUTHORS: Xiangyi Chen, Sijia Liu, Ruoyu Sun, Mingyi Hong

TL;DR: We analyze convergence of Adam-type algorithms and provide mild sufficient conditions to guarantee their convergence, we also show violating the conditions can makes an algorithm diverge.

HIGHLIGHT: In this paper, we develop an analysis framework and a set of mild sufficient conditions that guarantee the convergence of the Adam-type methods, with a convergence rate of order  $\mathcal{O}(\log\{T\}^{\frac{1}{2}}\sqrt{T})$  for non-convex stochastic optimization.

146, TITLE: Minimum Divergence vs. Maximum Margin: an Empirical Comparison on Seq2Seq Models

<https://openreview.net/forum?id=H1xD9sR5Fm>

AUTHORS: Huan Zhang, Hai Zhao

TL;DR: We analyze convergence of Adam-type algorithms and provide mild sufficient conditions to guarantee their convergence, we also show violating the conditions can makes an algorithm diverge.

HIGHLIGHT: We introduce a new training criterion based on the analysis of existing work, and empirically compare models in the two categories.

147, TITLE: GANSynth: Adversarial Neural Audio Synthesis

<https://openreview.net/forum?id=H1xQVn09FX>

AUTHORS: Jesse Engel, Kumar Krishna Agrawal, Shuo Chen, Ishaan Gulrajani, Chris Donahue, Adam Roberts

TL;DR: High-quality audio synthesis with GANs

HIGHLIGHT: Herein, we demonstrate that GANs can in fact generate high-fidelity and locally-coherent audio by modeling log magnitudes and instantaneous frequencies with sufficient frequency resolution in the spectral domain.

148, TITLE: Sliced Wasserstein Auto-Encoders

<https://openreview.net/forum?id=H1xaJn05FQ>

AUTHORS: Soheil Kolouri, Phillip E. Pope, Charles E. Martin, Gustavo K. Rohde

TL;DR: In this paper we use the sliced-Wasserstein distance to shape the latent distribution of an auto-encoder into any samplable prior distribution.

HIGHLIGHT: In this paper we use the geometric properties of the optimal transport (OT) problem and the Wasserstein distances to define a prior distribution for the latent space of an auto-encoder.

149, TITLE: Learning Two-layer Neural Networks with Symmetric Inputs

<https://openreview.net/forum?id=H1xipsA5K7>

AUTHORS: Rong Ge, Rohith Kuditipudi, Zhize Li, Xiang Wang

TL;DR: We give an algorithm for learning a two-layer neural network with symmetric input distribution.

HIGHLIGHT: We give a new algorithm for learning a two-layer neural network under a very general class of input distributions.

150, TITLE: Learning to Understand Goal Specifications by Modelling Reward

<https://openreview.net/forum?id=H1xsSjC9Ym>

AUTHORS: Dzmitry Bahdanau, Felix Hill, Jan Leike, Edward Hughes, Arian Hosseini, Pushmeet Kohli, Edward Grefenstette  
TL;DR: We propose AGILE, a framework for training agents to perform instructions from examples of respective goal-states.  
HIGHLIGHT: To overcome this limitation, we present a framework within which instruction-conditional RL agents are trained using rewards obtained not from the environment, but from reward models which are jointly trained from expert examples.

151, TITLE: Do Deep Generative Models Know What They Don't Know?  
<https://openreview.net/forum?id=H1xwNhCcYm>  
AUTHORS: Eric Nalisnick, Akihiro Matsukawa, Yee Whye Teh, Dilan Gorur, Balaji Lakshminarayanan  
TL;DR: We propose AGILE, a framework for training agents to perform instructions from examples of respective goal-states.  
HIGHLIGHT: In this paper we challenge this assumption.

152, TITLE: Identifying and Controlling Important Neurons in Neural Machine Translation  
<https://openreview.net/forum?id=H1z-PsR5KX>  
AUTHORS: Anthony Bau, Yonatan Belinkov, Hassan Sajjad, Nadir Durrani, Fahim Dalvi, James Glass  
TL;DR: Unsupervised methods for finding, analyzing, and controlling important neurons in NMT  
HIGHLIGHT: We develop unsupervised methods for discovering important neurons in NMT models.

153, TITLE: Representing Formal Languages: A Comparison Between Finite Automata and Recurrent Neural Networks  
<https://openreview.net/forum?id=H1zeHnA9KX>  
AUTHORS: Joshua J. Michalenko, Ameesh Shah, Abhinav Verma, Richard G. Baraniuk, Swarat Chaudhuri, Ankit B. Patel  
TL;DR: Finite Automata Can be Linearly decoded from Language-Recognizing RNNs using low coarseness abstraction functions and high accuracy decoders.  
HIGHLIGHT: We investigate the internal representations that a recurrent neural network (RNN) uses while learning to recognize a regular formal language.

154, TITLE: Visual Explanation by Interpretation: Improving Visual Feedback Capabilities of Deep Neural Networks  
<https://openreview.net/forum?id=H1ziPjC5Fm>  
AUTHORS: Jose Oramas, Kaili Wang, Tinne Tuytelaars  
TL;DR: Interpretation by Identifying model-learned features that serve as indicators for the task of interest. Explain model decisions by highlighting the response of these features in test data. Evaluate explanations objectively with a controlled dataset.  
HIGHLIGHT: In this paper, we propose a novel scheme for both interpretation as well as explanation in which, given a pretrained model, we automatically identify internal features relevant for the set of classes considered by the model, without relying on additional annotations.

155, TITLE: Don't let your Discriminator be fooled  
<https://openreview.net/forum?id=HJE6X305Fm>  
AUTHORS: Brady Zhou, Philipp Kr?henb?hl  
TL;DR: A discriminator that is not easily fooled by adversarial example makes GAN training more robust and leads to a smoother objective.  
HIGHLIGHT: In this paper, we show that the Wasserstein distance is just one out of a large family of objective functions that yield these properties.

156, TITLE: Latent Convolutional Models  
<https://openreview.net/forum?id=HJGciiR5Y7>  
AUTHORS: ShahRukh Athar, Evgeny Burnaev, Victor Lempitsky  
TL;DR: We present a new deep latent model of natural images that can be trained from unlabeled datasets and can be utilized to solve various image restoration tasks.  
HIGHLIGHT: We present a new latent model of natural images that can be learned on large-scale datasets.

157, TITLE: A Universal Music Translation Network  
<https://openreview.net/forum?id=HJGkisCcKm>  
AUTHORS: Noam Mor, Lior Wolf, Adam Polyak, Yaniv Taigman  
TL;DR: An automatic method for converting music between instruments and styles  
HIGHLIGHT: We present a method for translating music across musical instruments and styles.

158, TITLE: How to train your MAML  
<https://openreview.net/forum?id=HJGven05Y7>  
AUTHORS: Antreas Antoniou, Harrison Edwards, Amos Storkey  
TL;DR: MAML is great, but it has many problems, we solve many of those problems and as a result we learn most hyper parameters end to end, speed-up training and inference and set a new SOTA in few-shot learning  
HIGHLIGHT: In this paper, we propose various modifications to MAML that not only stabilize the system, but also substantially improve the generalization performance, convergence speed and computational overhead of MAML, which we call MAML++.

159, TITLE: Learning a SAT Solver from Single-Bit Supervision  
[https://openreview.net/forum?id=HJMC\\_iA5tm](https://openreview.net/forum?id=HJMC_iA5tm)  
AUTHORS: Daniel Selsam, Matthew Lamm, Benedikt B" {u}nz, Percy Liang, Leonardo de Moura, David L. Dill  
TL;DR: We train a graph network to predict boolean satisfiability and show that it learns to search for solutions, and that the solutions it finds can be decoded from its activations.  
HIGHLIGHT: We present NeuroSAT, a message passing neural network that learns to solve SAT problems after only being trained as a classifier to predict satisfiability.

160, TITLE: Learning Representations of Sets through Optimized Permutations  
<https://openreview.net/forum?id=HJMCcjAcYX>  
AUTHORS: Yan Zhang, Jonathon Hare, Adam Pr?gel-Bennett  
TL;DR: Learn how to permute a set, then encode permuted set with RNN to obtain a set representation.  
HIGHLIGHT: To this end, we propose a Permutation-Optimisation module that learns how to permute a set end-to-end.

161, TITLE: Big-Little Net: An Efficient Multi-Scale Feature Representation for Visual and Speech Recognition  
<https://openreview.net/forum?id=HJMHpjC9Ym>  
AUTHORS: Chun-Fu (Richard) Chen, Quanfu Fan, Neil Mallinar, Tom Sercu, Rogerio Feris  
TL;DR: Learn how to permute a set, then encode permuted set with RNN to obtain a set representation.  
HIGHLIGHT: In this paper, we propose a novel Convolutional Neural Network (CNN) architecture for learning multi-scale feature representations with good tradeoffs between speed and accuracy.

162, TITLE: Unsupervised Hyper-alignment for Multilingual Word Embeddings  
<https://openreview.net/forum?id=HJe62s09tX>  
AUTHORS: Jean Alaux, Edouard Grave, Marco Cuturi, Armand Joulin  
TL;DR: Learn how to permute a set, then encode permuted set with RNN to obtain a set representation.  
HIGHLIGHT: We thus propose a novel formulation that ensures composable mappings, leading to better alignments.

163, TITLE: Visual Semantic Navigation using Scene Priors  
<https://openreview.net/forum?id=HJeRkh05Km>  
AUTHORS: Wei Yang, Xiaolong Wang, Ali Farhadi, Abhinav Gupta, Roozbeh Mottaghi  
TL;DR: Learn how to permute a set, then encode permuted set with RNN to obtain a set representation.  
HIGHLIGHT: In this work, we focus on incorporating semantic priors in the task of semantic navigation.

164, TITLE: NOODL: Provable Online Dictionary Learning and Sparse Coding  
<https://openreview.net/forum?id=HJeu43ActQ>  
AUTHORS: Sirisha Rambhatla, Xingguo Li, Jarvis Haupt  
TL;DR: We present a provable algorithm for exactly recovering both factors of the dictionary learning model.  
HIGHLIGHT: This was a major challenge until recently, when provable algorithms for dictionary learning were proposed.

165, TITLE: Stochastic Gradient/Mirror Descent: Minimax Optimality and Implicit Regularization  
<https://openreview.net/forum?id=HJf9ZhC9FX>  
AUTHORS: Navid Azizan, Babak Hassibi  
TL;DR: We present a provable algorithm for exactly recovering both factors of the dictionary learning model.  
HIGHLIGHT: Stochastic Gradient/Mirror Descent: Minimax Optimality and Implicit Regularization.

166, TITLE: Active Learning with Partial Feedback  
<https://openreview.net/forum?id=HJfSEnRqKQ>  
AUTHORS: Peiyun Hu, Zachary C. Lipton, Anima Anandkumar, Deva Ramanan  
TL;DR: We provide a new perspective on training a machine learning model from scratch in hierarchical label setting, i.e. thinking of it as two-way communication between human and algorithms, and study how we can both measure and improve the efficiency.  
HIGHLIGHT: To address this more realistic setting, we propose active learning with partial feedback (ALPF), where the learner must actively choose both which example to label and which binary question to ask.

167, TITLE: Gradient descent aligns the layers of deep linear networks  
<https://openreview.net/forum?id=HJflg30qKX>  
AUTHORS: Ziwei Ji, Matus Telgarsky  
TL;DR: We provide a new perspective on training a machine learning model from scratch in hierarchical label setting, i.e. thinking of it as two-way communication between human and algorithms, and study how we can both measure and improve the efficiency.  
HIGHLIGHT: This paper establishes risk convergence and asymptotic weight matrix alignment --- a form of implicit regularization --- of gradient flow and gradient descent when applied to deep linear networks on linearly separable data.

168, TITLE: Data-Dependent Coresets for Compressing Neural Networks with Applications to Generalization Bounds  
<https://openreview.net/forum?id=HJfwJ2A5KX>  
AUTHORS: Cenk Baykal, Lucas Liebenwein, Igor Gilitschenski, Dan Feldman, Daniela Rus  
TL;DR: We provide a new perspective on training a machine learning model from scratch in hierarchical label setting, i.e. thinking of it as two-way communication between human and algorithms, and study how we can both measure and improve the efficiency.  
HIGHLIGHT: We present an efficient coresets-based neural network compression algorithm that sparsifies the parameters of a trained fully-connected neural network in a manner that provably approximates the network's output.

169, TITLE: On the loss landscape of a class of deep neural networks with no bad local valleys  
<https://openreview.net/forum?id=HJgXsjA5tQ>  
AUTHORS: Quynh Nguyen, Mahesh Chandra Mukkamala, Matthias Hein  
TL;DR: We provide a new perspective on training a machine learning model from scratch in hierarchical label setting, i.e. thinking of it as two-way communication between human and algorithms, and study how we can both measure and improve the efficiency.  
HIGHLIGHT: We identify a class of over-parameterized deep neural networks with standard activation functions and cross-entropy loss which provably have no bad local valley, in the sense that from any point in parameter space there exists a continuous path on which the cross-entropy loss is non-increasing and gets arbitrarily close to zero.

170, TITLE: DOM-Q-NET: Grounded RL on Structured Language  
<https://openreview.net/forum?id=HJgd1nAqFX>  
AUTHORS: Sheng Jia, Jamie Ryan Kiros, Jimmy Ba  
TL;DR: Graph-based Deep Q Network for Web Navigation  
HIGHLIGHT: In this work, we introduce DOM-Q-NET, a novel architecture for RL-based web navigation to address both of these problems.

171, TITLE: Boosting Robustness Certification of Neural Networks  
<https://openreview.net/forum?id=HJgeEh09KQ>  
AUTHORS: Gagandeep Singh, Timon Gehr, Markus P?schel, Martin Vechev  
TL;DR: We refine the over-approximation results from incomplete verifiers using MILP solvers to prove more robustness properties than state-of-the-art.  
HIGHLIGHT: We present a novel approach for the certification of neural networks against adversarial perturbations which combines scalable overapproximation methods with precise (mixed integer) linear programming.

172, TITLE: Learning To Simulate  
<https://openreview.net/forum?id=HJgkx2Aqt7>  
AUTHORS: Nataniel Ruiz, Samuel Schuler, Manmohan Chandraker  
TL;DR: We propose an algorithm that automatically adjusts parameters of a simulation engine to generate training data for a neural network such that validation accuracy is maximized.  
HIGHLIGHT: In this work, we propose a reinforcement learning-based method for automatically adjusting the parameters of any (non-differentiable) simulator, thereby controlling the distribution of synthesized data in order to maximize the accuracy of a model trained on that data.

173, TITLE: Towards Understanding Regularization in Batch Normalization  
<https://openreview.net/forum?id=HJILKjR9FQ>  
AUTHORS: Ping Luo, Xinjiang Wang, Wenqi Shao, Zhanglin Peng

TL;DR: We propose an algorithm that automatically adjusts parameters of a simulation engine to generate training data for a neural network such that validation accuracy is maximized.

HIGHLIGHT: We analyze BN by using a basic block of neural networks, consisting of a kernel layer, a BN layer, and a nonlinear activation function.

174, TITLE: The Laplacian in RL: Learning Representations with Efficient Approximations

<https://openreview.net/forum?id=HJINpoA5YQ>

AUTHORS: Yifan Wu, George Tucker, Ofir Nachum

TL;DR: We propose a scalable method to approximate the eigenvectors of the Laplacian in the reinforcement learning context and we show that the learned representations can improve the performance of an RL agent.

HIGHLIGHT: In this paper, we present a fully general and scalable method for approximating the eigenvectors of the Laplacian in a model-free RL context.

175, TITLE: Predicting the Generalization Gap in Deep Networks with Margin Distributions

<https://openreview.net/forum?id=HJlQfnCqKX>

AUTHORS: Yiding Jiang, Dilip Krishnan, Hossein Mobahi, Samy Bengio

TL;DR: We develop a new scheme to predict the generalization gap in deep networks with high accuracy.

HIGHLIGHT: In this paper, we propose such a measure, and conduct extensive empirical studies on how well it can predict the generalization gap.

176, TITLE: Adversarial Imitation via Variational Inverse Reinforcement Learning

<https://openreview.net/forum?id=HJlmHoR5tQ>

AUTHORS: Ahmed H. Qureshi, Byron Boots, Michael C. Yip

TL;DR: Our method introduces the empowerment-regularized maximum-entropy inverse reinforcement learning to learn near-optimal rewards and policies from expert demonstrations.

HIGHLIGHT: Our proposed method builds on the framework of generative adversarial networks and introduces the empowerment-regularized maximum-entropy inverse reinforcement learning to learn near-optimal rewards and policies.

177, TITLE: Reasoning About Physical Interactions with Object-Oriented Prediction and Planning

<https://openreview.net/forum?id=HJx9EhC9tQ>

AUTHORS: Michael Janner, Sergey Levine, William T. Freeman, Joshua B. Tenenbaum, Chelsea Finn, Jiajun Wu

TL;DR: We present a framework for learning object-centric representations suitable for planning in tasks that require an understanding of physics.

HIGHLIGHT: We present a paradigm for learning object-centric representations for physical scene understanding without direct supervision of object properties.

178, TITLE: LayoutGAN: Generating Graphic Layouts with Wireframe Discriminators

<https://openreview.net/forum?id=HJxB5sRcFQ>

AUTHORS: Jianan Li, Jimei Yang, Aaron Hertzmann, Jianming Zhang, Tingfa Xu

TL;DR: We present a framework for learning object-centric representations suitable for planning in tasks that require an understanding of physics.

HIGHLIGHT: We propose a novel Generative Adversarial Network, called LayoutGAN, that synthesizes layouts by modeling geometric relations of different types of 2D elements.

179, TITLE: Learning Mixed-Curvature Representations in Product Spaces

<https://openreview.net/forum?id=HJxeWnCcF7>

AUTHORS: Albert Gu, Frederic Sala, Beliz Gunel, Christopher R?

TL;DR: Product manifold embedding spaces with heterogenous curvature yield improved representations compared to traditional embedding spaces for a variety of structures.

HIGHLIGHT: The quality of the representations achieved by embeddings is determined by how well the geometry of the embedding space matches the structure of the data. Euclidean space has been the workhorse for embeddings; recently hyperbolic and spherical spaces have gained popularity due to their ability to better embed new types of structured data---such as hierarchical data---but most data is not structured so uniformly. We address this problem by proposing learning embeddings in a product manifold combining multiple copies of these model spaces (spherical, hyperbolic, Euclidean), providing a space of heterogeneous curvature suitable for a wide variety of structures. We introduce a heuristic to estimate the sectional curvature of graph data and directly determine an appropriate signature---the number of component spaces and their dimensions---of the product manifold. Empirically, we jointly learn the curvature and the embedding in the product space via Riemannian optimization. We discuss how to define and compute intrinsic quantities such as means---a challenging notion for product manifolds---and provably learnable optimization functions. On a range of datasets and reconstruction tasks, our product space embeddings outperform single Euclidean or hyperbolic spaces used in previous works, reducing distortion by 32.55% on a Facebook social network dataset.

180, TITLE: StrokeNet: A Neural Painting Environment

<https://openreview.net/forum?id=HJxwDiActX>

AUTHORS: Ningyuan Zheng, Yifan Jiang, Dingjiang Huang

TL;DR: StrokeNet is a novel architecture where the agent is trained to draw by strokes on a differentiable simulation of the environment, which could effectively exploit the power of back-propagation.

HIGHLIGHT: In this paper we try to address the discrete nature of software environment with an intermediate, differentiable simulation.

181, TITLE: Harmonizing Maximum Likelihood with GANs for Multimodal Conditional Generation

<https://openreview.net/forum?id=HJxyAjRcFX>

AUTHORS: Soochan Lee, Junsoo Ha, Gunhee Kim

TL;DR: We prove that the mode collapse in conditional GANs is largely attributed to a mismatch between reconstruction loss and GAN loss and introduce a set of novel loss functions as alternatives for reconstruction loss.

HIGHLIGHT: In order to accomplish both training stability and multimodal output generation, we propose novel training schemes with a new set of losses named moment reconstruction losses that simply replace the reconstruction loss.

182, TITLE: Measuring Compositionality in Representation Learning

<https://openreview.net/forum?id=HJz05o0qK7>

AUTHORS: Jacob Andreas

TL;DR: This paper proposes a simple procedure for evaluating compositional structure in learned representations, and uses the procedure to explore the role of compositionality in four learning problems.

HIGHLIGHT: We describe a procedure for evaluating compositionality by measuring how well the true representation-producing model can be approximated by a model that explicitly composes a collection of inferred representational primitives.

183, TITLE: Benchmarking Neural Network Robustness to Common Corruptions and Perturbations

<https://openreview.net/forum?id=HJz6tiCqYm>

AUTHORS: Dan Hendrycks, Thomas Dietterich

TL;DR: We propose ImageNet-C to measure classifier corruption robustness and ImageNet-P to measure perturbation robustness

HIGHLIGHT: In this paper we establish rigorous benchmarks for image classifier robustness. Then we propose a new dataset called ImageNet-P which enables researchers to benchmark a classifier's robustness to common perturbations.

184, TITLE: ADef: an Iterative Algorithm to Construct Adversarial Deformations  
<https://openreview.net/forum?id=Hk4dFjR5K7>  
AUTHORS: Rima Alaifari, Giovanni S. Albeti, Tandri Gauksson  
TL;DR: We propose a new, efficient algorithm to construct adversarial examples by means of deformations, rather than additive perturbations.  
HIGHLIGHT: In this paper, we propose the ADef algorithm to construct a different kind of adversarial attack created by iteratively applying small deformations to the image, found through a gradient descent step.

185, TITLE: Discriminator-Actor-Critic: Addressing Sample Inefficiency and Reward Bias in Adversarial Imitation Learning  
<https://openreview.net/forum?id=Hk4fpoA5Km>  
AUTHORS: Ilya Kostrikov, Kumar Krishna Agrawal, Debidatta Dwibedi, Sergey Levine, Jonathan Tompson  
TL;DR: We address sample inefficiency and reward bias in adversarial imitation learning algorithms such as GAIL and AIRL.  
HIGHLIGHT: In order to address these issues, we propose a new algorithm called Discriminator-Actor-Critic that uses off-policy Reinforcement Learning to reduce policy-environment interaction sample complexity by an average factor of 10.

186, TITLE: Doubly Reparameterized Gradient Estimators for Monte Carlo Objectives  
<https://openreview.net/forum?id=HkG3e205K7>  
AUTHORS: George Tucker, Dieterich Lawson, Shixiang Gu, Chris J. Maddison  
TL;DR: Doubly reparameterized gradient estimators provide unbiased variance reduction which leads to improved performance.  
HIGHLIGHT: In particular, we show that this estimator reduces the variance of the IWAE gradient, the reweighted wake-sleep update (RWS) (Bornschein & Bengio 2014), and the jackknife variational inference (JVI) gradient (Nowozin 2018).

187, TITLE: Learning Recurrent Binary/Ternary Weights  
<https://openreview.net/forum?id=HkNGYjR9FX>  
AUTHORS: Arash Ardakani, Zhengyun Ji, Sean C. Smithson, Brett H. Meyer, Warren J. Gross  
TL;DR: We propose high-performance LSTMs with binary/ternary weights, that can greatly reduce implementation complexity  
HIGHLIGHT: To address the above issues, we introduce a method that can learn binary and ternary weights during the training phase to facilitate hardware implementations of RNNs.

188, TITLE: Learning concise representations for regression by evolving networks of trees  
<https://openreview.net/forum?id=Hke-JhA9Y7>  
AUTHORS: William La Cava, Tilak Raj Singh, James Taggart, Srinivas Suri  
TL;DR: Representing the network architecture as a set of syntax trees and optimizing their structure leads to accurate and concise regression models.  
HIGHLIGHT: We propose and study a method for learning interpretable representations for the task of regression.

189, TITLE: Efficient Training on Very Large Corpora via Gramian Estimation  
<https://openreview.net/forum?id=Hke20iA9Y7>  
AUTHORS: Walid Krichene, Nicolas Mayoraz, Steffen Rendle, Li Zhang, Xinyang Yi, Lichan Hong, Ed Chi, John Anderson

TL;DR: We develop efficient methods to train neural embedding models with a dot-product structure, by reformulating the objective function in terms of generalized Gram matrices, and maintaining estimates of those matrices.

HIGHLIGHT: These models are typically trained using SGD with random sampling of unobserved pairs, with a sample size that grows quadratically with the corpus size, making it expensive to scale. We propose new efficient methods to train these models without having to sample unobserved pairs.

190, TITLE: MAE: Mutual Posterior-Divergence Regularization for Variational AutoEncoders

<https://openreview.net/forum?id=Hke4I2AcKQ>

AUTHORS: Xuezhe Ma, Chunting Zhou, Eduard Hovy

TL;DR: We develop efficient methods to train neural embedding models with a dot-product structure, by reformulating the objective function in terms of generalized Gram matrices, and maintaining estimates of those matrices.

HIGHLIGHT: In this work, we introduce mutual posterior-divergence regularization, a novel regularization that is able to control the geometry of the latent space to accomplish meaningful representation learning, while achieving comparable or superior capability of density estimation. Experiments on three image benchmark datasets demonstrate that, when equipped with powerful decoders, our model performs well both on density estimation and representation learning.

191, TITLE: Residual Non-local Attention Networks for Image Restoration

<https://openreview.net/forum?id=HkeGhoA5FX>

AUTHORS: Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, Yun Fu

TL;DR: New state-of-the-art framework for image restoration

HIGHLIGHT: In this paper, we propose a residual non-local attention network for high-quality image restoration.

192, TITLE: Meta-Learning For Stochastic Gradient MCMC

<https://openreview.net/forum?id=HkeoOo09YX>

AUTHORS: Wenbo Gong, Yingzhen Li, Jos? Miguel Hern?andez-Lobato

TL;DR: This paper proposes a method to automate the design of stochastic gradient MCMC proposal using meta learning approach.

HIGHLIGHT: This paper presents the first meta-learning algorithm that allows automated design for the underlying continuous dynamics of an SG-MCMC sampler.

193, TITLE: Systematic Generalization: What Is Required and Can It Be Learned?

<https://openreview.net/forum?id=HkezXnA9YX>

AUTHORS: Dzmitry Bahdanau\*, Shikhar Murty\*, Michael Noukhovitch, Thien Huu Nguyen, Harm de Vries, Aaron Courville

TL;DR: We show that modular structured models are the best in terms of systematic generalization and that their end-to-end versions don't generalize as well.

HIGHLIGHT: Numerous models for grounded language understanding have been recently proposed, including (i) generic models that can be easily adapted to any given task and (ii) intuitively appealing modular models that require background knowledge to be instantiated.

194, TITLE: Efficient Lifelong Learning with A-GEM

[https://openreview.net/forum?id=Hkf2\\_sC5FX](https://openreview.net/forum?id=Hkf2_sC5FX)

AUTHORS: Arslan Chaudhry, Marc? Aurelio Ranzato, Marcus Rohrbach, Mohamed Elhoseiny

TL;DR: An efficient lifelong learning algorithm that provides a better trade-off between accuracy and time/ memory complexity compared to other algorithms.

HIGHLIGHT: In this work, we investigate the efficiency of current lifelong approaches, in terms of sample complexity, computational and memory cost.

195, TITLE: Multi-step Retriever-Reader Interaction for Scalable Open-domain Question Answering  
<https://openreview.net/forum?id=HkfPSH05K7>

AUTHORS: Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Andrew McCallum

TL;DR: Paragraph retriever and machine reader interacts with each other via reinforcement learning to yield large improvements on open domain datasets

HIGHLIGHT: This paper introduces a new framework for open-domain question answering in which the retriever and the reader `\emph{iteratively interact}` with each other.

196, TITLE: Double Viterbi: Weight Encoding for High Compression Ratio and Fast On-Chip Reconstruction for Deep Neural Network

<https://openreview.net/forum?id=HkfYOoCcYX>

AUTHORS: Daehyun Ahn, Dongsoo Lee, Taesu Kim, Jae-Joon Kim

TL;DR: We present a new weight encoding scheme which enables high compression ratio and fast sparse-to-dense matrix conversion.

HIGHLIGHT: In this paper, we propose a new sparse matrix format in order to enable a highly parallel decoding process of the entire sparse matrix.

197, TITLE: Overcoming the Disentanglement vs Reconstruction Trade-off via Jacobian Supervision  
<https://openreview.net/forum?id=Hkg4W2AcFm>

AUTHORS: Jos? Lezama

TL;DR: A method for learning image representations that are good for both disentangling factors of variation and obtaining faithful reconstructions.

HIGHLIGHT: In this work, we propose to overcome this trade-off by progressively growing the dimension of the latent code, while constraining the Jacobian of the output image with respect to the disentangled variables to remain the same.

198, TITLE: RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space

<https://openreview.net/forum?id=HkgEQnRqYQ>

AUTHORS: Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, Jian Tang

TL;DR: A new state-of-the-art approach for knowledge graph embedding.

HIGHLIGHT: In this paper, we present a new approach for knowledge graph embedding called RotatE, which is able to model and infer various relation patterns including: symmetry/antisymmetry, inversion, and composition.

199, TITLE: Guiding Policies with Language via Meta-Learning

<https://openreview.net/forum?id=HkgSEnA5KQ>

AUTHORS: John D. Co-Reyes, Abhishek Gupta, Suvansh Sanjeev, Nick Altieri, Jacob Andreas, John DeNero, Pieter Abbeel, Sergey Levine

TL;DR: We propose a meta-learning method for interactively correcting policies with natural language.

HIGHLIGHT: In this work, we propose an interactive formulation of the task specification problem, where iterative language corrections are provided to an autonomous agent, guiding it in acquiring the desired skill.

200, TITLE: AdaShift: Decorrelation and Convergence of Adaptive Learning Rate Methods

<https://openreview.net/forum?id=HkgTkRcKQ>

AUTHORS: Zhiming Zhou\*, Qingru Zhang\*, Guansong Lu, Hongwei Wang, Weinan Zhang, Yong Yu

TL;DR: We analysis and solve the non-convergence issue of Adam.

**HIGHLIGHT:** In this paper, we provide a new insight into the non-convergence issue of Adam as well as other adaptive learning rate methods.

201, **TITLE:** AD-VAT: An Asymmetric Dueling mechanism for learning Visual Active Tracking  
<https://openreview.net/forum?id=HkgYmhR9KX>  
**AUTHORS:** Fangwei Zhong, Peng Sun, Wenhan Luo, Tingyun Yan, Yizhou Wang  
**TL;DR:** We propose AD-VAT, where the tracker and the target object, viewed as two learnable agents, are opponents and can mutually enhance during training.  
**HIGHLIGHT:** To learn a robust tracker for VAT, in this paper, we propose a novel adversarial RL method which adopts an Asymmetric Dueling mechanism, referred to as AD-VAT.

202, **TITLE:** Marginal Policy Gradients: A Unified Family of Estimators for Bounded Action Spaces with Applications  
<https://openreview.net/forum?id=HkgqFiAcFm>  
**AUTHORS:** Carson Eisenach, Haichuan Yang, Ji Liu, Han Liu  
**TL;DR:** We propose AD-VAT, where the tracker and the target object, viewed as two learnable agents, are opponents and can mutually enhance during training.  
**HIGHLIGHT:** To this end we introduce the angular policy gradient (APG), a stochastic policy gradient method for directional control.

203, **TITLE:** On Self Modulation for Generative Adversarial Networks  
<https://openreview.net/forum?id=Hkl5aoR5tm>  
**AUTHORS:** Ting Chen, Mario Lucic, Neil Houlsby, Sylvain Gelly  
**TL;DR:** A simple GAN modification that improves performance across many losses, architectures, regularization schemes, and datasets.  
**HIGHLIGHT:** We propose and study an architectural modification, self-modulation, which improves GAN performance across different data sets, architectures, losses, regularizers, and hyperparameter settings.

204, **TITLE:** Off-Policy Evaluation and Learning from Logged Bandit Feedback: Error Reduction via Surrogate Policy  
<https://openreview.net/forum?id=HklKui0ct7>  
**AUTHORS:** Yuan Xie, Boyi Liu, Qiang Liu, Zhaoran Wang, Yuan Zhou, Jian Peng  
**TL;DR:** A simple GAN modification that improves performance across many losses, architectures, regularization schemes, and datasets.  
**HIGHLIGHT:** In this work, we introduce a new approach named Maximum Likelihood Inverse Propensity Scoring (MLIPS) for batch learning from logged bandit feedback.

205, **TITLE:** Subgradient Descent Learns Orthogonal Dictionaries  
<https://openreview.net/forum?id=HklSf3CqKm>  
**AUTHORS:** Yu Bai, Qijia Jiang, Ju Sun  
**TL;DR:** Efficient dictionary learning by L1 minimization via a novel analysis of the non-convex non-smooth geometry.  
**HIGHLIGHT:** We show that a subgradient descent algorithm, with random initialization, can recover orthogonal dictionaries on a natural nonsmooth, nonconvex L1 minimization formulation of the problem, under mild statistical assumption on the data.

206, **TITLE:** ClariNet: Parallel Wave Generation in End-to-End Text-to-Speech  
<https://openreview.net/forum?id=HklY120cYm>  
**AUTHORS:** Wei Ping, Kainan Peng, Jitong Chen

TL;DR: Efficient dictionary learning by L1 minimization via a novel analysis of the non-convex non-smooth geometry.

HIGHLIGHT: In this work, we propose a new solution for parallel wave generation by WaveNet.

207, TITLE: MARGINALIZED AVERAGE ATTENTIONAL NETWORK FOR WEAKLY-SUPERVISED LEARNING

<https://openreview.net/forum?id=HkljioCcFQ>

AUTHORS: Yuan Yuan, Yueming Lyu, Xi Shen, Ivor W. Tsang, Dit-Yan Yeung

TL;DR: A novel marginalized average attentional network for weakly-supervised temporal action localization

HIGHLIGHT: To alleviate this issue, we propose a marginalized average attentional network (MAAN) to suppress the dominant response of the most salient regions in a principled manner.

208, TITLE: Towards GAN Benchmarks Which Require Generalization

<https://openreview.net/forum?id=HkxKH2AcFm>

AUTHORS: Ishaan Gulrajani, Colin Raffel, Luke Metz

TL;DR: We argue that GAN benchmarks must require a large sample from the model to penalize memorization and investigate whether neural network divergences have this property.

HIGHLIGHT: For many evaluation metrics commonly used as benchmarks for unconditional image generation, trivially memorizing the training set attains a better score than models which are considered state-of-the-art; we consider this problematic. We clarify a necessary condition for an evaluation metric not to behave this way: estimating the function must require a large sample from the model.

209, TITLE: A Closer Look at Few-shot Classification

<https://openreview.net/forum?id=HkxLXnAcFQ>

AUTHORS: Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, Jia-Bin Huang

TL;DR: A detailed empirical study in few-shot classification that revealing challenges in standard evaluation setting and showing a new direction.

HIGHLIGHT: In this paper, we present 1) a consistent comparative analysis of several representative few-shot classification algorithms, with results showing that deeper backbones significantly reduce the gap across methods including the baseline, 2) a slightly modified baseline method that surprisingly achieves competitive performance when compared with the state-of-the-art on both the mini-ImageNet and the CUB datasets, and 3) a new experimental setting for evaluating the cross-domain generalization ability for few-shot classification algorithms.

210, TITLE: Meta-Learning Probabilistic Inference for Prediction

<https://openreview.net/forum?id=HkxStoC5F7>

AUTHORS: Jonathan Gordon, John Bronskill, Matthias Bauer, Sebastian Nowozin, Richard Turner

TL;DR: Novel framework for meta-learning that unifies and extends a broad class of existing few-shot learning methods. Achieves strong performance on few-shot learning benchmarks without requiring iterative test-time inference.

HIGHLIGHT: This paper introduces a new framework for data efficient and versatile learning.

211, TITLE: Deep reinforcement learning with relational inductive biases

<https://openreview.net/forum?id=HkxaFoC9KQ>

AUTHORS: 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

TL;DR: Relational inductive biases improve out-of-distribution generalization capacities in model-free reinforcement learning agents

**HIGHLIGHT:** We introduce an approach for augmenting model-free deep reinforcement learning agents with a mechanism for relational reasoning over structured representations, which improves performance, learning efficiency, generalization, and interpretability.

212, **TITLE:** Relaxed Quantization for Discretized Neural Networks  
<https://openreview.net/forum?id=HkxjYoCqKX>  
**AUTHORS:** Christos Louizos, Matthias Reisser, Tijmen Blankevoort, Efstratios Gavves, Max Welling  
**TL;DR:** We introduce a technique that allows for gradient based training of quantized neural networks.  
**HIGHLIGHT:** In order to train networks that can be effectively discretized without loss of performance, we introduce a differentiable quantization procedure.

213, **TITLE:** Tree-Structured Recurrent Switching Linear Dynamical Systems for Multi-Scale Modeling  
<https://openreview.net/forum?id=HkzRQhR9YX>  
**AUTHORS:** Josue Nassar, Scott Linderman, Monica Bugallo, Il Memming Park  
**TL;DR:** We introduce a technique that allows for gradient based training of quantized neural networks.  
**HIGHLIGHT:** To fit this model, we present a fully-Bayesian sampling procedure using Polya-Gamma data augmentation to allow for fast and conjugate Gibbs sampling.

214, **TITLE:** STCN: Stochastic Temporal Convolutional Networks  
<https://openreview.net/forum?id=HkzSQhCcK7>  
**AUTHORS:** Emre Aksan, Otmar Hilliges  
**TL;DR:** We combine the computational advantages of temporal convolutional architectures with the expressiveness of stochastic latent variables.  
**HIGHLIGHT:** In this work, we propose stochastic temporal convolutional networks (STCNs), a novel architecture that combines the computational advantages of temporal convolutional networks (TCN) with the representational power and robustness of stochastic latent spaces.

215, **TITLE:** Soft Q-Learning with Mutual-Information Regularization  
<https://openreview.net/forum?id=HyEtjoCqFX>  
**AUTHORS:** Jordi Grau-Moya, Felix Leibfried, Peter Vrancx  
**TL;DR:** We combine the computational advantages of temporal convolutional architectures with the expressiveness of stochastic latent variables.  
**HIGHLIGHT:** We propose a reinforcement learning (RL) algorithm that uses mutual-information regularization to optimize a prior action distribution for better performance and exploration.

216, **TITLE:** On the Turing Completeness of Modern Neural Network Architectures  
<https://openreview.net/forum?id=HyGBdo0qFm>  
**AUTHORS:** Jorge P?rez, Javier Marinkovic, Pablo Barcel?  
**TL;DR:** We show that the Transformer architecture and the Neural GPU are Turing complete.  
**HIGHLIGHT:** We show both models to be Turing complete exclusively based on their capacity to compute and access internal dense representations of the data.

217, **TITLE:** Improving Differentiable Neural Computers Through Memory Masking, De-allocation, and Link Distribution Sharpness Control  
<https://openreview.net/forum?id=HyGEM3C9KQ>  
**AUTHORS:** Robert Csordas, Juergen Schmidhuber  
**TL;DR:** We show that the Transformer architecture and the Neural GPU are Turing complete.

**HIGHLIGHT:** An analysis of its internal activation patterns reveals three problems: Most importantly, the lack of key-value separation makes the address distribution resulting from content-based look-up noisy and flat, since the value influences the score calculation, although only the key should.

218, **TITLE:** Evaluating Robustness of Neural Networks with Mixed Integer Programming

<https://openreview.net/forum?id=HyGldiRqtm>

**AUTHORS:** Vincent Tjeng, Kai Y. Xiao, Russ Tedrake

**TL;DR:** We efficiently verify the robustness of deep neural models with over 100,000 ReLUs, certifying more samples than the state-of-the-art and finding more adversarial examples than a strong first-order attack.

**HIGHLIGHT:** We achieve this computational speedup via tight formulations for non-linearities, as well as a novel presolve algorithm that makes full use of all information available.

219, **TITLE:** Random mesh projectors for inverse problems

<https://openreview.net/forum?id=HyGcghRct7>

**AUTHORS:** Konik Kothari\*, Sidharth Gupta\*, Maarten v. de Hoop, Ivan Dokmanic

**TL;DR:** We solve ill-posed inverse problems with scarce ground truth examples by estimating an ensemble of random projections of the model instead of the model itself.

**HIGHLIGHT:** We propose a new learning-based approach to solve ill-posed inverse problems in imaging.

220, **TITLE:** Multi-Agent Dual Learning

<https://openreview.net/forum?id=HyGhN2A5tm>

**AUTHORS:** Yiren Wang, Yingce Xia, Tianyu He, Fei Tian, Tao Qin, ChengXiang Zhai, Tie-Yan Liu

**TL;DR:** We solve ill-posed inverse problems with scarce ground truth examples by estimating an ensemble of random projections of the model instead of the model itself.

**HIGHLIGHT:** In this paper, we extend this framework by introducing multiple primal and dual models, and propose the multi-agent dual learning framework.

221, **TITLE:** Complement Objective Training

<https://openreview.net/forum?id=HyM7AiA5YX>

**AUTHORS:** Hao-Yun Chen, Pei-Hsin Wang, Chun-Hao Liu, Shih-Chieh Chang, Jia-Yu Pan, Yu-Ting Chen, Wei Wei, Da-Cheng Juan

**TL;DR:** We propose Complement Objective Training (COT), a new training paradigm that optimizes both the primary and complement objectives for effectively learning the parameters of neural networks.

**HIGHLIGHT:** We conduct extensive experiments on multiple tasks ranging from computer vision to natural language understanding.

222, **TITLE:** Mode Normalization

<https://openreview.net/forum?id=HyN-M2Rctm>

**AUTHORS:** Lucas Deecke, Iain Murray, Hakan Bilen

**TL;DR:** We present a novel normalization method for deep neural networks that is robust to multi-modalities in intermediate feature distributions.

**HIGHLIGHT:** As a remedy, we propose a more flexible approach: by extending the normalization to more than a single mean and variance, we detect modes of data on-the-fly, jointly normalizing samples that share common features.

223, **TITLE:** Detecting Egregious Responses in Neural Sequence-to-sequence Models

<https://openreview.net/forum?id=HyNA5iRcFQ>

**AUTHORS:** Tianxing He, James Glass

TL;DR: This paper aims to provide an empirical answer to the question of whether well-trained dialogue response model can output malicious responses.

HIGHLIGHT: In this work, we attempt to answer a critical question: whether there exists some input sequence that will cause a well-trained discrete-space neural network sequence-to-sequence (seq2seq) model to generate egregious outputs (aggressive, malicious, attacking, etc.).

224, TITLE: Learning Actionable Representations with Goal Conditioned Policies

<https://openreview.net/forum?id=Hye9lnCct7>

AUTHORS: Dibya Ghosh, Abhishek Gupta, Sergey Levine

TL;DR: Learning state representations which capture factors necessary for control

HIGHLIGHT: In this paper, we instead aim to learn functionally salient representations: representations that are not necessarily complete in terms of capturing all factors of variation in the observation space, but rather aim to capture those factors of variation that are important for decision making -- that are "actionable".

225, TITLE: Verification of Non-Linear Specifications for Neural Networks

<https://openreview.net/forum?id=HyeFAsRctQ>

AUTHORS: Chongli Qin, Krishnamurthy (Dj) Dvijotham, Brendan O'Donoghue, Rudy Bunel, Robert Stanforth, Sven Gowal, Jonathan Uesato, Grzegorz Swirszcz, Pushmeet Kohli

TL;DR: Learning state representations which capture factors necessary for control

HIGHLIGHT: In this paper, we extend verification algorithms to be able to certify richer properties of neural networks.

226, TITLE: Generating Liquid Simulations with Deformation-aware Neural Networks

<https://openreview.net/forum?id=HyeGBj09Fm>

AUTHORS: Lukas Prantl, Boris Bonev, Nils Thuerey

TL;DR: Learning weighting and deformations of space-time data sets for highly efficient approximations of liquid behavior.

HIGHLIGHT: We propose a novel approach for deformation-aware neural networks that learn the weighting and synthesis of dense volumetric deformation fields.

227, TITLE: DyRep: Learning Representations over Dynamic Graphs

<https://openreview.net/forum?id=HyePrhR5KX>

AUTHORS: Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, Hongyuan Zha

TL;DR: Models Representation Learning over dynamic graphs as latent hidden process bridging two observed processes of Topological Evolution of and Interactions on dynamic graphs.

HIGHLIGHT: We present DyRep - a novel modeling framework for dynamic graphs that posits representation learning as a latent mediation process bridging two observed processes namely -- dynamics of the network (realized as topological evolution) and dynamics on the network (realized as activities between nodes).

228, TITLE: Trellis Networks for Sequence Modeling

<https://openreview.net/forum?id=HyeVtoRqtQ>

AUTHORS: Shaojie Bai, J. Zico Kolter, Vladlen Koltun

TL;DR: Trellis networks are a new sequence modeling architecture that bridges recurrent and convolutional models and sets a new state of the art on word- and character-level language modeling.

HIGHLIGHT: We present trellis networks, a new architecture for sequence modeling.

229, TITLE: Scalable Unbalanced Optimal Transport using Generative Adversarial Networks

<https://openreview.net/forum?id=HyexAiA5Fm>

AUTHORS: Karren D. Yang, Caroline Uhler  
TL;DR: We propose new methodology for unbalanced optimal transport using generative adversarial networks.  
HIGHLIGHT: In this paper, we present a scalable method for unbalanced optimal transport (OT) based on the generative-adversarial framework.

230, TITLE: Solving the Rubik's Cube with Approximate Policy Iteration  
<https://openreview.net/forum?id=Hyfn2jCcKm>  
AUTHORS: Stephen McAleer, Forest Agostinelli, Alexander Shmakov, Pierre Baldi  
TL;DR: We solve the Rubik's Cube with pure reinforcement learning  
HIGHLIGHT: We introduce Autodidactic Iteration: an API algorithm that overcomes the problem of sparse rewards by training on a distribution of states that allows the reward to propagate from the goal state to states farther away.

231, TITLE: Variance Reduction for Reinforcement Learning in Input-Driven Environments  
<https://openreview.net/forum?id=Hyg1G2AqtQ>  
AUTHORS: Hongzi Mao, Shaileshh Bojja Venkatakrisnan, Malte Schwarzkopf, Mohammad Alizadeh  
TL;DR: For environments dictated partially by external input processes, we derive an input-dependent baseline that provably reduces the variance for policy gradient methods and improves the policy performance in a wide range of RL tasks.  
HIGHLIGHT: We consider reinforcement learning in input-driven environments, where an exogenous, stochastic input process affects the dynamics of the system.

232, TITLE: Model-Predictive Policy Learning with Uncertainty Regularization for Driving in Dense Traffic  
<https://openreview.net/forum?id=HygQBn0cYm>  
AUTHORS: Mikael Henaff, Alfredo Canziani, Yann LeCun  
TL;DR: A model-based RL approach which uses a differentiable uncertainty penalty to learn driving policies from purely observational data.  
HIGHLIGHT: In this work, we propose to train a policy while explicitly penalizing the mismatch between these two distributions over a fixed time horizon.

233, TITLE: GAN Dissection: Visualizing and Understanding Generative Adversarial Networks  
[https://openreview.net/forum?id=Hyg\\_X2C5FX](https://openreview.net/forum?id=Hyg_X2C5FX)  
AUTHORS: David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, Antonio Torralba  
TL;DR: GAN representations are examined in detail, and sets of representation units are found that control the generation of semantic concepts in the output.  
HIGHLIGHT: Answering such questions could enable us to develop new insights and better models. In this work, we present an analytic framework to visualize and understand GANs at the unit-, object-, and scene-level.

234, TITLE: Improving MMD-GAN Training with Repulsive Loss Function  
<https://openreview.net/forum?id=HygqjR9Km>  
AUTHORS: Wei Wang, Yuan Sun, Saman Halgamuge  
TL;DR: Rearranging the terms in maximum mean discrepancy yields a much better loss function for the discriminator of generative adversarial nets  
HIGHLIGHT: To address this issue, we propose a repulsive loss function to actively learn the difference among the real data by simply rearranging the terms in MMD.

235, TITLE: Deterministic PAC-Bayesian generalization bounds for deep networks via generalizing noise-resilience

<https://openreview.net/forum?id=Hygn2o0qKX>

AUTHORS: Vaishnavh Nagarajan, Zico Kolter

TL;DR: We provide a PAC-Bayes based generalization guarantee for uncompressed, deterministic deep networks by generalizing noise-resilience of the network on the training data to the test data.

HIGHLIGHT: In this work, we present a general PAC-Bayesian framework that leverages this observation to provide a bound on the original network learned -- a network that is deterministic and uncompressed.

236, TITLE: Recall Traces: Backtracking Models for Efficient Reinforcement Learning

<https://openreview.net/forum?id=HygsfnR9Ym>

AUTHORS: Anirudh Goyal, Philemon Brakel, William Fedus, Soumye Singhal, Timothy Lillicrap, Sergey Levine, Hugo Larochelle, Yoshua Bengio

TL;DR: A backward model of previous (state, action) given the next state, i.e.  $P(s_t, a_t | s_{t+1})$ , can be used to simulate additional trajectories terminating at states of interest! Improves RL learning efficiency.

HIGHLIGHT: Hence, we may want to preferentially train on those high-reward states and the probable trajectories leading to them. To this end, we advocate for the use of a `backtracking model` that predicts the preceding states that terminate at a given high-reward state.

237, TITLE: Stable Recurrent Models

<https://openreview.net/forum?id=Hygxb2CqKm>

AUTHORS: John Miller, Moritz Hardt

TL;DR: Stable recurrent models can be approximated by feed-forward networks and empirically perform as well as unstable models on benchmark tasks.

HIGHLIGHT: In this work, we conduct a thorough investigation of stable recurrent models.

238, TITLE: The Limitations of Adversarial Training and the Blind-Spot Attack

<https://openreview.net/forum?id=HylTBhA5tQ>

AUTHORS: Huan Zhang\*, Hongge Chen\*, Zhao Song, Duane Boning, Inderjit S. Dhillon, Cho-Jui Hsieh

TL;DR: We show that even the strongest adversarial training methods cannot defend against adversarial examples crafted on slightly scaled and shifted test images.

HIGHLIGHT: In our paper, we shed some lights on the practicality and the hardness of adversarial training by showing that the effectiveness (robustness on test set) of adversarial training has a strong correlation with the distance between a test point and the manifold of training data embedded by the network.

239, TITLE: Efficiently testing local optimality and escaping saddles for ReLU networks

<https://openreview.net/forum?id=HylTXn0qYX>

AUTHORS: Chulhee Yun, Suvrit Sra, Ali Jadbabaie

TL;DR: A theoretical algorithm for testing local optimality and extracting descent directions at nondifferentiable points of empirical risks of one-hidden-layer ReLU networks.

HIGHLIGHT: We provide a theoretical algorithm for checking local optimality and escaping saddles at nondifferentiable points of empirical risks of two-layer ReLU networks.

240, TITLE: ProxylessNAS: Direct Neural Architecture Search on Target Task and Hardware

<https://openreview.net/forum?id=HylVB3AqYm>

AUTHORS: Han Cai, Ligeng Zhu, Song Han

TL;DR: Proxy-less neural architecture search for directly learning architectures on large-scale target task (ImageNet) while reducing the cost to the same level of normal training.

HIGHLIGHT: In this paper, we present ProxylessNAS that can directly learn the architectures for large-scale target tasks and target hardware platforms.

241, TITLE: Hierarchical Reinforcement Learning via Advantage-Weighted Information Maximization

[https://openreview.net/forum?id=Hyl\\_vjC5KQ](https://openreview.net/forum?id=Hyl_vjC5KQ)

AUTHORS: Takayuki Osa, Voot Tangkaratt, Masashi Sugiyama

TL;DR: This paper presents a hierarchical reinforcement learning framework based on deterministic option policies and mutual information maximization.

HIGHLIGHT: In this paper, we propose an HRL method that learns a latent variable of a hierarchical policy using mutual information maximization.

242, TITLE: Generalizable Adversarial Training via Spectral Normalization

<https://openreview.net/forum?id=Hyx4knR9Ym>

AUTHORS: Farzan Farnia, Jesse Zhang, David Tse

TL;DR: This paper presents a hierarchical reinforcement learning framework based on deterministic option policies and mutual information maximization.

HIGHLIGHT: In this work, we extend the notion of margin loss to adversarial settings and bound the generalization error for DNNs trained under several well-known gradient-based attack schemes, motivating an effective regularization scheme based on spectral normalization of the DNN's weight matrices.

243, TITLE: Adversarial Domain Adaptation for Stable Brain-Machine Interfaces

<https://openreview.net/forum?id=Hyx6Bi0qYm>

AUTHORS: Ali Farshchian, Juan A. Gallego, Joseph P. Cohen, Yoshua Bengio, Lee E. Miller, Sara A. Solla

TL;DR: We implement an adversarial domain adaptation network to stabilize a fixed Brain-Machine Interface against gradual changes in the recorded neural signals.

HIGHLIGHT: Here, we introduce a new computational approach that decodes movement intent from a low-dimensional latent representation of the neural data.

244, TITLE: Deep Online Learning Via Meta-Learning: Continual Adaptation for Model-Based RL

<https://openreview.net/forum?id=HyxAfnA5tm>

AUTHORS: Anusha Nagabandi, Chelsea Finn, Sergey Levine

TL;DR: We implement an adversarial domain adaptation network to stabilize a fixed Brain-Machine Interface against gradual changes in the recorded neural signals.

HIGHLIGHT: The goal in this paper is to develop a method for continual online learning from an incoming stream of data, using deep neural network models.

245, TITLE: Deep Anomaly Detection with Outlier Exposure

<https://openreview.net/forum?id=HyxCxhRcY7>

AUTHORS: Dan Hendrycks, Mantas Mazeika, Thomas Dietterich

TL;DR: OE teaches anomaly detectors to learn heuristics for detecting unseen anomalies; experiments are in classification, density estimation, and calibration in NLP and vision settings; we do not tune on test distribution samples, unlike previous work

HIGHLIGHT: We propose leveraging these data to improve deep anomaly detection by training anomaly detectors against an auxiliary dataset of outliers, an approach we call Outlier Exposure (OE).

246, TITLE: Contingency-Aware Exploration in Reinforcement Learning

<https://openreview.net/forum?id=HyxGB2AcY7>

AUTHORS: Jongwook Choi, Yijie Guo, Marcin Moczulski, Junhyuk Oh, Neal Wu, Mohammad Norouzi, Honglak Lee

TL;DR: We investigate contingency-awareness and controllable aspects in exploration and achieve state-of-the-art performance on Montezuma's Revenge without expert demonstrations.

HIGHLIGHT: In this study, we develop an attentive dynamics model (ADM) that discovers controllable elements of the observations, which are often associated with the location of the character in Atari games.

247, TITLE: Context-adaptive Entropy Model for End-to-end Optimized Image Compression

<https://openreview.net/forum?id=HyxKIiAqYQ>

AUTHORS: Jooyoung Lee, Seunghyun Cho, Seung-Kwon Beack

TL;DR: Context-adaptive entropy model for use in end-to-end optimized image compression, which significantly improves compression performance

HIGHLIGHT: We propose a context-adaptive entropy model for use in end-to-end optimized image compression.

248, TITLE: Variational Discriminator Bottleneck: Improving Imitation Learning, Inverse RL, and GANs by Constraining Information Flow

<https://openreview.net/forum?id=HyxPx3R9tm>

AUTHORS: Xue Bin Peng, Angjoo Kanazawa, Sam Toyer, Pieter Abbeel, Sergey Levine

TL;DR: Regularizing adversarial learning with an information bottleneck, applied to imitation learning, inverse reinforcement learning, and generative adversarial networks.

HIGHLIGHT: In this work, we propose a simple and general technique to constrain information flow in the discriminator by means of an information bottleneck.

249, TITLE: Meta-learning with differentiable closed-form solvers

<https://openreview.net/forum?id=HyxnZh0ct7>

AUTHORS: Luca Bertinetto, Joao F. Henriques, Philip Torr, Andrea Vedaldi

TL;DR: We propose a meta-learning approach for few-shot classification that achieves strong performance at high-speed by back-propagating through the solution of fast solvers, such as ridge regression or logistic regression.

HIGHLIGHT: Adapting deep networks to new concepts from a few examples is challenging, due to the high computational requirements of standard fine-tuning procedures. Most work on few-shot learning has thus focused on simple learning techniques for adaptation, such as nearest neighbours or gradient descent. Nonetheless, the machine learning literature contains a wealth of methods that learn non-deep models very efficiently. In this paper, we propose to use these fast convergent methods as the main adaptation mechanism for few-shot learning. The main idea is to teach a deep network to use standard machine learning tools, such as ridge regression, as part of its own internal model, enabling it to quickly adapt to novel data. This requires back-propagating errors through the solver steps. While normally the cost of the matrix operations involved in such a process would be significant, by using the Woodbury identity we can make the small number of examples work to our advantage. We propose both closed-form and iterative solvers, based on ridge regression and logistic regression components. Our methods constitute a simple and novel approach to the problem of few-shot learning and achieve performance competitive with or superior to the state of the art on three benchmarks.

250, TITLE: Learning Self-Imitating Diverse Policies

<https://openreview.net/forum?id=HyxzRsR9Y7>

AUTHORS: Tanmay Gangwani, Qiang Liu, Jian Peng

TL;DR: Policy optimization by using past good rollouts from the agent; learning shaped rewards via divergence minimization; SVPG with JS-kernel for population-based exploration.

HIGHLIGHT: In this work, we introduce a self-imitation learning algorithm that exploits and explores well in the sparse and episodic reward settings.

251, TITLE: ProxQuant: Quantized Neural Networks via Proximal Operators  
<https://openreview.net/forum?id=HyzMhCcK7>  
AUTHORS: Yu Bai, Yu-Xiang Wang, Edo Liberty  
TL;DR: A principled framework for model quantization using the proximal gradient method, with empirical evaluation and theoretical convergence analyses.  
HIGHLIGHT: Despite its empirical success, little is understood about why the straight-through gradient method works. Building upon a novel observation that the straight-through gradient method is in fact identical to the well-known Nesterov's dual-averaging algorithm on a quantization constrained optimization problem, we propose a more principled alternative approach, called ProxQuant, that formulates quantized network training as a regularized learning problem instead and optimizes it via the prox-gradient method.

252, TITLE: Universal Transformers  
<https://openreview.net/forum?id=HyzdRiR9Y7>  
AUTHORS: Mostafa Dehghani, Stephan Gouws, Oriol Vinyals, Jakob Uszkoreit, Lukasz Kaiser  
TL;DR: We introduce the Universal Transformer, a self-attentive parallel-in-time recurrent sequence model that outperforms Transformers and LSTMs on a wide range of sequence-to-sequence tasks, including machine translation.  
HIGHLIGHT: We propose the Universal Transformer (UT), a parallel-in-time self-attentive recurrent sequence model which can be cast as a generalization of the Transformer model and which addresses these issues.

253, TITLE: Learning to Adapt in Dynamic, Real-World Environments through Meta-Reinforcement Learning  
<https://openreview.net/forum?id=HyztsoC5Y7>  
AUTHORS: Anusha Nagabandi, Ignasi Clavera, Simin Liu, Ronald S. Fearing, Pieter Abbeel, Sergey Levine, Chelsea Finn  
TL;DR: A model-based meta-RL algorithm that enables a real robot to adapt online in dynamic environments  
HIGHLIGHT: Given that it is impractical to train separate policies to accommodate all situations the agent may see in the real world, this work proposes to learn how to quickly and effectively adapt online to new tasks.

254, TITLE: L-Shapley and C-Shapley: Efficient Model Interpretation for Structured Data  
<https://openreview.net/forum?id=S1E3Ko09F7>  
AUTHORS: Jianbo Chen, Le Song, Martin J. Wainwright, Michael I. Jordan  
TL;DR: We develop two linear-complexity algorithms for model-agnostic model interpretation based on the Shapley value, in the settings where the contribution of features to the target is well-approximated by a graph-structured factorization.  
HIGHLIGHT: Methods based on the Shapley score have been proposed as a fair way of computing feature attributions, but incur an exponential complexity in the number of features.

255, TITLE: Discovery of Natural Language Concepts in Individual Units of CNNs  
<https://openreview.net/forum?id=S1EERs09YQ>  
AUTHORS: Seil Na, Yo Joong Choe, Dong-Hyun Lee, Gunhee Kim  
TL;DR: We show that individual units in CNN representations learned in NLP tasks are selectively responsive to natural language concepts.  
HIGHLIGHT: In order to quantitatively analyze such intriguing phenomenon, we propose a concept alignment method based on how units respond to replicated text.

256, TITLE: Towards the first adversarially robust neural network model on MNIST  
<https://openreview.net/forum?id=S1EH0sC9tX>  
AUTHORS: Lukas Schott, Jonas Rauber, Matthias Bethge, Wieland Brendel  
TL;DR: We show that individual units in CNN representations learned in NLP tasks are selectively responsive to natural language concepts.  
HIGHLIGHT: We present a novel robust classification model that performs analysis by synthesis using learned class-conditional data distributions.

257, TITLE: Discriminator Rejection Sampling  
<https://openreview.net/forum?id=S1GkToR5tm>  
AUTHORS: Samaneh Azadi, Catherine Olsson, Trevor Darrell, Ian Goodfellow, Augustus Odena  
TL;DR: We use a GAN discriminator to perform an approximate rejection sampling scheme on the output of the GAN generator.  
HIGHLIGHT: We propose a rejection sampling scheme using the discriminator of a GAN to approximately correct errors in the GAN generator distribution.

258, TITLE: Harmonic Unpaired Image-to-image Translation  
<https://openreview.net/forum?id=S1M6Z2Cctm>  
AUTHORS: Rui Zhang, Tomas Pfister, Jia Li  
TL;DR: Smooth regularization over sample graph for unpaired image-to-image translation results in significantly improved consistency  
HIGHLIGHT: In this paper, we take a manifold view of the problem by introducing a smoothness term over the sample graph to attain harmonic functions to enforce consistent mappings during the translation.

259, TITLE: Universal Successor Features Approximators  
<https://openreview.net/forum?id=S1VWjiRcKX>  
AUTHORS: Diana Borsa, Andre Barreto, John Quan, Daniel J. Mankowitz, Hado van Hasselt, Remi Munos, David Silver, Tom Schaul  
TL;DR: Smooth regularization over sample graph for unpaired image-to-image translation results in significantly improved consistency  
HIGHLIGHT: We discuss the challenges involved in training a USFA, its generalisation properties and demonstrate its practical benefits and transfer abilities on a large-scale domain in which the agent has to navigate in a first-person perspective three-dimensional environment.

260, TITLE: Gradient Descent Provably Optimizes Over-parameterized Neural Networks  
<https://openreview.net/forum?id=S1eK3i09YQ>  
AUTHORS: Simon S. Du, Xiyu Zhai, Barnabas Poczos, Aarti Singh  
TL;DR: We prove gradient descent achieves zero training loss with a linear rate on over-parameterized neural networks.  
HIGHLIGHT: One of the mysteries in the success of neural networks is randomly initialized first order methods like gradient descent can achieve zero training loss even though the objective function is non-convex and non-smooth.

261, TITLE: Opportunistic Learning: Budgeted Cost-Sensitive Learning from Data Streams  
<https://openreview.net/forum?id=S1eOH09KX>  
AUTHORS: Mohammad Kachuee, Orpaz Goldstein, Kimmo K?rkk?inen, Sajad Darabi, Majid Sarrafzadeh  
TL;DR: An online algorithm for cost-aware feature acquisition and prediction  
HIGHLIGHT: In this paper, we propose a novel approach for cost-sensitive feature acquisition at the prediction-time.

262, TITLE: DARTS: Differentiable Architecture Search  
<https://openreview.net/forum?id=S1eYHoC5FX>  
AUTHORS: Hanxiao Liu, Karen Simonyan, Yiming Yang  
TL;DR: We propose a differentiable architecture search algorithm for both convolutional and recurrent networks, achieving competitive performance with the state of the art using orders of magnitude less computation resources.  
HIGHLIGHT: This paper addresses the scalability challenge of architecture search by formulating the task in a differentiable manner.

263, TITLE: Feature-Wise Bias Amplification  
<https://openreview.net/forum?id=S1ecm2C9K7>  
AUTHORS: Klas Leino, Matt Fredrikson, Emily Black, Shayak Sen, Anupam Datta  
TL;DR: We propose a differentiable architecture search algorithm for both convolutional and recurrent networks, achieving competitive performance with the state of the art using orders of magnitude less computation resources.  
HIGHLIGHT: We present two new feature selection algorithms for mitigating bias amplification in linear models, and show how they can be adapted to convolutional neural networks efficiently.

264, TITLE: The relativistic discriminator: a key element missing from standard GAN  
<https://openreview.net/forum?id=S1erHoR5t7>  
AUTHORS: Alexia Jolicoeur-Martineau  
TL;DR: Improving the quality and stability of GANs using a relativistic discriminator; IPM GANs (such as WGAN-GP) are a special case.  
HIGHLIGHT: We generalize both approaches to non-standard GAN loss functions and we refer to them respectively as Relativistic GANs (RGANs) and Relativistic average GANs (RaGANs).

265, TITLE: Understanding and Improving Interpolation in Autoencoders via an Adversarial Regularizer  
<https://openreview.net/forum?id=S1fQSiCcYm>  
AUTHORS: David Berthelot\*, Colin Raffel\*, Aurko Roy, Ian Goodfellow  
TL;DR: We propose a regularizer that improves interpolation and autoencoders and show that it also improves the learned representation for downstream tasks.  
HIGHLIGHT: In this paper, we propose a regularization procedure which encourages interpolated outputs to appear more realistic by fooling a critic network which has been trained to recover the mixing coefficient from interpolated data.

266, TITLE: Quasi-hyperbolic momentum and Adam for deep learning  
<https://openreview.net/forum?id=S1fUpoR5FQ>  
AUTHORS: Jerry Ma, Denis Yarats  
TL;DR: Mix plain SGD and momentum (or do something similar with Adam) for great profit.  
HIGHLIGHT: We propose the quasi-hyperbolic momentum algorithm (QHM) as an extremely simple alteration of momentum SGD, averaging a plain SGD step with a momentum step.

267, TITLE: Local SGD Converges Fast and Communicates Little  
<https://openreview.net/forum?id=S1g2JnRcFX>  
AUTHORS: Sebastian U. Stich  
TL;DR: We prove that parallel local SGD achieves linear speedup with much lesser communication than parallel mini-batch SGD.

**HIGHLIGHT:** To overcome this communication bottleneck recent works propose to reduce the communication frequency.

268, **TITLE:** Learning Finite State Representations of Recurrent Policy Networks  
<https://openreview.net/forum?id=S1gOpsCctm>  
**AUTHORS:** Anurag Koul, Alan Fern, Sam Greydanus  
**TL;DR:** Extracting a finite state machine from a recurrent neural network via quantization for the purpose of interpretability with experiments on Atari.  
**HIGHLIGHT:** In this paper, we introduce a new technique, Quantized Bottleneck Insertion, to learn finite representations of these vectors and features.

269, **TITLE:** Multilingual Neural Machine Translation with Knowledge Distillation  
<https://openreview.net/forum?id=S1gUsoR9YX>  
**AUTHORS:** Xu Tan, Yi Ren, Di He, Tao Qin, Zhou Zhao, Tie-Yan Liu  
**TL;DR:** We proposed a knowledge distillation based method to boost the accuracy of multilingual neural machine translation.  
**HIGHLIGHT:** In this paper, we propose a distillation-based approach to boost the accuracy of multilingual machine translation.

270, **TITLE:** MisGAN: Learning from Incomplete Data with Generative Adversarial Networks  
<https://openreview.net/forum?id=S1IDV3RcKm>  
**AUTHORS:** Steven Cheng-Xian Li, Bo Jiang, Benjamin Marlin  
**TL;DR:** This paper presents a GAN-based framework for learning the distribution from high-dimensional incomplete data.  
**HIGHLIGHT:** In this paper, we present a GAN-based framework for learning from complex, high-dimensional incomplete data.

271, **TITLE:** A Direct Approach to Robust Deep Learning Using Adversarial Networks  
<https://openreview.net/forum?id=S1IIMn05F7>  
**AUTHORS:** Huaxia Wang, Chun-Nam Yu  
**TL;DR:** Jointly train an adversarial noise generating network with a classification network to provide better robustness to adversarial attacks.  
**HIGHLIGHT:** In this paper we propose a new defensive mechanism under the generative adversarial network-(GAN) framework.

272, **TITLE:** Combinatorial Attacks on Binarized Neural Networks  
<https://openreview.net/forum?id=S1ITEh09FQ>  
**AUTHORS:** Elias B Khalil, Amrita Gupta, Bistra Dilkina  
**TL;DR:** Gradient-based attacks on binarized neural networks are not effective due to the non-differentiability of such networks; Our IPROP algorithm solves this problem using integer optimization  
**HIGHLIGHT:** In this work, we study the problem of attacking a BNN through the lens of combinatorial and integer optimization.

273, **TITLE:** Exemplar Guided Unsupervised Image-to-Image Translation with Semantic Consistency  
<https://openreview.net/forum?id=S1ITg3RqYQ>  
**AUTHORS:** Liqian Ma, Xu Jia, Stamatios Georgoulis, Tinne Tuytelaars, Luc Van Gool  
**TL;DR:** We propose the Exemplar Guided & Semantically Consistent Image-to-image Translation (EGSC-IT) network which conditions the translation process on an exemplar image in the target domain.

**HIGHLIGHT:** To alleviate these issues, we propose the Exemplar Guided & Semantically Consistent Image-to-image Translation (EGSC-IT) network which conditions the translation process on an exemplar image in the target domain.

274, **TITLE:** ARM: Augment-REINFORCE-Merge Gradient for Stochastic Binary Networks  
<https://openreview.net/forum?id=S1lg0jAcYm>  
**AUTHORS:** Mingzhang Yin, Mingyuan Zhou  
**TL;DR:** An unbiased and low-variance gradient estimator for discrete latent variable models  
**HIGHLIGHT:** To backpropagate the gradients through stochastic binary layers, we propose the augment-REINFORCE-merge (ARM) estimator that is unbiased, exhibits low variance, and has low computational complexity.

275, **TITLE:** Building Dynamic Knowledge Graphs from Text using Machine Reading Comprehension  
<https://openreview.net/forum?id=S1lhbnRqF7>  
**AUTHORS:** Rajarshi Das, Tsendsuren Munkhdalai, Xingdi Yuan, Adam Trischler, Andrew McCallum  
**TL;DR:** An unbiased and low-variance gradient estimator for discrete latent variable models  
**HIGHLIGHT:** We propose a neural machine-reading model that constructs dynamic knowledge graphs from procedural text.

276, **TITLE:** Information asymmetry in KL-regularized RL  
<https://openreview.net/forum?id=S1lqMn05Ym>  
**AUTHORS:** Alexandre Galashov, Siddhant M. Jayakumar, Leonard Hasenclever, Dhruva Tirumala, Jonathan Schwarz, Guillaume Desjardins, Wojciech M. Czarnecki, Yee Whye Teh, Razvan Pascanu, Nicolas Heess  
**TL;DR:** Limiting state information for the default policy can improve performance, in a KL-regularized RL framework where both agent and default policy are optimized together  
**HIGHLIGHT:** In this work we study the possibility of leveraging such repeated structure to speed up and regularize learning.

277, **TITLE:** TimbreTron: A WaveNet(CycleGAN(CQT(Audio))) Pipeline for Musical Timbre Transfer  
<https://openreview.net/forum?id=S1lv305YQ>  
**AUTHORS:** Sicong Huang, Qiyang Li, Cem Anil, Xuchan Bao, Sageev Oore, Roger B. Grosse  
**TL;DR:** We present the TimbreTron, a pipeline for performing high-quality timbre transfer on musical waveforms using CQT-domain style transfer.  
**HIGHLIGHT:** In this work, we address the problem of musical timbre transfer, where the goal is to manipulate the timbre of a sound sample from one instrument to match another instrument while preserving other musical content, such as pitch, rhythm, and loudness.

278, **TITLE:** Whitening and Coloring Batch Transform for GANs  
<https://openreview.net/forum?id=S1x2Fj0qKQ>  
**AUTHORS:** Aliaksandr Siarohin, Enver Sangineto, Nicu Sebe  
**TL;DR:** We present the TimbreTron, a pipeline for performing high-quality timbre transfer on musical waveforms using CQT-domain style transfer.  
**HIGHLIGHT:** In this paper we propose to generalize both BN and cBN using a Whitening and Coloring based batch normalization.

279, **TITLE:** Learnable Embedding Space for Efficient Neural Architecture Compression

<https://openreview.net/forum?id=S1xLN3C9YX>

AUTHORS: Shengcao Cao, Xiaofang Wang, Kris M. Kitani

TL;DR: We propose a method to incrementally learn an embedding space over the domain of network architectures, to enable the careful selection of architectures for evaluation during compressed architecture search.

HIGHLIGHT: We propose a method to incrementally learn an embedding space over the domain of network architectures, to enable the careful selection of architectures for evaluation during compressed architecture search.

280, TITLE: On the Sensitivity of Adversarial Robustness to Input Data Distributions

<https://openreview.net/forum?id=S1xNEhR9KX>

AUTHORS: Gavin Weiguang Ding, Kry Yik Chau Lui, Xiaomeng Jin, Luyu Wang, Ruitong Huang

TL;DR: Robustness performance of PGD trained models are sensitive to semantics-preserving transformation of image datasets, which implies the trickiness of evaluation of robust learning algorithms in practice.

HIGHLIGHT: In this paper, we demonstrate an intriguing phenomenon about the most popular robust training method in the literature, adversarial training: Adversarial robustness, unlike clean accuracy, is sensitive to the input data distribution.

281, TITLE: Minimal Images in Deep Neural Networks: Fragile Object Recognition in Natural Images

<https://openreview.net/forum?id=S1xNb2A9YX>

AUTHORS: Sanjana Srivastava, Guy Ben-Yosef, Xavier Boix

TL;DR: Robustness performance of PGD trained models are sensitive to semantics-preserving transformation of image datasets, which implies the trickiness of evaluation of robust learning algorithms in practice.

HIGHLIGHT: In this paper, we demonstrate that such drops in accuracy due to changes of the visible region are a common phenomenon between humans and existing state-of-the-art deep neural networks (DNNs), and are much more prominent in DNNs.

282, TITLE: A Statistical Approach to Assessing Neural Network Robustness

<https://openreview.net/forum?id=S1xcx3C5FX>

AUTHORS: Stefan Webb, Tom Rainforth, Yee Whye Teh, M. Pawan Kumar

TL;DR: We introduce a statistical approach to assessing neural network robustness that provides an informative notion of how robust a network is, rather than just the conventional binary assertion of whether or not of property is violated.

HIGHLIGHT: We present a new approach to assessing the robustness of neural networks based on estimating the proportion of inputs for which a property is violated.

283, TITLE: Improving Sequence-to-Sequence Learning via Optimal Transport

<https://openreview.net/forum?id=S1xtAjR5tX>

AUTHORS: Liqun Chen, Yizhe Zhang, Ruiyi Zhang, Chenyang Tao, Zhe Gan, Haichao Zhang, Bai Li, Dinghan Shen, Changyou Chen, Lawrence Carin

TL;DR: We introduce a statistical approach to assessing neural network robustness that provides an informative notion of how robust a network is, rather than just the conventional binary assertion of whether or not of property is violated.

HIGHLIGHT: We present a novel solution to alleviate these issues.

284, TITLE: PATE-GAN: Generating Synthetic Data with Differential Privacy Guarantees

<https://openreview.net/forum?id=S1zk9iRqF7>

AUTHORS: James Jordon, Jinsung Yoon, Mihaela van der Schaar

TL;DR: We introduce a statistical approach to assessing neural network robustness that provides an informative notion of how robust a network is, rather than just the conventional binary assertion of whether or not of property is violated.

HIGHLIGHT: In this paper, we investigate a method for ensuring (differential) privacy of the generator of the Generative Adversarial Nets (GAN) framework.

285, TITLE: Integer Networks for Data Compression with Latent-Variable Models

<https://openreview.net/forum?id=S1zz2i0cY7>

AUTHORS: Johannes Ball?, Nick Johnston, David Minnen

TL;DR: We train variational models with quantized networks for computational determinism. This enables using them for cross-platform data compression.

HIGHLIGHT: We propose using integer networks as a universal solution to this problem, and demonstrate that they enable reliable cross-platform encoding and decoding of images using variational models.

286, TITLE: Value Propagation Networks

<https://openreview.net/forum?id=SJG6G2RqtX>

AUTHORS: Nantas Nardelli, Gabriel Synnaeve, Zeming Lin, Pushmeet Kohli, Philip H. S. Torr, Nicolas Usunier

TL;DR: We present planners based on convnets that are sample-efficient and that generalize to larger instances of navigation and pathfinding problems.

HIGHLIGHT: We present Value Propagation (VProp), a set of parameter-efficient differentiable planning modules built on Value Iteration which can successfully be trained using reinforcement learning to solve unseen tasks, has the capability to generalize to larger map sizes, and can learn to navigate in dynamic environments.

287, TITLE: Bayesian Policy Optimization for Model Uncertainty

<https://openreview.net/forum?id=SJGvns0qK7>

AUTHORS: Gilwoo Lee, Brian Hou, Aditya Mandalika, Jeongseok Lee, Sanjiban Choudhury, Siddhartha S. Srinivasa

TL;DR: We formulate model uncertainty in Reinforcement Learning as a continuous Bayes-Adaptive Markov Decision Process and present a method for practical and scalable Bayesian policy optimization.

HIGHLIGHT: To address challenges from discretizing the continuous latent parameter space, we propose a new policy network architecture that encodes the belief distribution independently from the observable state.

288, TITLE: Variational Bayesian Phylogenetic Inference

<https://openreview.net/forum?id=SJVmjjR9FX>

AUTHORS: Cheng Zhang, Frederick A. Matsen IV

TL;DR: The first variational Bayes formulation of phylogenetic inference, a challenging inference problem over structures with intertwined discrete and continuous components

HIGHLIGHT: In this paper we present an alternative approach: a variational framework for Bayesian phylogenetic analysis.

289, TITLE: LEARNING FACTORIZED REPRESENTATIONS FOR OPEN-SET DOMAIN ADAPTATION

<https://openreview.net/forum?id=SJe3HiC5KX>

AUTHORS: Mahsa Baktashmotlagh, Masoud Faraki, Tom Drummond, Mathieu Salzmann

TL;DR: The first variational Bayes formulation of phylogenetic inference, a challenging inference problem over structures with intertwined discrete and continuous components

**HIGHLIGHT:** In this paper, we tackle the more challenging, yet more realistic case of open-set domain adaptation, where new, unknown classes can be present in the target data.

290, **TITLE:** On the Universal Approximability and Complexity Bounds of Quantized ReLU Neural Networks

<https://openreview.net/forum?id=SJe9rh0cFX>

**AUTHORS:** Yukun Ding, Jinglan Liu, Jinjun Xiong, Yiyu Shi

**TL;DR:** This paper proves the universal approximability of quantized ReLU neural networks and puts forward the complexity bound given arbitrary error.

**HIGHLIGHT:** In this paper, we study the representation power of quantized neural networks.

291, **TITLE:** Learning Localized Generative Models for 3D Point Clouds via Graph Convolution

<https://openreview.net/forum?id=SJeXSo09FQ>

**AUTHORS:** Diego Valsesia, Giulia Fracastoro, Enrico Magli

**TL;DR:** A GAN using graph convolution operations with dynamically computed graphs from hidden features

**HIGHLIGHT:** We focus on the generator of a GAN and define methods for graph convolution when the graph is not known in advance as it is the very output of the generator.

292, **TITLE:** ACCELERATING NONCONVEX LEARNING VIA REPLICA EXCHANGE LANGEVIN DIFFUSION

<https://openreview.net/forum?id=SJfPFjA9Fm>

**AUTHORS:** Yi Chen, Jinglin Chen, Jing Dong, Jian Peng, Zhaoran Wang

**TL;DR:** A GAN using graph convolution operations with dynamically computed graphs from hidden features

**HIGHLIGHT:** To attain the advantages of both regimes, we propose to use replica exchange, which swaps between two Langevin diffusions with different temperatures.

293, **TITLE:** Dynamically Unfolding Recurrent Restorer: A Moving Endpoint Control Method for Image Restoration

<https://openreview.net/forum?id=SJfZKiC5FX>

**AUTHORS:** Xiaoshuai Zhang, Yiping Lu, Jiaying Liu, Bin Dong

**TL;DR:** We propose a novel method to handle image degradations of different levels by learning a diffusion terminal time. Our model can generalize to unseen degradation level and different noise statistic.

**HIGHLIGHT:** In this paper, we propose a new control framework called the moving endpoint control to restore images corrupted by different degradation levels in one model.

294, **TITLE:** Bias-Reduced Uncertainty Estimation for Deep Neural Classifiers

<https://openreview.net/forum?id=SJfb5jCqKm>

**AUTHORS:** Yonatan Geifman, Guy Uziel, Ran El-Yaniv

**TL;DR:** We use snapshots from the training process to improve any uncertainty estimation method of a DNN classifier.

**HIGHLIGHT:** We consider the problem of uncertainty estimation in the context of (non-Bayesian) deep neural classification.

295, **TITLE:** CAMOU: Learning Physical Vehicle Camouflages to Adversarially Attack Detectors in the Wild

<https://openreview.net/forum?id=SJgEl3A5tm>

**AUTHORS:** Yang Zhang, Hassan Foroosh, Philip David, Boqing Gong

TL;DR: We propose a method to learn physical vehicle camouflage to adversarially attack object detectors in the wild. We find our camouflage effective and transferable.

HIGHLIGHT: In this paper, we conduct an intriguing experimental study about the physical adversarial attack on object detectors in the wild.

296, TITLE: Learning Latent Superstructures in Variational Autoencoders for Deep Multidimensional Clustering

<https://openreview.net/forum?id=SJgNwi09Km>

AUTHORS: Xiaopeng Li, Zhouong Chen, Leonard K. M. Poon, Nevin L. Zhang

TL;DR: We investigate a variant of variational autoencoders where there is a superstructure of discrete latent variables on top of the latent features.

HIGHLIGHT: We investigate a variant of variational autoencoders where there is a superstructure of discrete latent variables on top of the latent features.

297, TITLE: Learning Programmatically Structured Representations with Perceptor Gradients

<https://openreview.net/forum?id=SJggZnRcFQ>

AUTHORS: Svetlin Penkov, Subramanian Ramamoorthy

TL;DR: We investigate a variant of variational autoencoders where there is a superstructure of discrete latent variables on top of the latent features.

HIGHLIGHT: We present the perceptor gradients algorithm -- a novel approach to learning symbolic representations based on the idea of decomposing an agent's policy into i) a perceptor network extracting symbols from raw observation data and ii) a task encoding program which maps the input symbols to output actions.

298, TITLE: Variational Autoencoders with Jointly Optimized Latent Dependency Structure

<https://openreview.net/forum?id=SJgsCjCqt7>

AUTHORS: Jiawei He, Yu Gong, Joseph Marino, Greg Mori, Andreas Lehmann

TL;DR: We propose a method for learning latent dependency structure in variational autoencoders.

HIGHLIGHT: We propose a method for learning the dependency structure between latent variables in deep latent variable models.

299, TITLE: The Unusual Effectiveness of Averaging in GAN Training

[https://openreview.net/forum?id=SJgw\\_sRqFQ](https://openreview.net/forum?id=SJgw_sRqFQ)

AUTHORS: Yasin Yazıcı, Chuan-Sheng Foo, Stefan Winkler, Kim-Hui Yap, Georgios Piliouras, Vijay Chandrasekhar

TL;DR: We propose a method for learning latent dependency structure in variational autoencoders.

HIGHLIGHT: We examine two different techniques for parameter averaging in GAN training.

300, TITLE: Beyond Pixel Norm-Balls: Parametric Adversaries using an Analytically Differentiable Renderer

<https://openreview.net/forum?id=SJI2niR9KQ>

AUTHORS: Hsueh-Ti Derek Liu, Michael Tao, Chun-Liang Li, Derek Nowrouzezahrai, Alec Jacobson

TL;DR: Enabled by a novel differentiable renderer, we propose a new metric that has real-world implications for evaluating adversarial machine learning algorithms, resolving the lack of realism of the existing metric based on pixel norms.

HIGHLIGHT: As such, we propose a novel evaluation measure, parametric norm-balls, by directly perturbing physical parameters that underly image formation.

301, TITLE: Diversity is All You Need: Learning Skills without a Reward Function

<https://openreview.net/forum?id=SJx63jRqFm>

AUTHORS: Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, Sergey Levine

TL;DR: We propose an algorithm for learning useful skills without a reward function, and show how these skills can be used to solve downstream tasks.

HIGHLIGHT: Intelligent creatures can explore their environments and learn useful skills without supervision. In this paper, we propose "Diversity is All You Need" (DIAYN), a method for learning useful skills without a reward function.

302, TITLE: Supervised Policy Update for Deep Reinforcement Learning

<https://openreview.net/forum?id=SJxTroR9F7>

AUTHORS: Quan Vuong, Yiming Zhang, Keith W. Ross

TL;DR: first posing and solving the sample efficiency optimization problem in the non-parameterized policy space, and then solving a supervised regression problem to find a parameterized policy that is near the optimal non-parameterized policy.

HIGHLIGHT: We propose a new sample-efficient methodology, called Supervised Policy Update (SPU), for deep reinforcement learning.

303, TITLE: Learning sparse relational transition models

<https://openreview.net/forum?id=SJxsV2R5FQ>

AUTHORS: Victoria Xia, Zi Wang, Kelsey Allen, Tom Silver, Leslie Pack Kaelbling

TL;DR: A new approach that learns a representation for describing transition models in complex uncertain domains using relational rules.

HIGHLIGHT: We present a representation for describing transition models in complex uncertain domains using relational rules.

304, TITLE: Learning to Schedule Communication in Multi-agent Reinforcement Learning

<https://openreview.net/forum?id=SJxu5iR9KQ>

AUTHORS: Daewoo Kim, Sangwoo Moon, David Hostallero, Wan Ju Kang, Taeyoung Lee, Kyunghwan Son, Yung Yi

TL;DR: A new approach that learns a representation for describing transition models in complex uncertain domains using relational rules.

HIGHLIGHT: In this paper, we study a practical scenario when (i) the communication bandwidth is limited and (ii) the agents share the communication medium so that only a restricted number of agents are able to simultaneously use the medium, as in the state-of-the-art wireless networking standards.

305, TITLE: Hierarchical RL Using an Ensemble of Proprioceptive Periodic Policies

<https://openreview.net/forum?id=SJz1x20cFQ>

AUTHORS: Kenneth Marino, Abhinav Gupta, Rob Fergus, Arthur Szlam

TL;DR: A new approach that learns a representation for describing transition models in complex uncertain domains using relational rules.

HIGHLIGHT: In this paper we introduce a simple, robust approach to hierarchically training an agent in the setting of sparse reward tasks. The agent is split into a low-level and a high-level policy.

306, TITLE: Multi-class classification without multi-class labels

<https://openreview.net/forum?id=SJzR2iRcK7>

AUTHORS: Yen-Chang Hsu, Zhaoyang Lv, Joel Schlosser, Phillip Odom, Zsolt Kira

TL;DR: A new approach that learns a representation for describing transition models in complex uncertain domains using relational rules.

HIGHLIGHT: We formulate this approach, present a probabilistic graphical model for it, and derive a surprisingly simple loss function that can be used to learn neural network-based models.

307, TITLE: What do you learn from context? Probing for sentence structure in contextualized word representations

<https://openreview.net/forum?id=SJzSgnRcKX>

AUTHORS: 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

TL;DR: We probe for sentence structure in ELMo and related contextual embedding models. We find existing models efficiently encode syntax and show evidence of long-range dependencies, but only offer small improvements on semantic tasks.

HIGHLIGHT: Building on recent token-level probing work, we introduce a novel edge probing task design and construct a broad suite of sub-sentence tasks derived from the traditional structured NLP pipeline.

308, TITLE: Spectral Inference Networks: Unifying Deep and Spectral Learning

<https://openreview.net/forum?id=SJzqpj09YQ>

AUTHORS: David Pfau, Stig Petersen, Ashish Agarwal, David G. T. Barrett, Kimberly L. Stachenfeld

TL;DR: We show how to learn spectral decompositions of linear operators with deep learning, and use it for unsupervised learning without a generative model.

HIGHLIGHT: We present Spectral Inference Networks, a framework for learning eigenfunctions of linear operators by stochastic optimization.

309, TITLE: PeerNets: Exploiting Peer Wisdom Against Adversarial Attacks

<https://openreview.net/forum?id=Sk4jFoA9K7>

AUTHORS: Jan Svoboda, Jonathan Masci, Federico Monti, Michael Bronstein, Leonidas Guibas

TL;DR: We show how to learn spectral decompositions of linear operators with deep learning, and use it for unsupervised learning without a generative model.

HIGHLIGHT: Unfortunately, it has been shown that such systems are vulnerable to adversarial attacks, making them prone to potential unlawful uses. Designing deep neural networks that are robust to adversarial attacks is a fundamental step in making such systems safer and deployable in a broader variety of applications (e.g. autonomous driving), but more importantly is a necessary step to design novel and more advanced architectures built on new computational paradigms rather than marginally building on the existing ones. In this paper we introduce PeerNets, a novel family of convolutional networks alternating classical Euclidean convolutions with graph convolutions to harness information from a graph of peer samples.

310, TITLE: Attentive Neural Processes

<https://openreview.net/forum?id=SkE6PjC9KX>

AUTHORS: Hyunjik Kim, Andriy Mnih, Jonathan Schwarz, Marta Garnelo, Ali Eslami, Dan Rosenbaum, Oriol Vinyals, Yee Whye Teh

TL;DR: A model for regression that learns conditional distributions of a stochastic process, by incorporating attention into Neural Processes.

HIGHLIGHT: We address this issue by incorporating attention into NPs, allowing each input location to attend to the relevant context points for the prediction.

311, TITLE: Representation Degeneration Problem in Training Natural Language Generation Models

<https://openreview.net/forum?id=SkEYojRqtm>

AUTHORS: Jun Gao, Di He, Xu Tan, Tao Qin, Liwei Wang, Tieyan Liu

TL;DR: A model for regression that learns conditional distributions of a stochastic process, by incorporating attention into Neural Processes.

HIGHLIGHT: We analyze the conditions and causes of this problem and propose a novel regularization method to address it.

312, TITLE: Hierarchical interpretations for neural network predictions  
<https://openreview.net/forum?id=SkEqro0ctQ>  
AUTHORS: Chandan Singh, W. James Murdoch, Bin Yu  
TL;DR: We introduce and validate hierarchical local interpretations, the first technique to automatically search for and display important interactions for individual predictions made by LSTMs and CNNs.  
HIGHLIGHT: To ameliorate this problem, we introduce the use of hierarchical interpretations to explain DNN predictions through our proposed method: agglomerative contextual decomposition (ACD).

313, TITLE: Spreading vectors for similarity search  
<https://openreview.net/forum?id=SkGuG2R5tm>  
AUTHORS: Alexandre Sablayrolles, Matthijs Douze, Cordelia Schmid, Hervé Jégou  
TL;DR: We learn a neural network that uniformizes the input distribution, which leads to competitive indexing performance in high-dimensional space  
HIGHLIGHT: In this work, we propose to reverse this paradigm and adapt the data to the quantizer: we train a neural net whose last layers form a fixed parameter-free quantizer, such as pre-defined points of a sphere.

314, TITLE: A Convergence Analysis of Gradient Descent for Deep Linear Neural Networks  
<https://openreview.net/forum?id=SkMQg3C5K7>  
AUTHORS: Sanjeev Arora, Nadav Cohen, Noah Golowich, Wei Hu  
TL;DR: We analyze gradient descent for deep linear neural networks, providing a guarantee of convergence to global optimum at a linear rate.  
HIGHLIGHT: We analyze speed of convergence to global optimum for gradient descent training a deep linear neural network by minimizing the L2 loss over whitened data.

315, TITLE: Feed-forward Propagation in Probabilistic Neural Networks with Categorical and Max Layers  
<https://openreview.net/forum?id=SkMuPjRcKQ>  
AUTHORS: Alexander Shekhovtsov, Boris Flach  
TL;DR: Approximating mean and variance of the NN output over noisy input / dropout / uncertain parameters. Analytic approximations for argmax, softmax and max layers.  
HIGHLIGHT: Probabilistic Neural Networks deal with various sources of stochasticity: input noise, dropout, stochastic neurons, parameter uncertainties modeled as random variables, etc. In this paper we revisit a feed-forward propagation approach that allows one to estimate for each neuron its mean and variance w.r.t. all mentioned sources of stochasticity.

316, TITLE: Measuring and regularizing networks in function space  
<https://openreview.net/forum?id=SkMwpiR9Y7>  
AUTHORS: Ari Benjamin, David Rolnick, Konrad Kording  
TL;DR: We find movement in function space is not proportional to movement in parameter space during optimization. We propose a new natural-gradient style optimizer to address this.  
HIGHLIGHT: Here, we show that it is simple and computationally feasible to calculate distances between functions in a  $L^2$  Hilbert space.

317, TITLE: Fluctuation-dissipation relations for stochastic gradient descent  
<https://openreview.net/forum?id=SkNksoRctQ>  
AUTHORS: Sho Yaida  
TL;DR: We prove fluctuation-dissipation relations for SGD, which can be used to (i) adaptively set learning rates and (ii) probe loss surfaces.

**HIGHLIGHT:** Here, we derive stationary fluctuation-dissipation relations that link measurable quantities and hyperparameters in the stochastic gradient descent algorithm.

318, **TITLE:** Poincare Glove: Hyperbolic Word Embeddings

<https://openreview.net/forum?id=Ske5r3AqK7>

**AUTHORS:** Alexandru Tifrea\*, Gary Becigneul\*, Octavian-Eugen Ganea\*

**TL;DR:** We embed words in the hyperbolic space and make the connection with the Gaussian word embeddings.

**HIGHLIGHT:** In this paper, justified by the notion of delta-hyperbolicity or tree-likeness of a space, we propose to embed words in a Cartesian product of hyperbolic spaces which we theoretically connect to the Gaussian word embeddings and their Fisher geometry.

319, **TITLE:** Episodic Curiosity through Reachability

<https://openreview.net/forum?id=SkeK3s0qKQ>

**AUTHORS:** Nikolay Savinov, Anton Raichuk, Damien Vincent, Raphael Marinier, Marc Pollefeys, Timothy Lillicrap, Sylvain Gelly

**TL;DR:** We propose a novel model of curiosity based on episodic memory and the ideas of reachability which allows us to overcome the known "couch-potato" issues of prior work.

**HIGHLIGHT:** We propose a new curiosity method which uses episodic memory to form the novelty bonus.

320, **TITLE:** Phase-Aware Speech Enhancement with Deep Complex U-Net

<https://openreview.net/forum?id=SkeRTsAcYm>

**AUTHORS:** Hyeong-Seok Choi, Jang-Hyun Kim, Jaesung Huh, Adrian Kim, Jung-Woo Ha, Kyogu Lee

**TL;DR:** This paper proposes a novel complex masking method for speech enhancement along with a loss function for efficient phase estimation.

**HIGHLIGHT:** To improve speech enhancement performance, we tackle the phase estimation problem in three ways.

321, **TITLE:** Generative predecessor models for sample-efficient imitation learning

<https://openreview.net/forum?id=SkeVsiAcYm>

**AUTHORS:** Yannick Schroecker, Mel Vecerik, Jon Scholz

**TL;DR:** This paper proposes a novel complex masking method for speech enhancement along with a loss function for efficient phase estimation.

**HIGHLIGHT:** We propose Generative Predecessor Models for Imitation Learning (GPRIL), a novel imitation learning algorithm that matches the state-action distribution to the distribution observed in expert demonstrations, using generative models to reason probabilistically about alternative histories of demonstrated states.

322, **TITLE:** Adaptive Estimators Show Information Compression in Deep Neural Networks

<https://openreview.net/forum?id=SkeZisA5t7>

**AUTHORS:** Ivan Chelombiev, Conor Houghton, Cian O'Donnell

**TL;DR:** We developed robust mutual information estimates for DNNs and used them to observe compression in networks with non-saturating activation functions

**HIGHLIGHT:** In this paper we developed more robust mutual information estimation techniques, that adapt to hidden activity of neural networks and produce more sensitive measurements of activations from all functions, especially unbounded functions.

323, **TITLE:** Multilingual Neural Machine Translation With Soft Decoupled Encoding

<https://openreview.net/forum?id=Skeke3C5Fm>

AUTHORS: Xinyi Wang, Hieu Pham, Philip Arthur, Graham Neubig

TL;DR: We developed robust mutual information estimates for DNNs and used them to observe compression in networks with non-saturating activation functions

HIGHLIGHT: In this paper, we propose Soft Decoupled Encoding (SDE), a multilingual lexicon encoding framework specifically designed to share lexical-level information intelligently without requiring heuristic preprocessing such as pre-segmenting the data.

324, TITLE: Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet

<https://openreview.net/forum?id=SkfMWhAqYQ>

AUTHORS: Wieland Brendel, Matthias Bethge

TL;DR: Aggregating class evidence from many small image patches suffices to solve ImageNet, yields more interpretable models and can explain aspects of the decision-making of popular DNNs.

HIGHLIGHT: We here introduce a high-performance DNN architecture on ImageNet whose decisions are considerably easier to explain.

325, TITLE: Reward Constrained Policy Optimization

<https://openreview.net/forum?id=SkfrvsA9FX>

AUTHORS: Chen Tessler, Daniel J. Mankowitz, Shie Mannor

TL;DR: For complex constraints in which it is not easy to estimate the gradient, we use the discounted penalty as a guiding signal. We prove that under certain assumptions it converges to a feasible solution.

HIGHLIGHT: In this work we present a novel multi-timescale approach for constrained policy optimization, called 'Reward Constrained Policy Optimization' (RCPO), which uses an alternative penalty signal to guide the policy towards a constraint satisfying one.

326, TITLE: On the Relation Between the Sharpest Directions of DNN Loss and the SGD Step Length

<https://openreview.net/forum?id=SkfEaj05t7>

AUTHORS: Stanislaw Jastrzebski, Zachary Kenton, Nicolas Ballas, Asja Fischer, Yoshua Bengio, Amos Storkey

TL;DR: SGD is steered early on in training towards a region in which its step is too large compared to curvature, which impacts the rest of training.

HIGHLIGHT: In this paper we extend previous work by investigating the curvature of the loss surface along the whole training trajectory, rather than only at the endpoint.

327, TITLE: Modeling the Long Term Future in Model-Based Reinforcement Learning

<https://openreview.net/forum?id=SkfQBn0cF7>

AUTHORS: Nan Rosemary Ke, Amanpreet Singh, Ahmed Touati, Anirudh Goyal, Yoshua Bengio, Devi Parikh, Dhruv Batra

TL;DR: incorporating, in the model, latent variables that encode future content improves the long-term prediction accuracy, which is critical for better planning in model-based RL.

HIGHLIGHT: To this end, we build a latent-variable autoregressive model by leveraging recent ideas in variational inference.

328, TITLE: Understanding Straight-Through Estimator in Training Activation Quantized Neural Nets

<https://openreview.net/forum?id=Skh4jRcKQ>

AUTHORS: Penghang Yin, Jiancheng Lyu, Shuai Zhang, Stanley Osher, Yingyong Qi, Jack Xin

TL;DR: We make theoretical justification for the concept of straight-through estimator.

HIGHLIGHT: In this paper, we provide the theoretical justification of the concept of STE by answering this question.

329, TITLE: DISTRIBUTIONAL CONCAVITY REGULARIZATION FOR GANS  
<https://openreview.net/forum?id=SkIEEnC5tQ>  
AUTHORS: Shoichiro Yamaguchi, Masanori Koyama  
TL;DR: We make theoretical justification for the concept of straight-through estimator.  
HIGHLIGHT: We propose Distributional Concavity (DC) regularization for Generative Adversarial Networks (GANs), a functional gradient-based method that promotes the entropy of the generator distribution and works against mode collapse. Our DC regularization is an easy-to-implement method that can be used in combination with the current state of the art methods like Spectral Normalization and Wasserstein GAN with gradient penalty to further improve the performance. We will not only show that our DC regularization can achieve highly competitive results on ILSVRC2012 and CIFAR datasets in terms of Inception score and Fréchet inception distance, but also provide a mathematical guarantee that our method can always increase the entropy of the generator distribution.

330, TITLE: LeMoNAde: Learned Motif and Neuronal Assembly Detection in calcium imaging videos  
<https://openreview.net/forum?id=SkloDjAqYm>  
AUTHORS: Elke Kirschbaum, Manuel Hausmann, Steffen Wolf, Hannah Sonntag, Justus Schneider, Shehabeldin Elzoheiry, Oliver Kann, Daniel Durstewitz, Fred A Hamprecht  
TL;DR: We present LeMoNAde, an end-to-end learned motif detection method directly operating on calcium imaging videos.  
HIGHLIGHT: We here propose LeMoNAde, a new exploratory data analysis method that facilitates hunting for motifs in calcium imaging videos, the dominant microscopic functional imaging modality in neurophysiology.

331, TITLE: Competitive experience replay  
<https://openreview.net/forum?id=Skism20ctX>  
AUTHORS: Hao Liu, Alexander Trott, Richard Socher, Caiming Xiong  
TL;DR: a novel method to learn with sparse reward using adversarial reward re-labeling  
HIGHLIGHT: We propose a novel method called competitive experience replay, which efficiently supplements a sparse reward by placing learning in the context of an exploration competition between a pair of agents.

332, TITLE: Multi-Domain Adversarial Learning  
<https://openreview.net/forum?id=SkIv5iRqYX>  
AUTHORS: Alice Schoenauer-Sebag, Louise Heinrich, Marc Schoenauer, Michele Sebag, Lani F. Wu, Steve J. Altschuler  
TL;DR: Adversarial Domain adaptation and Multi-domain learning: a new loss to handle multi- and single-domain classes in the semi-supervised setting.  
HIGHLIGHT: This paper presents a multi-domain adversarial learning approach, MuLANN, to leverage multiple datasets with overlapping but distinct class sets, in a semi-supervised setting.

333, TITLE: ProMP: Proximal Meta-Policy Search  
<https://openreview.net/forum?id=SkxXCi0qFX>  
AUTHORS: Jonas Rothfuss, Dennis Lee, Ignasi Clavera, Tamim Asfour, Pieter Abbeel  
TL;DR: A novel and theoretically grounded meta-reinforcement learning algorithm  
HIGHLIGHT: Building on the gained insights we develop a novel meta-learning algorithm that overcomes both the issue of poor credit assignment and previous difficulties in estimating meta-policy gradients.

334, TITLE: Don't Settle for Average, Go for the Max: Fuzzy Sets and Max-Pooled Word Vectors

<https://openreview.net/forum?id=SkxXg2C5FX>

AUTHORS: Vitalii Zhelezniak, Aleksandar Savkov, April Shen, Francesco Moramarco, Jack Flann, Nils Y. Hammerla

TL;DR: Max-pooled word vectors with fuzzy Jaccard set similarity are an extremely competitive baseline for semantic similarity; we propose a simple dynamic variant that performs even better.

HIGHLIGHT: We propose a novel fuzzy bag-of-words (FBoW) representation for text that contains all the words in the vocabulary simultaneously but with different degrees of membership, which are derived from similarities between word vectors.

335, TITLE: Stable Opponent Shaping in Differentiable Games

<https://openreview.net/forum?id=SyGjjsC5tQ>

AUTHORS: Alistair Letcher, Jakob Foerster, David Balduzzi, Tim Rocktäschel, Shimon Whiteson

TL;DR: Opponent shaping is a powerful approach to multi-agent learning but can prevent convergence; our SOS algorithm fixes this with strong guarantees in all differentiable games.

HIGHLIGHT: In this paper we present Stable Opponent Shaping (SOS), a new method that interpolates between LOLA and a stable variant named LookAhead.

336, TITLE: A Mean Field Theory of Batch Normalization

<https://openreview.net/forum?id=SyMDXnCcF7>

AUTHORS: Greg Yang, Jeffrey Pennington, Vinay Rao, Jascha Sohl-Dickstein, Samuel S. Schoenholz

TL;DR: Batch normalization causes exploding gradients in vanilla feedforward networks.

HIGHLIGHT: We develop a mean field theory for batch normalization in fully-connected feedforward neural networks.

337, TITLE: Learning Exploration Policies for Navigation

<https://openreview.net/forum?id=SyMWn05F7>

AUTHORS: Tao Chen, Saurabh Gupta, Abhinav Gupta

TL;DR: Batch normalization causes exploding gradients in vanilla feedforward networks.

HIGHLIGHT: In this work, we study how agents can autonomously explore realistic and complex 3D environments without the context of task-rewards.

338, TITLE: Distribution-Interpolation Trade off in Generative Models

<https://openreview.net/forum?id=SyMhLo0qKQ>

AUTHORS: Damian Lesniak, Igor Sieradzki, Igor Podolak

TL;DR: We theoretically prove that linear interpolations are unsuitable for analysis of trained implicit generative models.

HIGHLIGHT: We investigate the properties of multidimensional probability distributions in the context of latent space prior distributions of implicit generative models.

339, TITLE: Learning to Describe Scenes with Programs

<https://openreview.net/forum?id=SyNpk2R9K7>

AUTHORS: Yunchao Liu, Zheng Wu, Daniel Ritchie, William T. Freeman, Joshua B. Tenenbaum, Jiajun Wu

TL;DR: We present scene programs, a structured scene representation that captures both low-level object appearance and high-level regularity in the scene.

HIGHLIGHT: In this paper, we present scene programs, representing a scene via a symbolic program for its objects, attributes, and their relations.

340, TITLE: Visceral Machines: Risk-Aversion in Reinforcement Learning with Intrinsic Physiological Rewards  
<https://openreview.net/forum?id=SyNvti09KQ>  
AUTHORS: Daniel McDuff, Ashish Kapoor  
TL;DR: We present a novel approach to reinforcement learning that leverages a task-independent intrinsic reward function trained on peripheral pulse measurements that are correlated with human autonomic nervous system responses.  
HIGHLIGHT: We present a novel approach to reinforcement learning that leverages a task-independent intrinsic reward function trained on peripheral pulse measurements that are correlated with human autonomic nervous system responses.

341, TITLE: Deep Frank-Wolfe For Neural Network Optimization  
<https://openreview.net/forum?id=SyVU6s05K7>  
AUTHORS: Leonard Berrada, Andrew Zisserman, M. Pawan Kumar  
TL;DR: We train neural networks by locally linearizing them and using a linear SVM solver (Frank-Wolfe) at each iteration.  
HIGHLIGHT: We present an optimization method that offers empirically the best of both worlds: our algorithm yields good generalization performance while requiring only one hyper-parameter.

342, TITLE: LEARNING TO PROPAGATE LABELS: TRANSDUCTIVE PROPAGATION NETWORK FOR FEW-SHOT LEARNING  
<https://openreview.net/forum?id=SyVuRiC5K7>  
AUTHORS: Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, Eunho Yang, Sung Ju Hwang, Yi Yang  
TL;DR: We propose a novel meta-learning framework for transductive inference that classifies the entire test set at once to alleviate the low-data problem.  
HIGHLIGHT: In this paper, we propose Transductive Propagation Network (TPN), a novel meta-learning framework for transductive inference that classifies the entire test set at once to alleviate the low-data problem.

343, TITLE: Improving the Generalization of Adversarial Training with Domain Adaptation  
<https://openreview.net/forum?id=SyflfnC5Ym>  
AUTHORS: Chuanbiao Song, Kun He, Liwei Wang, John E. Hopcroft  
TL;DR: We propose a novel adversarial training with domain adaptation method that significantly improves the generalization ability on adversarial examples from different attacks.  
HIGHLIGHT: To alleviate this problem, we propose a novel Adversarial Training with Domain Adaptation (ATDA) method.

344, TITLE: Dimensionality Reduction for Representing the Knowledge of Probabilistic Models  
<https://openreview.net/forum?id=SygD-hCcF7>  
AUTHORS: Marc T Law, Jake Snell, Amir-massoud Farahmand, Raquel Urtasun, Richard S Zemel  
TL;DR: dimensionality reduction for cases where examples can be represented as soft probability distributions  
HIGHLIGHT: We propose a simple, intuitive and scalable dimension reduction framework that takes into account the soft probabilistic interpretation of standard deep models for classification.

345, TITLE: Learning protein sequence embeddings using information from structure  
<https://openreview.net/forum?id=SygLehCqtm>  
AUTHORS: Tristan Bepler, Bonnie Berger

TL;DR: We present a method for learning protein sequence embedding models using structural information in the form of global structural similarity between proteins and within protein residue-residue contacts.

HIGHLIGHT: We introduce a framework that maps any protein sequence to a sequence of vector embeddings --- one per amino acid position --- that encode structural information.

346, TITLE: Variational Smoothing in Recurrent Neural Network Language Models

<https://openreview.net/forum?id=SygQvs0cFQ>

AUTHORS: Lingpeng Kong, Gabor Melis, Wang Ling, Lei Yu, Dani Yogatama

TL;DR: We present a method for learning protein sequence embedding models using structural information in the form of global structural similarity between proteins and within protein residue-residue contacts.

HIGHLIGHT: We present a new theoretical perspective of data noising in recurrent neural network language models (Xie et al., 2017).

347, TITLE: Biologically-Plausible Learning Algorithms Can Scale to Large Datasets

<https://openreview.net/forum?id=SylvZ209F7>

AUTHORS: Will Xiao, Honglin Chen, Qianli Liao, Tomaso Poggio

TL;DR: Biologically plausible learning algorithms, particularly sign-symmetry, work well on ImageNet

HIGHLIGHT: To address this ?weight transport problem? (Grossberg, 1987), two biologically-plausible algorithms, proposed by Liao et al. (2016) and Lillicrap et al. (2016), relax BP?s weight symmetry requirements and demonstrate comparable learning capabilities to that of BP on small datasets.

348, TITLE: Coarse-grain Fine-grain Coattention Network for Multi-evidence Question Answering

<https://openreview.net/forum?id=Syl7OsRqY7>

AUTHORS: Victor Zhong, Caiming Xiong, Nitish Shirish Keskar, Richard Socher

TL;DR: A new state-of-the-art model for multi-evidence question answering using coarse-grain fine-grain hierarchical attention.

HIGHLIGHT: In this work, we propose the Coarse-grain Fine-grain Coattention Network (CFC), a new question answering model that combines information from evidence across multiple documents.

349, TITLE: Learning a Meta-Solver for Syntax-Guided Program Synthesis

<https://openreview.net/forum?id=Syl8Sn0cK7>

AUTHORS: Xujie Si, Yuan Yang, Hanjun Dai, Mayur Naik, Le Song

TL;DR: We propose a meta-learning framework that learns a transferable policy from only weak supervision to solve synthesis tasks with different logical specifications and grammars.

HIGHLIGHT: To address these challenges, we propose a meta-learning framework that learns a transferable policy from only weak supervision.

350, TITLE: Towards Robust, Locally Linear Deep Networks

<https://openreview.net/forum?id=SylCrmCcFX>

AUTHORS: Guang-He Lee, David Alvarez-Melis, Tommi S. Jaakkola

TL;DR: A scalable algorithm to establish robust derivatives of deep networks w.r.t. the inputs.

HIGHLIGHT: In this paper, we propose a new learning problem to encourage deep networks to have stable derivatives over larger regions.

351, TITLE: How Important is a Neuron

<https://openreview.net/forum?id=SylKoo0cKm>

AUTHORS: Kedar Dhamdhere, Mukund Sundararajan, Qiqi Yan

TL;DR: A scalable algorithm to establish robust derivatives of deep networks w.r.t. the inputs.

**HIGHLIGHT:** We introduce the notion of conductance to extend the notion of attribution to understanding the importance of hidden units. Informally, the conductance of a hidden unit of a deep network is the flow of attribution via this hidden unit.

352, **TITLE:** Learning to Make Analogies by Contrasting Abstract Relational Structure  
<https://openreview.net/forum?id=SylLYsCcFm>  
**AUTHORS:** Felix Hill, Adam Santoro, David Barrett, Ari Morcos, Timothy Lillicrap  
**TL;DR:** The most robust capacity for analogical reasoning is induced when networks learn analogies by contrasting abstract relational structures in their input domains.  
**HIGHLIGHT:** Here, we study how analogical reasoning can be induced in neural networks that learn to perceive and reason about raw visual data.

353, **TITLE:** Learning what you can do before doing anything  
<https://openreview.net/forum?id=SylPMnR9Ym>  
**AUTHORS:** Oleh Rybkin, Karl Pertsch, Konstantinos G. Derpanis, Kostas Daniilidis, Andrew Jaegle  
**TL;DR:** We learn a representation of an agent's action space from pure visual observations. We use a recurrent latent variable approach with a novel composability loss.  
**HIGHLIGHT:** In this work, we address the problem of learning an agent's action space purely from visual observation.

354, **TITLE:** Learning Grid Cells as Vector Representation of Self-Position Coupled with Matrix Representation of Self-Motion  
<https://openreview.net/forum?id=Syx0Mh05YQ>  
**AUTHORS:** Ruiqi Gao, Jianwen Xie, Song-Chun Zhu, Ying Nian Wu  
**TL;DR:** We learn a representation of an agent's action space from pure visual observations. We use a recurrent latent variable approach with a novel composability loss.  
**HIGHLIGHT:** This paper proposes a representational model for grid cells.

355, **TITLE:** Universal Stagewise Learning for Non-Convex Problems with Convergence on Averaged Solutions  
<https://openreview.net/forum?id=Syx5V2CcFm>  
**AUTHORS:** Zaiyi Chen, Zhuoning Yuan, Jinfeng Yi, Bowen Zhou, Enhong Chen, Tianbao Yang  
**TL;DR:** We learn a representation of an agent's action space from pure visual observations. We use a recurrent latent variable approach with a novel composability loss.  
**HIGHLIGHT:** We propose a universal stagewise optimization framework for a broad family of non-smooth non-convex problems with the following key features: (i) at each stage any suitable stochastic convex optimization algorithms (e.g., SGD or AdaGrad) that return an averaged solution can be employed for minimizing a regularized convex problem; (ii) the step size is decreased in a stagewise manner; (iii) an averaged solution is returned as the final solution.

356, **TITLE:** Invariant and Equivariant Graph Networks  
<https://openreview.net/forum?id=Syx72jC9tm>  
**AUTHORS:** Haggai Maron, Heli Ben-Hamu, Nadav Shamir, Yaron Lipman  
**TL;DR:** The paper provides a full characterization of permutation invariant and equivariant linear layers for graph data.  
**HIGHLIGHT:** A basic challenge in developing such networks is finding the maximal collection of invariant and equivariant `\emph{linear}` layers.

357, **TITLE:** Robustness May Be at Odds with Accuracy  
<https://openreview.net/forum?id=SyxAb30cY7>

AUTHORS: Dimitris Tsipras, Shibani Santurkar, Logan Engstrom, Alexander Turner, Aleksander Madry

TL;DR: We show that adversarial robustness might come at the cost of standard classification performance, but also yields unexpected benefits.

HIGHLIGHT: We show that there exists an inherent tension between the goal of adversarial robustness and that of standard generalization. Specifically, training robust models may not only be more resource-consuming, but also lead to a reduction of standard accuracy.

358, TITLE: Feature Intertwiner for Object Detection

<https://openreview.net/forum?id=SyxZJn05YX>

AUTHORS: Hongyang Li, Bo Dai, Shaoshuai Shi, Wanli Ouyang, Xiaogang Wang

TL;DR: (Camera-ready version) A feature intertwiner module to leverage features from one accurate set to help the learning of another less reliable set.

HIGHLIGHT: In this paper, we address this problem via a new perspective.

359, TITLE: Adversarial Reprogramming of Neural Networks

[https://openreview.net/forum?id=Syx\\_Ss05tm](https://openreview.net/forum?id=Syx_Ss05tm)

AUTHORS: Gamaleldin F. Elsayed, Ian Goodfellow, Jascha Sohl-Dickstein

TL;DR: We introduce the first instance of adversarial attacks that reprogram the target model to perform a task chosen by the attacker---without the attacker needing to specify or compute the desired output for each test-time input.

HIGHLIGHT: We introduce attacks that instead reprogram the target model to perform a task chosen by the attacker without the attacker needing to specify or compute the desired output for each test-time input.

360, TITLE: G-SGD: Optimizing ReLU Neural Networks in its Positively Scale-Invariant Space

<https://openreview.net/forum?id=SyxfEn09Y7>

AUTHORS: Qi Meng, Shuxin Zheng, Huishuai Zhang, Wei Chen, Qiwei Ye, Zhi-Ming Ma, Nenghai Yu, Tie-Yan Liu

TL;DR: We introduce the first instance of adversarial attacks that reprogram the target model to perform a task chosen by the attacker---without the attacker needing to specify or compute the desired output for each test-time input.

HIGHLIGHT: In this paper, we provide our positive answer to this question.

361, TITLE: From Hard to Soft: Understanding Deep Network Nonlinearities via Vector Quantization and Statistical Inference

<https://openreview.net/forum?id=Syxt2jC5FX>

AUTHORS: Randall Balestriero, Richard Baraniuk

TL;DR: Reformulate deep networks nonlinearities from a vector quantization scope and bridge most known nonlinearities together.

HIGHLIGHT: This paper extends the MASO framework to these and an infinitely large class of new nonlinearities by linking deterministic MASOs with probabilistic Gaussian Mixture Models (GMMs).

362, TITLE: Aggregated Momentum: Stability Through Passive Damping

<https://openreview.net/forum?id=Syxt5oC5YQ>

AUTHORS: James Lucas, Shengyang Sun, Richard Zemel, Roger Grosse

TL;DR: We introduce a simple variant of momentum optimization which is able to outperform classical momentum, Nesterov, and Adam on deep learning tasks with minimal hyperparameter tuning.

HIGHLIGHT: We propose Aggregated Momentum (AggMo), a variant of momentum which combines multiple velocity vectors with different damping coefficients.

363, TITLE: Variational Autoencoder with Arbitrary Conditioning  
<https://openreview.net/forum?id=SyxtJh0qYm>  
AUTHORS: Oleg Ivanov, Michael Figurnov, Dmitry Vetrov  
TL;DR: We propose an extension of conditional variational autoencoder that allows conditioning on an arbitrary subset of the features and sampling the remaining ones.  
HIGHLIGHT: We propose a single neural probabilistic model based on variational autoencoder that can be conditioned on an arbitrary subset of observed features and then sample the remaining features in "one shot".

364, TITLE: Time-Agnostic Prediction: Predicting Predictable Video Frames  
<https://openreview.net/forum?id=SyzVb3CcFX>  
AUTHORS: Dinesh Jayaraman, Frederik Ebert, Alexei Efros, Sergey Levine  
TL;DR: In visual prediction tasks, letting your predictive model choose which times to predict does two things: (i) improves prediction quality, and (ii) leads to semantically coherent "bottleneck state" predictions, which are useful for planning.  
HIGHLIGHT: We evaluate our approach for future and intermediate frame prediction across three robotic manipulation tasks.

365, TITLE: A Closer Look at Deep Learning Heuristics: Learning rate restarts, Warmup and Distillation  
<https://openreview.net/forum?id=r14EOsCqKX>  
AUTHORS: Akhilesh Gotmare, Nitish Shirish Keskar, Caiming Xiong, Richard Socher  
TL;DR: We use empirical tools of mode connectivity and SVCCA to investigate neural network training heuristics of learning rate restarts, warmup and knowledge distillation.  
HIGHLIGHT: Instead, we revisit the empirical analysis of heuristics through the lens of recently proposed methods for loss surface and representation analysis, viz. mode connectivity and canonical correlation analysis (CCA), and hypothesize reasons why the heuristics succeed.

366, TITLE: Self-Monitoring Navigation Agent via Auxiliary Progress Estimation  
<https://openreview.net/forum?id=r1GAsjC5Fm>  
AUTHORS: Chih-Yao Ma, Jiasen Lu, Zuxuan Wu, Ghassan AlRegib, Zsolt Kira, Richard Socher, Caiming Xiong  
TL;DR: We propose a self-monitoring agent for the Vision-and-Language Navigation task.  
HIGHLIGHT: In this paper, we introduce a self-monitoring agent with two complementary components: (1) visual-textual co-grounding module to locate the instruction completed in the past, the instruction required for the next action, and the next moving direction from surrounding images and (2) progress monitor to ensure the grounded instruction correctly reflects the navigation progress.

367, TITLE: Kernel Change-point Detection with Auxiliary Deep Generative Models  
<https://openreview.net/forum?id=r1GbfhRqF7>  
AUTHORS: Wei-Cheng Chang, Chun-Liang Li, Yiming Yang, Barnab's P?czos  
TL;DR: In this paper, we propose KL-CPD, a novel kernel learning framework for time series CPD that optimizes a lower bound of test power via an auxiliary generative model as a surrogate to the abnormal distribution.  
HIGHLIGHT: In this paper, we propose KL-CPD, a novel kernel learning framework for time series CPD that optimizes a lower bound of test power via an auxiliary generative model.

368, TITLE: Unsupervised Learning via Meta-Learning  
<https://openreview.net/forum?id=r1My6sR9tX>  
AUTHORS: Kyle Hsu, Sergey Levine, Chelsea Finn

TL;DR: An unsupervised learning method that uses meta-learning to enable efficient learning of downstream image classification tasks, outperforming state-of-the-art methods.  
HIGHLIGHT: Many prior unsupervised learning works aim to do so by developing proxy objectives based on reconstruction, disentanglement, prediction, and other metrics.

369, TITLE: Auxiliary Variational MCMC  
<https://openreview.net/forum?id=r1NJqsRctX>  
AUTHORS: Raza Habib, David Barber  
TL;DR: An unsupervised learning method that uses meta-learning to enable efficient learning of downstream image classification tasks, outperforming state-of-the-art methods.  
HIGHLIGHT: We introduce Auxiliary Variational MCMC, a novel framework for learning MCMC kernels that combines recent advances in variational inference with insights drawn from traditional auxiliary variable MCMC methods such as Hamiltonian Monte Carlo.

370, TITLE: Neural network gradient-based learning of black-box function interfaces  
<https://openreview.net/forum?id=r1e13s05YX>  
AUTHORS: Alon Jacovi, Guy Hadash, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, Jonathan Berant  
TL;DR: Training DNNs to interface w\ black box functions w\o intermediate labels by using an estimator sub-network that can be replaced with the black box after training  
HIGHLIGHT: We propose a method for end-to-end training of a base neural network that integrates calls to existing black-box functions.

371, TITLE: Self-Tuning Networks: Bilevel Optimization of Hyperparameters using Structured Best-Response Functions  
<https://openreview.net/forum?id=r1eEG20qKQ>  
AUTHORS: Matthew Mackay, Paul Vicol, Jonathan Lorraine, David Duvenaud, Roger Grosse  
TL;DR: We use a hypernetwork to predict optimal weights given hyperparameters, and jointly train everything together.  
HIGHLIGHT: We aim to adapt regularization hyperparameters for neural networks by fitting compact approximations to the best-response function, which maps hyperparameters to optimal weights and biases.

372, TITLE: Unsupervised Control Through Non-Parametric Discriminative Rewards  
<https://openreview.net/forum?id=r1eVMnA9K7>  
AUTHORS: David Warde-Farley, Tom Van de Wiele, Tejas Kulkarni, Catalin Ionescu, Steven Hansen, Volodymyr Mnih  
TL;DR: Unsupervised reinforcement learning method for learning a policy to robustly achieve perceptually specified goals.  
HIGHLIGHT: We present an unsupervised learning algorithm to train agents to achieve perceptually-specified goals using only a stream of observations and actions.

373, TITLE: Interpolation-Prediction Networks for Irregularly Sampled Time Series  
<https://openreview.net/forum?id=r1efr3C9Ym>  
AUTHORS: Satya Narayan Shukla, Benjamin Marlin  
TL;DR: This paper presents a new deep learning architecture for addressing the problem of supervised learning with sparse and irregularly sampled multivariate time series.  
HIGHLIGHT: In this paper, we present a new deep learning architecture for addressing the problem of supervised learning with sparse and irregularly sampled multivariate time series.

374, TITLE: Riemannian Adaptive Optimization Methods

<https://openreview.net/forum?id=r1eiqi09K7>

AUTHORS: Gary Becigneul, Octavian-Eugen Ganea

TL;DR: Adapting Adam, Amsgrad, Adagrad to Riemannian manifolds.

HIGHLIGHT: Several first order stochastic optimization methods commonly used in the Euclidean domain such as stochastic gradient descent (SGD), accelerated gradient descent or variance reduced methods have already been adapted to certain Riemannian settings.

375, TITLE: Minimal Random Code Learning: Getting Bits Back from Compressed Model Parameters

<https://openreview.net/forum?id=r1f0YiCctm>

AUTHORS: Marton Havasi, Robert Peharz, Jos? Miguel Hern?andez-Lobato

TL;DR: This paper proposes an effective method to compress neural networks based on recent results in information theory.

HIGHLIGHT: A typical approach is to train a set of deterministic weights, while applying certain techniques such as pruning and quantization, in order that the empirical weight distribution becomes amenable to Shannon-style coding schemes.

376, TITLE: Characterizing Audio Adversarial Examples Using Temporal Dependency

<https://openreview.net/forum?id=r1g4E3C9t7>

AUTHORS: Zhuolin Yang, Bo Li, Pin-Yu Chen, Dawn Song

TL;DR: Adversarial audio discrimination using temporal dependency

HIGHLIGHT: Nonetheless, as unique data properties have inspired distinct and powerful learning principles, this paper aims to explore their potentials towards mitigating adversarial inputs.

377, TITLE: Equi-normalization of Neural Networks

<https://openreview.net/forum?id=r1gEqiC9FX>

AUTHORS: Pierre Stock, Benjamin Graham, R?mi Gribonval, Herv? J?gou

TL;DR: Fast iterative algorithm to balance the energy of a network while staying in the same functional equivalence class

HIGHLIGHT: Inspired by the Sinkhorn-Knopp algorithm, we introduce a fast iterative method for minimizing the  $l_2$  norm of the weights, equivalently the weight decay regularizer.

378, TITLE: Generalized Tensor Models for Recurrent Neural Networks

<https://openreview.net/forum?id=r1gNni0qtm>

AUTHORS: Valentin Khruikov, Oleksii Hrinchuk, Ivan Oseledets

TL;DR: Analysis of expressivity and generality of recurrent neural networks with ReLU nonlinearities using Tensor-Train decomposition.

HIGHLIGHT: In this work, we attempt to reduce the gap between theory and practice by extending the theoretical analysis to RNNs which employ various nonlinearities, such as Rectified Linear Unit (ReLU), and show that they also benefit from properties of universality and depth efficiency.

379, TITLE: Wizard of Wikipedia: Knowledge-Powered Conversational Agents

<https://openreview.net/forum?id=r1173iRqKm>

AUTHORS: Emily Dinan, Stephen Roller, Kurt Shuster, Angela Fan, Michael Auli, Jason Weston

TL;DR: We build knowledgeable conversational agents by conditioning on Wikipedia + a new supervised task.

HIGHLIGHT: To that end we collect and release a large dataset with conversations directly grounded with knowledge retrieved from Wikipedia.

380, TITLE: Are adversarial examples inevitable?

<https://openreview.net/forum?id=r1IWUoA9FQ>

**AUTHORS:** Ali Shafahi, W. Ronny Huang, Christoph Studer, Soheil Feizi, Tom Goldstein  
**TL;DR:** This paper identifies classes of problems for which adversarial examples are inescapable, and derives fundamental bounds on the susceptibility of any classifier to adversarial examples.  
**HIGHLIGHT:** Given the lack of success at generating robust defenses, we are led to ask a fundamental question: Are adversarial attacks inevitable? This paper analyzes adversarial examples from a theoretical perspective, and identifies fundamental bounds on the susceptibility of a classifier to adversarial attacks.

381, **TITLE:** A Variational Inequality Perspective on Generative Adversarial Networks  
<https://openreview.net/forum?id=r1laEnA5Ym>  
**AUTHORS:** Gauthier Gidel, Hugo Berard, Ga?tan Vignoud, Pascal Vincent, Simon Lacoste-Julien  
**TL;DR:** We cast GANs in the variational inequality framework and import techniques from this literature to optimize GANs better; we give algorithmic extensions and empirically test their performance for training GANs.  
**HIGHLIGHT:** In this work, we cast GAN optimization problems in the general variational inequality framework.

382, **TITLE:** Learning-Based Frequency Estimation Algorithms  
<https://openreview.net/forum?id=r1lohoCqY7>  
**AUTHORS:** Chen-Yu Hsu, Piotr Indyk, Dina Katabi, Ali Vakilian  
**TL;DR:** Data stream algorithms can be improved using deep learning, while retaining performance guarantees.  
**HIGHLIGHT:** We propose a new class of algorithms that automatically learn relevant patterns in the input data and use them to improve its frequency estimates.

383, **TITLE:** From Language to Goals: Inverse Reinforcement Learning for Vision-Based Instruction Following  
<https://openreview.net/forum?id=r1lq1hRqYQ>  
**AUTHORS:** Justin Fu, Anoop Korattikara, Sergey Levine, Sergio Guadarrama  
**TL;DR:** We ground language commands in a high-dimensional visual environment by learning language-conditioned rewards using inverse reinforcement learning.  
**HIGHLIGHT:** In this work, we investigate the problem of grounding language commands as reward functions using inverse reinforcement learning, and argue that language-conditioned rewards are more transferable than language-conditioned policies to new environments.

384, **TITLE:** Backpropamine: training self-modifying neural networks with differentiable neuromodulated plasticity  
<https://openreview.net/forum?id=r1lrAiA5Ym>  
**AUTHORS:** Thomas Miconi, Aditya Rawal, Jeff Clune, Kenneth O. Stanley  
**TL;DR:** Neural networks can be trained to modify their own connectivity, improving their online learning performance on challenging tasks.  
**HIGHLIGHT:** Extending previous work on differentiable Hebbian plasticity, we propose a differentiable formulation for the neuromodulation of plasticity.

385, **TITLE:** Recurrent Experience Replay in Distributed Reinforcement Learning  
<https://openreview.net/forum?id=r1lyTjAqYX>  
**AUTHORS:** Steven Kapturowski, Georg Ostrovski, John Quan, Remi Munos, Will Dabney  
**TL;DR:** Investigation on combining recurrent neural networks and experience replay leading to state-of-the-art agent on both Atari-57 and DMLab-30 using single set of hyper-parameters.  
**HIGHLIGHT:** Building on the recent successes of distributed training of RL agents, in this paper we investigate the training of RNN-based RL agents from distributed prioritized experience replay.

- 386, TITLE: A Generative Model For Electron Paths  
<https://openreview.net/forum?id=r1x4BnCqKX>  
AUTHORS: John Bradshaw, Matt J. Kusner, Brooks Paige, Marwin H. S. Segler, Jos? Miguel Hernandez-Lobato  
TL;DR: A generative model for reaction prediction that learns the mechanistic electron steps of a reaction directly from raw reaction data.  
HIGHLIGHT: We propose an electron path prediction model (ELECTRO) to learn these sequences directly from raw reaction data.
- 387, TITLE: Modeling Uncertainty with Hedged Instance Embeddings  
<https://openreview.net/forum?id=r1xQQhAqKX>  
AUTHORS: Seong Joon Oh, Kevin P. Murphy, Jiyan Pan, Joseph Roth, Florian Schroff, Andrew C. Gallagher  
TL;DR: The paper proposes using probability distributions instead of points for instance embeddings tasks such as recognition and verification.  
HIGHLIGHT: We introduce the hedged instance embedding (HIB) in which embeddings are modeled as random variables and the model is trained under the variational information bottleneck principle (Alemi et al., 2016; Achille & Soatto, 2018).
- 388, TITLE: Beyond Greedy Ranking: Slate Optimization via List-CVAE  
<https://openreview.net/forum?id=r1xX42R5Fm>  
AUTHORS: Ray Jiang, Sven Gowal, Yuqiu Qian, Timothy Mann, Danilo J. Rezende  
TL;DR: We used a CVAE type model structure to learn to directly generate slates/whole pages for recommendation systems.  
HIGHLIGHT: In this paper, we introduce List Conditional Variational Auto-Encoders (ListCVAE), which learn the joint distribution of documents on the slate conditioned on user responses, and directly generate full slates.
- 389, TITLE: Stochastic Prediction of Multi-Agent Interactions from Partial Observations  
<https://openreview.net/forum?id=r1xdH3CcKX>  
AUTHORS: Chen Sun, Per Karlsson, Jiajun Wu, Joshua B Tenenbaum, Kevin Murphy  
TL;DR: We present a method which learns to integrate temporal information and ambiguous visual information in the context of interacting agents.  
HIGHLIGHT: We present a method which learns to integrate temporal information, from a learned dynamics model, with ambiguous visual information, from a learned vision model, in the context of interacting agents.
- 390, TITLE: GamePad: A Learning Environment for Theorem Proving  
<https://openreview.net/forum?id=r1xwKoR9Y7>  
AUTHORS: Daniel Huang, Prafulla Dhariwal, Dawn Song, Ilya Sutskever  
TL;DR: We introduce a system called GamePad to explore the application of machine learning methods to theorem proving in the Coq proof assistant.  
HIGHLIGHT: In this paper, we introduce a system called GamePad that can be used to explore the application of machine learning methods to theorem proving in the Coq proof assistant.
- 391, TITLE: GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding  
<https://openreview.net/forum?id=rJ4km2R5t7>  
AUTHORS: Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, Samuel R. Bowman

TL;DR: We present a multi-task benchmark and analysis platform for evaluating generalization in natural language understanding systems.

HIGHLIGHT: In pursuit of this objective, we introduce the General Language Understanding Evaluation (GLUE) benchmark, a collection of tools for evaluating the performance of models across a diverse set of existing NLU tasks.

392, TITLE: On Computation and Generalization of Generative Adversarial Networks under Spectrum Control

<https://openreview.net/forum?id=rJNH6sAqY7>

AUTHORS: Haoming Jiang, Zhehui Chen, Minshuo Chen, Feng Liu, Dingding Wang, Tuo Zhao

TL;DR: We present a multi-task benchmark and analysis platform for evaluating generalization in natural language understanding systems.

HIGHLIGHT: Motivated by their discovery, we propose a new framework for training GANs, which allows more flexible spectrum control (e.g., making the weight matrices of the discriminator have slow singular value decays).

393, TITLE: Large-Scale Study of Curiosity-Driven Learning

<https://openreview.net/forum?id=rJNwDjAqYX>

AUTHORS: Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, Alexei A. Efros

TL;DR: An agent trained only with curiosity, and no extrinsic reward, does surprisingly well on 54 popular environments, including the suite of Atari games, Mario etc.

HIGHLIGHT: In this paper: (a) We perform the first large-scale study of purely curiosity-driven learning, i.e. {em without any extrinsic rewards}, across \$54\$ standard benchmark environments, including the Atari game suite.

394, TITLE: Unsupervised Discovery of Parts, Structure, and Dynamics

<https://openreview.net/forum?id=rJe10iC5K7>

AUTHORS: Zhenjia Xu, Zhijian Liu, Chen Sun, Kevin Murphy, William T. Freeman, Joshua B. Tenenbaum, Jiajun Wu

TL;DR: Learning object parts, hierarchical structure, and dynamics by watching how they move

HIGHLIGHT: In this paper, we propose a novel formulation that simultaneously learns a hierarchical, disentangled object representation and a dynamics model for object parts from unlabeled videos.

395, TITLE: Music Transformer: Generating Music with Long-Term Structure

<https://openreview.net/forum?id=rJe4ShAcF7>

AUTHORS: Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Ian Simon, Curtis Hawthorne, Noam Shazeer, Andrew M. Dai, Matthew D. Hoffman, Monica Dinulescu, Douglas Eck

TL;DR: We show the first successful use of Transformer in generating music that exhibits long-term structure.

HIGHLIGHT: We propose an algorithm that reduces the intermediate memory requirements to linear in the sequence length.

396, TITLE: BabyAI: A Platform to Study the Sample Efficiency of Grounded Language Learning

<https://openreview.net/forum?id=rJeXC0cYX>

AUTHORS: Maxime Chevalier-Boisvert, Dzmitry Bahdanau, Salem Lahlou, Lucas Willems, Chitwan Saharia, Thien Huu Nguyen, Yoshua Bengio

TL;DR: We present the BabyAI platform for studying data efficiency of language learning with a human in the loop

HIGHLIGHT: We introduce the BabyAI research platform, with the goal of supporting investigations towards including humans in the loop for grounded language learning.

397, TITLE: Analyzing Inverse Problems with Invertible Neural Networks  
<https://openreview.net/forum?id=rJed6j0cKX>  
AUTHORS: Lynton Ardizzone, Jakob Kruse, Carsten Rother, Ullrich K?the  
TL;DR: To analyze inverse problems with Invertible Neural Networks  
HIGHLIGHT: We prove theoretically and verify experimentally, onartificial data and real-world problems from medicine and astrophysics, thatINNs are a powerful analysis tool to find multi-modalities in parameter space,uncover parameter correlations, and identify unrecoverable parameters.

398, TITLE: RelGAN: Relational Generative Adversarial Networks for Text Generation  
<https://openreview.net/forum?id=rJedV3R5tm>  
AUTHORS: Weili Nie, Nina Narodytska, Ankit Patel  
TL;DR: To analyze inverse problems with Invertible Neural Networks  
HIGHLIGHT: In this work, we propose RelGAN, a new GAN architecture for text generation, consisting of three main components: a relational memory based generator for the long-distance dependency modeling, the Gumbel-Softmax relaxation for training GANs on discrete data, and multiple embedded representations in the discriminator to provide a more informative signal for the generator updates.

399, TITLE: The Singular Values of Convolutional Layers  
<https://openreview.net/forum?id=rJevYoA9Fm>  
AUTHORS: Hanie Sedghi, Vineet Gupta, Philip M. Long  
TL;DR: We characterize the singular values of the linear transformation associated with a standard 2D multi-channel convolutional layer, enabling their efficient computation.  
HIGHLIGHT: We characterize the singular values of the linear transformation associated with a standard 2D multi-channel convolutional layer, enabling their efficient computation.

400, TITLE: An Empirical study of Binary Neural Networks' Optimisation  
<https://openreview.net/forum?id=rJfUCoR5KX>  
AUTHORS: Milad Alizadeh, Javier Fern?ndez-Marqu?s, Nicholas D. Lane, Yarín Gal  
TL;DR: We characterize the singular values of the linear transformation associated with a standard 2D multi-channel convolutional layer, enabling their efficient computation.  
HIGHLIGHT: In this work, we empirically identify and study the effectiveness of the various ad-hoc techniques commonly used in the literature, providing best-practices for efficient training of binary models.

401, TITLE: Approximability of Discriminators Implies Diversity in GANs  
<https://openreview.net/forum?id=rJfW5oA5KQ>  
AUTHORS: Yu Bai, Tengyu Ma, Andrej Risteski  
TL;DR: GANs can in principle learn distributions sample-efficiently, if the discriminator class is compact and has strong distinguishing power against the particular generator class.  
HIGHLIGHT: The theoretical work of Arora et al. (2017a) suggests a dilemma about GANs? statistical properties: powerful discriminators cause overfitting, whereas weak discriminators cannot detect mode collapse.By contrast, we show in this paper that GANs can in principle learn distributions in Wasserstein distance (or KL-divergence in many cases) with polynomial sample complexity, if the discriminator class has strong distinguishing power against the particular generator class (instead of against all possible generators).

402, TITLE: Learning Embeddings into Entropic Wasserstein Spaces  
<https://openreview.net/forum?id=rJg4J3CqFm>  
AUTHORS: Charlie Frogner, Farzaneh Mirzazadeh, Justin Solomon

TL;DR: We show that Wasserstein spaces are good targets for embedding data with complex semantic structure.

HIGHLIGHT: We propose to exploit this flexibility by learning an embedding that captures the semantic information in the Wasserstein distance between embedded distributions.

403, TITLE: DeepOBS: A Deep Learning Optimizer Benchmark Suite

<https://openreview.net/forum?id=rJg6ssC5Y7>

AUTHORS: Frank Schneider, Lukas Balles, Philipp Hennig

TL;DR: We provide a software package that drastically simplifies, automates, and improves the evaluation of deep learning optimizers.

HIGHLIGHT: As the primary contribution, we present DeepOBS, a Python package of deep learning optimization benchmarks.

404, TITLE: InfoBot: Transfer and Exploration via the Information Bottleneck

<https://openreview.net/forum?id=rJg8yhAqKm>

AUTHORS: Anirudh Goyal, Riashat Islam, DJ Strouse, Zafarali Ahmed, Hugo Larochelle, Matthew Botvinick, Yoshua Bengio, Sergey Levine

TL;DR: Training agents with goal-policy information bottlenecks promotes transfer and yields a powerful exploration bonus

HIGHLIGHT: We propose to learn about decision states from prior experience.

405, TITLE: The Comparative Power of ReLU Networks and Polynomial Kernels in the Presence of Sparse Latent Structure

<https://openreview.net/forum?id=rJgTTjA9tX>

AUTHORS: Frederic Koehler, Andrej Risteski

TL;DR: Beyond-worst-case analysis of the representational power of ReLU nets & polynomial kernels -- in particular in the presence of sparse latent structure.

HIGHLIGHT: We give an almost-tight theoretical analysis of the performance of both neural networks and polynomials for this problem, as well as verify our theory with simulations.

406, TITLE: Learning Implicitly Recurrent CNNs Through Parameter Sharing

<https://openreview.net/forum?id=rJgYxn09Fm>

AUTHORS: Pedro Savarese, Michael Maire

TL;DR: We propose a method that enables CNN folding to create recurrent connections

HIGHLIGHT: We introduce a parameter sharing scheme, in which different layers of a convolutional neural network (CNN) are defined by a learned linear combination of parameter tensors from a global bank of templates.

407, TITLE: Learning Particle Dynamics for Manipulating Rigid Bodies, Deformable Objects, and Fluids

<https://openreview.net/forum?id=rJgbSn09Ym>

AUTHORS: Yunzhu Li, Jiajun Wu, Russ Tedrake, Joshua B. Tenenbaum, Antonio Torralba

TL;DR: Learning particle dynamics with dynamic interaction graphs for simulating and control rigid bodies, deformable objects, and fluids.

HIGHLIGHT: In this paper, we propose to learn a particle-based simulator for complex control tasks.

408, TITLE: Regularized Learning for Domain Adaptation under Label Shifts

<https://openreview.net/forum?id=rJl0r3R9KX>

AUTHORS: Kamyar Azizzadenesheli, Anqi Liu, Fanny Yang, Animashree Anandkumar

TL;DR: A practical and provably guaranteed approach for training efficiently classifiers in the presence of label shifts between Source and Target data sets

HIGHLIGHT: We propose Regularized Learning under Label shifts (RLLS), a principled and a practical domain-adaptation algorithm to correct for shifts in the label distribution between a source and a target domain.

409, TITLE: Von Mises-Fisher Loss for Training Sequence to Sequence Models with Continuous Outputs

<https://openreview.net/forum?id=rJIDnoA5Y7>

AUTHORS: Sachin Kumar, Yulia Tsvetkov

TL;DR: Language generation using seq2seq models which produce word embeddings instead of a softmax based distribution over the vocabulary at each step enabling much faster training while maintaining generation quality

HIGHLIGHT: We propose a general technique for replacing the softmax layer with a continuous embedding layer.

410, TITLE: Relational Forward Models for Multi-Agent Learning

<https://openreview.net/forum?id=rJIEojAqFm>

AUTHORS: Andrea Tacchetti, H. Francis Song, Pedro A. M. Mediano, Vinicius Zambaldi, J?nos Kram?r, Neil C. Rabinowitz, Thore Graepel, Matthew Botvinick, Peter W. Battaglia

TL;DR: Relational Forward Models for multi-agent learning make accurate predictions of agents' future behavior, they produce interpretable representations and can be used inside agents.

HIGHLIGHT: Here we introduce Relational Forward Models (RFM) for multi-agent learning, networks that can learn to make accurate predictions of agents' future behavior in multi-agent environments.

411, TITLE: Imposing Category Trees Onto Word-Embeddings Using A Geometric Construction

<https://openreview.net/forum?id=rJlWOj0qF7>

AUTHORS: Tiansi Dong, Chrisitan Bauckhage, Hailong Jin, Juanzi Li, Olaf Cremers, Daniel Speicher, Armin B. Cremers, Joerg Zimmermann

TL;DR: we show a geometric method to perfectly encode category tree information into pre-trained word-embeddings.

HIGHLIGHT: We present a novel method to precisely impose tree-structured category information onto word-embeddings, resulting in ball embeddings in higher dimensional spaces (N-balls for short).

412, TITLE: Two-Timescale Networks for Nonlinear Value Function Approximation

<https://openreview.net/forum?id=rJleN20qK7>

AUTHORS: Wesley Chung, Somjit Nath, Ajin Joseph, Martha White

TL;DR: We propose an architecture for learning value functions which allows the use of any linear policy evaluation algorithm in tandem with nonlinear feature learning.

HIGHLIGHT: In this work, we provide a two-timescale network (TTN) architecture that enables linear methods to be used to learn values, with a nonlinear representation learned at a slower timescale.

413, TITLE: Diversity-Sensitive Conditional Generative Adversarial Networks

<https://openreview.net/forum?id=rJliMh09F7>

AUTHORS: Dingdong Yang, Seunghoon Hong, Yunseok Jang, Tianchen Zhao, Honglak Lee

TL;DR: We propose a simple and general approach that avoids a mode collapse problem in various conditional GANs.

HIGHLIGHT: We propose a simple yet highly effective method that addresses the mode-collapse problem in the Conditional Generative Adversarial Network (cGAN).

414, TITLE: Query-Efficient Hard-label Black-box Attack: An Optimization-based Approach  
<https://openreview.net/forum?id=rJlk6iRqKX>  
AUTHORS: Minhao Cheng, Thong Le, Pin-Yu Chen, Huan Zhang, JinFeng Yi, Cho-Jui Hsieh  
TL;DR: We propose a simple and general approach that avoids a mode collapse problem in various conditional GANs.  
HIGHLIGHT: We study the problem of attacking machine learning models in the hard-label black-box setting, where no model information is revealed except that the attacker can make queries to probe the corresponding hard-label decisions.

415, TITLE: Rethinking the Value of Network Pruning  
<https://openreview.net/forum?id=rJlnB3C5Ym>  
AUTHORS: Zhuang Liu, Mingjie Sun, Tinghui Zhou, Gao Huang, Trevor Darrell  
TL;DR: In structured network pruning, fine-tuning a pruned model only gives comparable performance with training it from scratch.  
HIGHLIGHT: In this work, we make several surprising observations which contradict common beliefs.

416, TITLE: Hyperbolic Attention Networks  
<https://openreview.net/forum?id=rJxHsjRqFQ>  
AUTHORS: Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro  
TL;DR: We propose to incorporate inductive biases and operations coming from hyperbolic geometry to improve the attention mechanism of the neural networks.  
HIGHLIGHT: Recent approaches have successfully demonstrated the benefits of learning the parameters of shallow networks in hyperbolic space.

417, TITLE: Learning from Positive and Unlabeled Data with a Selection Bias  
<https://openreview.net/forum?id=rJzLciCqKm>  
AUTHORS: Masahiro Kato, Takeshi Teshima, Junya Honda  
TL;DR: We propose to incorporate inductive biases and operations coming from hyperbolic geometry to improve the attention mechanism of the neural networks.  
HIGHLIGHT: In this paper, we propose a method to partially identify the classifier.

418, TITLE: Adv-BNN: Improved Adversarial Defense through Robust Bayesian Neural Network  
<https://openreview.net/forum?id=rk4Qso0cKm>  
AUTHORS: Xuanqing Liu, Yao Li, Chongruo Wu, Cho-Jui Hsieh  
TL;DR: We design an adversarial training method to Bayesian neural networks, showing a much stronger defense to white-box adversarial attacks  
HIGHLIGHT: We present a new algorithm to train a robust neural network against adversarial attacks. Our algorithm is motivated by the following two ideas.

419, TITLE: Optimal Completion Distillation for Sequence Learning  
<https://openreview.net/forum?id=rkMW1hRqKX>  
AUTHORS: Sara Sabour, William Chan, Mohammad Norouzi  
TL;DR: Optimal Completion Distillation (OCD) is a training procedure for optimizing sequence to sequence models based on edit distance which achieves state-of-the-art on end-to-end Speech Recognition tasks.  
HIGHLIGHT: We present Optimal Completion Distillation (OCD), a training procedure for optimizing sequence to sequence models based on edit distance.

420, TITLE: Caveats for information bottleneck in deterministic scenarios

<https://openreview.net/forum?id=rke4HiAcY7>

AUTHORS: Artemy Kolchinsky, Brendan D. Tracey, Steven Van Kuyk

TL;DR: Information bottleneck behaves in surprising ways whenever the output is a deterministic function of the input.

HIGHLIGHT: To address problem (1), we propose a functional that, unlike the IB Lagrangian, can recover the IB curve in all cases.

421, TITLE: Deep Learning 3D Shapes Using Alt-az Anisotropic 2-Sphere Convolution

<https://openreview.net/forum?id=rkeSiiA5Fm>

AUTHORS: Min Liu, Fupin Yao, Chiho Choi, Ayan Sinha, Karthik Ramani

TL;DR: A method for applying deep learning to 3D surfaces using their spherical descriptors and alt-az anisotropic convolution on 2-sphere.

HIGHLIGHT: In this paper, we present a method for applying deep learning to 3D surfaces using their spherical descriptors and alt-az anisotropic convolution on 2-sphere.

422, TITLE: Small nonlinearities in activation functions create bad local minima in neural networks

[https://openreview.net/forum?id=rke\\_YiRct7](https://openreview.net/forum?id=rke_YiRct7)

AUTHORS: Chulhee Yun, Suvrit Sra, Ali Jadbabaie

TL;DR: We constructively prove that even the slightest nonlinear activation functions introduce spurious local minima, for general datasets and activation functions.

HIGHLIGHT: We investigate the loss surface of neural networks.

423, TITLE: Information Theoretic lower bounds on negative log likelihood

<https://openreview.net/forum?id=rkemqsC9Fm>

AUTHORS: Luis A. Lastras-Montaño

TL;DR: Use rate-distortion theory to bound how much a latent variable model can be improved

HIGHLIGHT: In this article we use rate-distortion theory, a branch of information theory devoted to the problem of lossy compression, to shed light on an important problem in latent variable modeling of data: is there room to improve the model?

424, TITLE: Preferences Implicit in the State of the World

<https://openreview.net/forum?id=rkevMnRqYQ>

AUTHORS: Rohin Shah, Dmitrii Krasheninnikov, Jordan Alexander, Pieter Abbeel, Anca Dragan

TL;DR: When a robot is deployed in an environment that humans have been acting in, the state of the environment is already optimized for what humans want, and we can use this to infer human preferences.

HIGHLIGHT: This motivates our key insight: when a robot is deployed in an environment that humans act in, the state of the environment is already optimized for what humans want.

425, TITLE: A Kernel Random Matrix-Based Approach for Sparse PCA

<https://openreview.net/forum?id=rkgBHoCqYX>

AUTHORS: Mohamed El Amine Seddik, Mohamed Tamaazousti, Romain Couillet

TL;DR: When a robot is deployed in an environment that humans have been acting in, the state of the environment is already optimized for what humans want, and we can use this to infer human preferences.

HIGHLIGHT: In this paper, we present a random matrix approach to recover sparse principal components from  $n$   $p$ -dimensional vectors.

426, TITLE: Bayesian Prediction of Future Street Scenes using Synthetic Likelihoods

<https://openreview.net/forum?id=rkgK3oC5Fm>

AUTHORS: Apratim Bhattacharyya, Mario Fritz, Bernt Schiele

TL;DR: Dropout based Bayesian inference is extended to deal with multi-modality and is evaluated on scene anticipation tasks.

HIGHLIGHT: In this work, we propose a novel Bayesian formulation for anticipating future scene states which leverages synthetic likelihoods that encourage the learning of diverse models to accurately capture the multi-modal nature of future scene states.

427, TITLE: There Are Many Consistent Explanations of Unlabeled Data: Why You Should Average  
<https://openreview.net/forum?id=rkgKBhA5Y7>

AUTHORS: Ben Athiwaratkun, Marc Finzi, Pavel Izmailov, Andrew Gordon Wilson

TL;DR: Consistency-based models for semi-supervised learning do not converge to a single point but continue to explore a diverse set of plausible solutions on the perimeter of a flat region. Weight averaging helps improve generalization performance.

HIGHLIGHT: Motivated by these observations, we propose to train consistency-based methods with Stochastic Weight Averaging (SWA), a recent approach which averages weights along the trajectory of SGD with a modified learning rate schedule.

428, TITLE: Large-Scale Answerer in Questioner's Mind for Visual Dialog Question Generation  
<https://openreview.net/forum?id=rkgT3jRct7>

AUTHORS: Sang-Woo Lee, Tong Gao, Sohee Yang, Jaejun Yoo, Jung-Woo Ha

TL;DR: Consistency-based models for semi-supervised learning do not converge to a single point but continue to explore a diverse set of plausible solutions on the perimeter of a flat region. Weight averaging helps improve generalization performance.

HIGHLIGHT: To address this, we propose AQM+ that can deal with a large-scale problem and ask a question that is more coherent to the current context of the dialog.

429, TITLE: Graph HyperNetworks for Neural Architecture Search

<https://openreview.net/forum?id=rkgW0oA9FX>

AUTHORS: Chris Zhang, Mengye Ren, Raquel Urtasun

TL;DR: Consistency-based models for semi-supervised learning do not converge to a single point but continue to explore a diverse set of plausible solutions on the perimeter of a flat region. Weight averaging helps improve generalization performance.

HIGHLIGHT: In this work, we propose the Graph HyperNetwork (GHN) to amortize the search cost: given an architecture, it directly generates the weights by running inference on a graph neural network.

430, TITLE: DELTA: DEEP LEARNING TRANSFER USING FEATURE MAP WITH ATTENTION FOR CONVOLUTIONAL NETWORKS

<https://openreview.net/forum?id=rkgbwsAcYm>

AUTHORS: Xingjian Li, Haoyi Xiong, Hanchao Wang, Yuxuan Rao, Liping Liu, Jun Huan

TL;DR: improving deep transfer learning with regularization using attention based feature maps

HIGHLIGHT: In this paper, we propose a novel regularized transfer learning framework DELTA, namely DEep Learning Transfer using Feature Map with Attention.

431, TITLE: textTOvec: DEEP CONTEXTUALIZED NEURAL AUTOREGRESSIVE TOPIC MODELS OF LANGUAGE WITH DISTRIBUTED COMPOSITIONAL PRIOR

<https://openreview.net/forum?id=rkgoy09KQ>

AUTHORS: Pankaj Gupta, Yatin Chaudhary, Florian Buettner, Hinrich Schuetze

TL;DR: Unified neural model of topic and language modeling to introduce language structure in topic models for contextualized topic vectors

HIGHLIGHT: In this work, we incorporate language structure by combining a neural autoregressive topic model (TM) with a LSTM based language model (LSTM-LM) in a single probabilistic framework.

432, TITLE: Amortized Bayesian Meta-Learning  
<https://openreview.net/forum?id=rkgpy3C5tX>  
AUTHORS: Sachin Ravi, Alex Beatson  
TL;DR: We propose a meta-learning method which efficiently amortizes hierarchical variational inference across training episodes.  
HIGHLIGHT: We propose a meta-learning method which efficiently amortizes hierarchical variational inference across tasks, learning a prior distribution over neural network weights so that a few steps of Bayes by Backprop will produce a good task-specific approximate posterior.

433, TITLE: Probabilistic Recursive Reasoning for Multi-Agent Reinforcement Learning  
<https://openreview.net/forum?id=rkl6As0cF7>  
AUTHORS: Ying Wen, Yaodong Yang, Rui Luo, Jun Wang, Wei Pan  
TL;DR: We proposed a novel probabilistic recursive reasoning (PR2) framework for multi-agent deep reinforcement learning tasks.  
HIGHLIGHT: In this paper, we start from level-1 recursion and introduce a probabilistic recursive reasoning (PR2) framework for multi-agent reinforcement learning.

434, TITLE: Learning Neural PDE Solvers with Convergence Guarantees  
<https://openreview.net/forum?id=rklaWn0qK7>  
AUTHORS: Jun-Ting Hsieh, Shengjia Zhao, Stephan Eismann, Lucia Mirabella, Stefano Ermon  
TL;DR: We learn a fast neural solver for PDEs that has convergence guarantees.  
HIGHLIGHT: In contrast to existing hand-crafted solutions, we propose an approach to learn a fast iterative solver tailored to a specific domain.

435, TITLE: A new dog learns old tricks: RL finds classic optimization algorithms  
<https://openreview.net/forum?id=rkluJ2R9KQ>  
AUTHORS: Weiwei Kong, Christopher Liaw, Aranyak Mehta, D. Sivakumar  
TL;DR: By combining ideas from traditional algorithms design and reinforcement learning, we introduce a novel framework for learning algorithms that solve online combinatorial optimization problems.  
HIGHLIGHT: This paper introduces a novel framework for learning algorithms to solve online combinatorial optimization problems.

436, TITLE: Deep Graph Infomax  
<https://openreview.net/forum?id=rklz9iAcKQ>  
AUTHORS: Petar Velickovic, William Fedus, William L. Hamilton, Pietro Li, Yoshua Bengio, R Devon Hjelm  
TL;DR: A new method for unsupervised representation learning on graphs, relying on maximizing mutual information between local and global representations in a graph. State-of-the-art results, competitive with supervised learning.  
HIGHLIGHT: We present Deep Graph Infomax (DGI), a general approach for learning node representations within graph-structured data in an unsupervised manner.

437, TITLE: Theoretical Analysis of Auto Rate-Tuning by Batch Normalization  
<https://openreview.net/forum?id=rkxQ-nA9FX>  
AUTHORS: Sanjeev Arora, Zhiyuan Li, Kaifeng Lyu  
TL;DR: We give a theoretical analysis of the ability of batch normalization to automatically tune learning rates, in the context of finding stationary points for a deep learning objective.  
HIGHLIGHT: It is shown that even if we fix the learning rate of scale-invariant parameters (e.g., weights of each layer with BN) to a constant (say, 0.3), gradient descent still approaches a stationary point

(i.e., a solution where gradient is zero) in the rate of  $T^{-1/2}$  in  $T$  iterations, asymptotically matching the best bound for gradient descent with well-tuned learning rates.

438, TITLE: Per-Tensor Fixed-Point Quantization of the Back-Propagation Algorithm  
<https://openreview.net/forum?id=rkxaNjA9Ym>  
AUTHORS: Charbel Sakr, Naresh Shanbhag  
TL;DR: We analyze and determine the precision requirements for training neural networks when all tensors, including back-propagated signals and weight accumulators, are quantized to fixed-point format.  
HIGHLIGHT: We describe a precision assignment methodology for neural network training in which all network parameters, i.e., activations and weights in the feedforward path, gradients and weight accumulators in the feedback path, are assigned close to minimal precision.

439, TITLE: FUNCTIONAL VARIATIONAL BAYESIAN NEURAL NETWORKS  
<https://openreview.net/forum?id=rkxacs0qY7>  
AUTHORS: Shengyang Sun, Guodong Zhang, Jiaxin Shi, Roger Grosse  
TL;DR: We perform functional variational inference on the stochastic processes defined by Bayesian neural networks.  
HIGHLIGHT: Based on this, we introduce a practical training objective which approximates the functional ELBO using finite measurement sets and the spectral Stein gradient estimator.

440, TITLE: NADPEX: An on-policy temporally consistent exploration method for deep reinforcement learning  
<https://openreview.net/forum?id=rkxciiC9tm>  
AUTHORS: Sirui Xie, Junning Huang, Lanxin Lei, Chunxiao Liu, Zheng Ma, Wei Zhang, Liang Lin  
TL;DR: We perform functional variational inference on the stochastic processes defined by Bayesian neural networks.  
HIGHLIGHT: In this work, we introduce a novel on-policy temporally consistent exploration strategy - Neural Adaptive Dropout Policy Exploration (NADPEX) - for deep reinforcement learning agents.

441, TITLE: SPIGAN: Privileged Adversarial Learning from Simulation  
<https://openreview.net/forum?id=rkxoNnC5FQ>  
AUTHORS: Kuan-Hui Lee, German Ros, Jie Li, Adrien Gaidon  
TL;DR: An unsupervised sim-to-real domain adaptation method for semantic segmentation using privileged information from a simulator with GAN-based image translation.  
HIGHLIGHT: We propose a new unsupervised domain adaptation algorithm, called SPIGAN, relying on Simulator Privileged Information (PI) and Generative Adversarial Networks (GAN).

442, TITLE: Generating Multi-Agent Trajectories using Programmatic Weak Supervision  
<https://openreview.net/forum?id=rkxw-hAcFQ>  
AUTHORS: Eric Zhan, Stephan Zheng, Yisong Yue, Long Sha, Patrick Lucey  
TL;DR: We blend deep generative models with programmatic weak supervision to generate coordinated multi-agent trajectories of significantly higher quality than previous baselines.  
HIGHLIGHT: We present a hierarchical framework that can effectively learn such sequential generative models.

443, TITLE: Label super-resolution networks  
<https://openreview.net/forum?id=rkxwShA9Ym>  
AUTHORS: Kolya Malkin, Caleb Robinson, Le Hou, Rachel Soobitsky, Jacob Czawlytko, Dimitris Samaras, Joel Saltz, Lucas Joppa, Nebojsa Jojic  
TL;DR: Super-resolving coarse labels into pixel-level labels, applied to aerial imagery and medical scans.

**HIGHLIGHT:** We present a deep learning-based method for super-resolving coarse (low-resolution) labels assigned to groups of image pixels into pixel-level (high-resolution) labels, given the joint distribution between those low- and high-resolution labels.

444, **TITLE:** ANYTIME MINIBATCH: EXPLOITING STRAGGLERS IN ONLINE DISTRIBUTED OPTIMIZATION

<https://openreview.net/forum?id=rkzDliA5YQ>

**AUTHORS:** Nuwan Ferdinand, Haider Al-Lawati, Stark Draper, Matthew Nokleby

**TL;DR:** Accelerate distributed optimization by exploiting stragglers.

**HIGHLIGHT:** To mitigate the impact of stragglers, we propose an online distributed optimization method called Anytime Minibatch.

445, **TITLE:** Sample Efficient Adaptive Text-to-Speech

<https://openreview.net/forum?id=rkzjUoAcFX>

**AUTHORS:** Yutian Chen, Yannis Assael, Brendan Shillingford, David Budden, Scott Reed, Heiga Zen, Quan Wang, Luis C. Cobo, Andrew Trask, Ben Laurie, Caglar Gulcehre, Aron van den Oord, Oriol Vinyals, Nando de Freitas

**TL;DR:** Sample efficient algorithms to adapt a text-to-speech model to a new voice style with the state-of-the-art performance.

**HIGHLIGHT:** We present a meta-learning approach for adaptive text-to-speech (TTS) with few data.

446, **TITLE:** Practical lossless compression with latent variables using bits back coding

<https://openreview.net/forum?id=ryE98iR5tm>

**AUTHORS:** James Townsend, Thomas Bird, David Barber

**TL;DR:** We do lossless compression of large image datasets using a VAE, beat existing compression algorithms.

**HIGHLIGHT:** We present 'Bits Back with ANS' (BB-ANS), a scheme to perform lossless compression with latent variable models at a near optimal rate.

447, **TITLE:** Kernel RNN Learning (KeRNL)

<https://openreview.net/forum?id=ryGfnoC5KQ>

**AUTHORS:** Christopher Roth, Ingmar Kanitscheider, Ila Fiete

**TL;DR:** A biologically plausible learning rule for training recurrent neural networks

**HIGHLIGHT:** We describe Kernel RNN Learning (KeRNL), a reduced-rank, temporal eligibility trace-based approximation to backpropagation through time (BPTT) for training recurrent neural networks (RNNs) that gives competitive performance to BPTT on long time-dependence tasks.

448, **TITLE:** Deep, Skinny Neural Networks are not Universal Approximators

<https://openreview.net/forum?id=ryGgSsAcFQ>

**AUTHORS:** Jesse Johnson

**TL;DR:** This paper proves that skinny neural networks cannot approximate certain functions, no matter how deep they are.

**HIGHLIGHT:** In this paper, we examine the topological constraints that the architecture of a neural network imposes on the level sets of all the functions that it is able to approximate.

449, **TITLE:** Large Scale Graph Learning From Smooth Signals

<https://openreview.net/forum?id=ryGkSo0qYm>

**AUTHORS:** Vassilis Kalofolias, Nathanael Perraudin

**TL;DR:** This paper proves that skinny neural networks cannot approximate certain functions, no matter how deep they are.

**HIGHLIGHT:** Our algorithm uses known approximate nearest neighbor techniques to reduce the number of variables, and automatically selects the correct parameters of the model, requiring a single intuitive input: the desired edge density.

450, **TITLE:** Overcoming Catastrophic Forgetting for Continual Learning via Model Adaptation  
<https://openreview.net/forum?id=ryGvcoA5YX>  
**AUTHORS:** Wenpeng Hu, Zhou Lin, Bing Liu, Chongyang Tao, Zhengwei Tao, Jinwen Ma, Dongyan Zhao, Rui Yan  
**TL;DR:** This paper proves that skinny neural networks cannot approximate certain functions, no matter how deep they are.  
**HIGHLIGHT:** In this paper, we propose a very different approach, called Parameter Generation and Model Adaptation (PGMA), to dealing with the problem.

451, **TITLE:** Analysis of Quantized Models  
[https://openreview.net/forum?id=ryM\\_IoAqYX](https://openreview.net/forum?id=ryM_IoAqYX)  
**AUTHORS:** Lu Hou, Ruiliang Zhang, James T. Kwok  
**TL;DR:** In this paper, we studied efficient training of loss-aware weight-quantized networks with quantized gradient in a distributed environment, both theoretically and empirically.  
**HIGHLIGHT:** In recent years, many weight-quantized models have been proposed.

452, **TITLE:** Deep learning generalizes because the parameter-function map is biased towards simple functions  
<https://openreview.net/forum?id=rye4g3AqFm>  
**AUTHORS:** Guillermo Valle-Perez, Chico Q. Camargo, Ard A. Louis  
**TL;DR:** The parameter-function map of deep networks is hugely biased; this can explain why they generalize. We use PAC-Bayes and Gaussian processes to obtain nonvacuous bounds.  
**HIGHLIGHT:** In this paper, we provide a new explanation.

453, **TITLE:** Learning when to Communicate at Scale in Multiagent Cooperative and Competitive Tasks  
<https://openreview.net/forum?id=rye7knCqK7>  
**AUTHORS:** Amanpreet Singh, Tushar Jain, Sainbayar Sukhbaatar  
**TL;DR:** We introduce IC3Net, a single network which can be used to train agents in cooperative, competitive and mixed scenarios. We also show that agents can learn when to communicate using our model.  
**HIGHLIGHT:** In this paper, we present Individualized Controlled Continuous Communication Model (IC3Net) which has better training efficiency than simple continuous communication model, and can be applied to semi-cooperative and competitive settings along with the cooperative settings.

454, **TITLE:** Synthetic Datasets for Neural Program Synthesis  
<https://openreview.net/forum?id=ryeOSnAqYm>  
**AUTHORS:** Richard Shin, Neel Kant, Kavi Gupta, Chris Bender, Brandon Trabucco, Rishabh Singh, Dawn Song  
**TL;DR:** We introduce IC3Net, a single network which can be used to train agents in cooperative, competitive and mixed scenarios. We also show that agents can learn when to communicate using our model.  
**HIGHLIGHT:** The goal of program synthesis is to automatically generate programs in a particular language from corresponding specifications, e.g. input-output behavior. Many current approaches achieve impressive results after training on randomly generated I/O examples in limited domain-specific languages (DSLs), as with string transformations in RobustFill. However, we empirically discover that applying test input generation techniques for languages with control flow and rich input space causes deep networks to

generalize poorly to certain data distributions; to correct this, we propose a new methodology for controlling and evaluating the bias of synthetic data distributions over both programs and specifications. We demonstrate, using the Karel DSL and a small Calculator DSL, that training deep networks on these distributions leads to improved cross-distribution generalization performance.

455, TITLE: DPSNet: End-to-end Deep Plane Sweep Stereo  
<https://openreview.net/forum?id=ryeYHi0ctQ>  
AUTHORS: Sunghoon Im, Hae-Gon Jeon, Stephen Lin, In So Kweon  
TL;DR: A convolution neural network for multi-view stereo matching whose design is inspired by best practices of traditional geometry-based approaches  
HIGHLIGHT: In this paper, we present a convolutional neural network called DPSNet (Deep Plane Sweep Network) whose design is inspired by best practices of traditional geometry-based approaches.

456, TITLE: Conditional Network Embeddings  
<https://openreview.net/forum?id=ryepUj0qtX>  
AUTHORS: Bo Kang, Jeffrey Lijffijt, Tijl De Bie  
TL;DR: We introduce a network embedding method that accounts for prior information about the network, yielding superior empirical performance.  
HIGHLIGHT: In recent years various methods for NE have been introduced, all following a similar strategy: defining a notion of similarity between nodes (typically some distance measure within the network), a distance measure in the embedding space, and a loss function that penalizes large distances for similar nodes and small distances for dissimilar nodes. A difficulty faced by existing methods is that certain networks are fundamentally hard to embed due to their structural properties: (approximate) multipartiteness, certain degree distributions, assortativity, etc.

457, TITLE: Defensive Quantization: When Efficiency Meets Robustness  
<https://openreview.net/forum?id=ryetZ20ctX>  
AUTHORS: Ji Lin, Chuang Gan, Song Han  
TL;DR: We designed a novel quantization methodology to jointly optimize the efficiency and robustness of deep learning models.  
HIGHLIGHT: This paper aims to raise people's awareness about the security of the quantized models, and we designed a novel quantization methodology to jointly optimize the efficiency and robustness of deep learning models.

458, TITLE: GO Gradient for Expectation-Based Objectives  
<https://openreview.net/forum?id=ryf6Fs09YX>  
AUTHORS: Yulai Cong, Miaoyun Zhao, Ke Bai, Lawrence Carin  
TL;DR: a Rep-like gradient for non-reparameterizable continuous/discrete distributions; further generalized to deep probabilistic models, yielding statistical back-propagation  
HIGHLIGHT: To address these limitations, we propose a General and One-sample (GO) gradient that ( $\mathcal{G}$ ) applies to many distributions associated with non-reparameterizable continuous or discrete random variables, and ( $\mathcal{G}$ ) has the same low-variance as the reparameterization trick.

459, TITLE: h-detach: Modifying the LSTM Gradient Towards Better Optimization  
<https://openreview.net/forum?id=ryf7ioRqFX>  
AUTHORS: Bhargav Kanuparthi, Devansh Arpit, Giancarlo Kerg, Nan Rosemary Ke, Ioannis Mitliagkas, Yoshua Bengio  
TL;DR: A simple algorithm to improve optimization and handling of long term dependencies in LSTM  
HIGHLIGHT: We introduce a simple stochastic algorithm (h-detach) that is specific to LSTM optimization and targeted towards addressing this problem.

460, TITLE: An analytic theory of generalization dynamics and transfer learning in deep linear networks  
<https://openreview.net/forum?id=ryfMLoCqtQ>  
AUTHORS: Andrew K. Lampinen, Surya Ganguli  
TL;DR: We provide many insights into neural network generalization from the theoretically tractable linear case.  
HIGHLIGHT: We develop an analytic theory of the nonlinear dynamics of generalization in deep linear networks, both within and across tasks.

461, TITLE: Differentiable Learning-to-Normalize via Switchable Normalization  
<https://openreview.net/forum?id=ryggIs0cYQ>  
AUTHORS: Ping Luo, Jiamin Ren, Zhanglin Peng, Ruimao Zhang, Jingyu Li  
TL;DR: We provide many insights into neural network generalization from the theoretically tractable linear case.  
HIGHLIGHT: We address a learning-to-normalize problem by proposing Switchable Normalization (SN), which learns to select different normalizers for different normalization layers of a deep neural network.

462, TITLE: SOM-VAE: Interpretable Discrete Representation Learning on Time Series  
<https://openreview.net/forum?id=rygicsR9Y7>  
AUTHORS: Vincent Fortuin, Matthias H?ser, Francesco Locatello, Heiko Strathmann, Gunnar R?tsch  
TL;DR: We present a method to learn interpretable representations on time series using ideas from variational autoencoders, self-organizing maps and probabilistic models.  
HIGHLIGHT: This is due to non-intuitive mappings from data features to salient properties of the representation and non-smoothness over time. To address this problem, we propose a new representation learning framework building on ideas from interpretable discrete dimensionality reduction and deep generative modeling.

463, TITLE: Hierarchical Generative Modeling for Controllable Speech Synthesis  
<https://openreview.net/forum?id=rygkk305YQ>  
AUTHORS: 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  
TL;DR: Building a TTS model with Gaussian Mixture VAEs enables fine-grained control of speaking style, noise condition, and more.  
HIGHLIGHT: This paper proposes a neural end-to-end text-to-speech (TTS) model which can control latent attributes in the generated speech that are rarely annotated in the training data, such as speaking style, accent, background noise, and recording conditions.

464, TITLE: Learning Factorized Multimodal Representations  
<https://openreview.net/forum?id=rygqqsA9KX>  
AUTHORS: Yao-Hung Hubert Tsai, Paul Pu Liang, Amir Zadeh, Louis-Philippe Morency, Ruslan Salakhutdinov  
TL;DR: We propose a model to learn factorized multimodal representations that are discriminative, generative, and interpretable.  
HIGHLIGHT: In this paper, we propose to optimize for a joint generative-discriminative objective across multimodal data and labels.

465, TITLE: Composing Complex Skills by Learning Transition Policies  
<https://openreview.net/forum?id=rygrBhC5tQ>  
AUTHORS: Youngwoon Lee\*, Shao-Hua Sun\*, Sriram Somasundaram, Edward S. Hu, Joseph J. Lim

TL;DR: Transition policies enable agents to compose complex skills by smoothly connecting previously acquired primitive skills.

HIGHLIGHT: To empower machines with this ability, we propose a method that can learn transition policies which effectively connect primitive skills to perform sequential tasks without handcrafted rewards.

466, TITLE: Human-level Protein Localization with Convolutional Neural Networks

<https://openreview.net/forum?id=ryl5khRcKm>

AUTHORS: Elisabeth Rumetshofer, Markus Hofmarcher, Clemens R?hrl, Sepp Hochreiter, G?nter Klambauer

TL;DR: Transition policies enable agents to compose complex skills by smoothly connecting previously acquired primitive skills.

HIGHLIGHT: We present the largest comparison of CNN architectures including GapNet-PL for protein localization in HTI images of human cells.

467, TITLE: Environment Probing Interaction Policies

<https://openreview.net/forum?id=ryl8-3AcFX>

AUTHORS: Wenxuan Zhou, Lerrel Pinto, Abhinav Gupta

TL;DR: Transition policies enable agents to compose complex skills by smoothly connecting previously acquired primitive skills.

HIGHLIGHT: In this work, we propose the ?Environment-Probing? Interaction (EPI) policy, a policy that probes a new environment to extract an implicit understanding of that environment?s behavior.

468, TITLE: Lagging Inference Networks and Posterior Collapse in Variational Autoencoders

<https://openreview.net/forum?id=rylDfnCqF7>

AUTHORS: Junxian He, Daniel Spokoyny, Graham Neubig, Taylor Berg-Kirkpatrick

TL;DR: To address posterior collapse in VAEs, we propose a novel yet simple training procedure that aggressively optimizes inference network with more updates. This new training procedure mitigates posterior collapse and leads to a better VAE model.

HIGHLIGHT: In this paper, we investigate posterior collapse from the perspective of training dynamics.

469, TITLE: A2BCD: Asynchronous Acceleration with Optimal Complexity

<https://openreview.net/forum?id=rylIAsCqYm>

AUTHORS: Robert Hannah, Fei Feng, Wotao Yin

TL;DR: In this paper, we propose the Asynchronous Accelerated Nonuniform Randomized Block Coordinate Descent algorithm (A2BCD). We prove A2BCD converges linearly to a solution of the convex minimization problem at the same rate as NU\_ACDM, so long as the maximum delay is not too large. This is the first asynchronous Nesterov-accelerated algorithm that attains any provable speedup. Moreover, we then prove that these algorithms both have optimal complexity. Asynchronous algorithms complete much faster iterations, and A2BCD has optimal complexity. Hence we observe in experiments that A2BCD is the top-performing coordinate descent algorithm, converging up to 4-5x faster than NU\_ACDM on some data sets in terms of wall-clock time. To motivate our theory and proof techniques, we also derive and analyze a continuous-time analog of our algorithm and prove it converges at the same rate.

HIGHLIGHT: A2BCD: Asynchronous Acceleration with Optimal Complexity.

470, TITLE: Learning to Infer and Execute 3D Shape Programs

<https://openreview.net/forum?id=rylNH20qFQ>

AUTHORS: Yonglong Tian, Andrew Luo, Xingyuan Sun, Kevin Ellis, William T. Freeman, Joshua B. Tenenbaum, Jiajun Wu

TL;DR: We propose 3D shape programs, a structured, compositional shape representation. Our model learns to infer and execute shape programs to explain 3D shapes.

**HIGHLIGHT:** In this paper, we propose 3D shape programs, integrating bottom-up recognition systems with top-down, symbolic program structure to capture both low-level geometry and high-level structural priors for 3D shapes.

471, **TITLE:** Deep Decoder: Concise Image Representations from Untrained Non-convolutional Networks

<https://openreview.net/forum?id=rylV-2C9KQ>

**AUTHORS:** Reinhard Heckel, Paul Hand

**TL;DR:** We introduce an underparameterized, nonconvolutional, and simple deep neural network that can, without training, effectively represent natural images and solve image processing tasks like compression and denoising competitively.

**HIGHLIGHT:** Contrary to classical tools such as wavelets, image-generating deep neural networks have a large number of parameters---typically a multiple of their output dimension---and need to be trained on large datasets. In this paper, we propose an untrained simple image model, called the deep decoder, which is a deep neural network that can generate natural images from very few weight parameters. The deep decoder has a simple architecture with no convolutions and fewer weight parameters than the output dimensionality.

472, **TITLE:** SNAS: stochastic neural architecture search

<https://openreview.net/forum?id=rylqooRqK7>

**AUTHORS:** Sirui Xie, Hehui Zheng, Chunxiao Liu, Liang Lin

**TL;DR:** We introduce an underparameterized, nonconvolutional, and simple deep neural network that can, without training, effectively represent natural images and solve image processing tasks like compression and denoising competitively.

**HIGHLIGHT:** In this work, NAS is reformulated as an optimization problem on parameters of a joint distribution for the search space in a cell.

473, **TITLE:** Revealing interpretable object representations from human behavior

<https://openreview.net/forum?id=ryxSrhC9KX>

**AUTHORS:** Charles Y. Zheng, Francisco Pereira, Chris I. Baker, Martin N. Hebart

**TL;DR:** Human behavioral judgments are used to obtain sparse and interpretable representations of objects that generalize to other tasks

**HIGHLIGHT:** Revealing interpretable object representations from human behavior

474, **TITLE:** AntisymmetricRNN: A Dynamical System View on Recurrent Neural Networks

<https://openreview.net/forum?id=ryxepo0cFX>

**AUTHORS:** Bo Chang, Minmin Chen, Eldad Haber, Ed H. Chi

**TL;DR:** Human behavioral judgments are used to obtain sparse and interpretable representations of objects that generalize to other tasks

**HIGHLIGHT:** In this paper, we draw connections between recurrent networks and ordinary differential equations.

475, **TITLE:** Global-to-local Memory Pointer Networks for Task-Oriented Dialogue

<https://openreview.net/forum?id=ryxnHhRqFm>

**AUTHORS:** Chien-Sheng Wu, Richard Socher, Caiming Xiong

**TL;DR:** GLMP: Global memory encoder (context RNN, global pointer) and local memory decoder (sketch RNN, local pointer) that share external knowledge (MemNN) are proposed to strengthen response generation in task-oriented dialogue.

**HIGHLIGHT:** We propose the global-to-local memory pointer (GLMP) networks to address this issue.

476, **TITLE:** InstaGAN: Instance-aware Image-to-Image Translation

<https://openreview.net/forum?id=ryxwJhC9YX>

AUTHORS: Sangwoo Mo, Minsu Cho, Jinwoo Shin

TL;DR: We propose a novel method to incorporate the set of instance attributes for image-to-image translation.

HIGHLIGHT: To tackle the issues, we propose a novel method, coined instance-aware GAN (InstaGAN), that incorporates the instance information (e.g., object segmentation masks) and improves multi-instance transfiguration.

477, TITLE: Deep Layers as Stochastic Solvers

<https://openreview.net/forum?id=ryxxCiRqYX>

AUTHORS: Adel Bibi, Bernard Ghanem, Vladlen Koltun, Rene Ranftl

TL;DR: A framework that links deep network layers to stochastic optimization algorithms; can be used to improve model accuracy and inform network design.

HIGHLIGHT: We provide a novel perspective on the forward pass through a block of layers in a deep network.

478, TITLE: Learning Multi-Level Hierarchies with Hindsight

<https://openreview.net/forum?id=ryzECoAcY7>

AUTHORS: Andrew Levy, George Konidaris, Robert Platt, Kate Saenko

TL;DR: We introduce the first Hierarchical RL approach to successfully learn 3-level hierarchies in parallel in tasks with continuous state and action spaces.

HIGHLIGHT: To address this problem, we introduce a framework that can learn multiple levels of policies in parallel.