1, TITLE: Minimal-Entropy Correlation Alignment for Unsupervised Deep Domain Adaptation
https://openreview.net/forum?id=rJWechg0Z
AUTHORS: Pietro Morerio, Jacopo Cavazza, Vittorio Murino
TL;DR: A new unsupervised deep domain adaptation technique which efficiently unifies correlation alignment and entropy minimization
HIGHLIGHT: In this work, we face the problem of unsupervised domain adaptation with a novel deep learning approach which leverages our finding that entropy minimization is induced by the optimal alignment of second order statistics between source and target domains.

2, TITLE: Large Scale Optimal Transport and Mapping Estimation
https://openreview.net/forum?id=B1zlp1bRW
AUTHORS: Vivien Seguy, Bharath Bhushan Damodaran, Remi Flamary, Nicolas Courty, Antoine Rolet, Mathieu Blondel
TL;DR: Learning optimal mapping with deepNN between distributions along with theoretical guarantees.
HIGHLIGHT: This paper presents a novel two-step approach for the fundamental problem of learning an optimal map from one distribution to another.

3, TITLE: TRUNCATED HORIZON POLICY SEARCH: COMBINING REINFORCEMENT LEARNING & IMITATION LEARNING
https://openreview.net/forum?id=ryUlhzWCZ
AUTHORS: Wen Sun, J. Andrew Bagnell, Byron Boots
TL;DR: Combining Imitation Learning and Reinforcement Learning to learn to outperform the expert
HIGHLIGHT: In this paper, we propose to combine imitation and reinforcement learning via the idea of reward shaping using an oracle.

4, TITLE: Model-Ensemble Trust-Region Policy Optimization
https://openreview.net/forum?id=SJJinbWRZ
AUTHORS: Thanard Kurutach, Ignasi Clavera, Yan Duan, Aviv Tamar, Pieter Abbeel
TL;DR: Deep Model-Based RL that works well.
HIGHLIGHT: In this paper, we analyze the behavior of vanilla model-based reinforcement learning methods when deep neural networks are used to learn both the model and the policy, and show that the learned policy tends to exploit regions where insufficient data is available for the model to be learned, causing instability in training.

5, TITLE: A Neural Representation of Sketch Drawings
https://openreview.net/forum?id=Hy6GHpkCW
AUTHORS: David Ha, Douglas Eck
TL;DR: We investigate alternative to traditional pixel image modelling approaches, and propose a generative model for vector images.
HIGHLIGHT: We outline a framework for conditional and unconditional sketch generation, and describe new robust training methods for generating coherent sketch drawings in a vector format.

6, TITLE: Deep Learning with Logged Bandit Feedback
https://openreview.net/forum?id=SJaP_-xAb
AUTHORS: Thorsten Joachims, Adith Swaminathan, Maarten de Rijke
TL;DR: The paper proposes a new output layer for deep networks that permits the use of logged contextual bandit feedback for training.
HIGHLIGHT: We propose a new output layer for deep neural networks that permits the use of logged contextual bandit feedback for training.

7, TITLE: Learning Latent Permutations with Gumbel-Sinkhorn Networks
https://openreview.net/forum?id=Byt3oJ-0W
AUTHORS: Gonzalo Mena, David Belanger, Scott Linderman, Jasper Snoek
TL;DR: A new method for gradient-descent inference of permutations, with applications to latent matching inference and supervised learning of permutations with neural networks
HIGHLIGHT: In response, this paper introduces a collection of new methods for end-to-end learning in such models that approximate discrete maximum-weight matching using the continuous Sinkhorn operator.

8, TITLE: Learning an Embedding Space for Transferable Robot Skills
https://openreview.net/forum?id=rk07ZXZRb
AUTHORS: Karol Hausman, Jost Tobias Springenberg, Ziyu Wang, Nicolas Heess, Martin Riedmiller
TL;DR: n/a
HIGHLIGHT: We present a method for reinforcement learning of closely related skills that are parameterized via a skill embedding space.

9, TITLE: Unsupervised Learning of Goal Spaces for Intrinsically Motivated Goal Exploration
https://openreview.net/forum?id=S1DWPP1A-
AUTHORS: Alexandre P?r?, S?bastien Forestier, Olivier Sigaud, Pierre-Yves Oudeyer
TL;DR: We propose a novel Intrinsically Motivated Goal Exploration architecture with unsupervised learning of goal space representations, and evaluate how various implementations enable the discovery of a diversity of policies.
HIGHLIGHT: In this work, we propose an approach using deep representation learning algorithms to learn an adequate goal space.

10, TITLE: Multi-View Data Generation Without View Supervision
https://openreview.net/forum?id=ryRh0bb0Z
AUTHORS: Mickael Chen, Ludovic Denoyer, Thierry Arti?res
TL;DR: We describe a novel multi-view generative model that can generate multiple views of the same object, or multiple objects in the same view with no need of label on views.
HIGHLIGHT: Therefore, we propose a generative model and a conditional variant built on such a disentangled latent space.

11, TITLE: Deep Bayesian Bandits Showdown: An Empirical Comparison of Bayesian Deep Networks for Thompson Sampling
https://openreview.net/forum?id=SyYe6k-CW
AUTHORS: Carlos Riquelme, George Tucker, Jasper Snoek
TL;DR: An Empirical Comparison of Bayesian Deep Networks for Thompson Sampling
HIGHLIGHT: Thompson Sampling and its extension to reinforcement learning provide an elegant approach to exploration that only requires access to posterior samples of the model.

12, TITLE: Semantic Interpolation in Implicit Models
https://openreview.net/forum?id=H15odZ-C-
AUTHORS: Yannic Kilcher, Aurelien Lucchi, Thomas Hofmann
TL;DR: n/a
HIGHLIGHT: As we show in this paper, care needs to be taken to match-up the distributional assumptions on code vectors with the geometry of the interpolating paths.

13, TITLE: Fidelity-Weighted Learning
https://openreview.net/forum?id=B1X0mzZCW
AUTHORS: Mostafa Dehghani, Arash Mehrjou, Stephan Gouws, Jaap Kamps, Bernhard Sch?lkopf
TL;DR: We propose Fidelity-weighted Learning, a semi-supervised teacher-student approach for training neural networks using weakly-labeled data.
HIGHLIGHT: To this end, we propose ?fidelity-weighted learning? (FWL), a semi-supervised student- teacher approach for training deep neural networks using weakly-labeled data.

14, TITLE: Latent Space Oddity: on the Curvature of Deep Generative Models
https://openreview.net/forum?id=SJzRZ-WCZ
AUTHORS: Georgios Arvanitidis, Lars Kai Hansen, S?ren Hauberg
TL;DR: n/a
HIGHLIGHT: Deep generative models provide a systematic way to learn nonlinear data distributions through a set of latent variables and a nonlinear "generator" function that maps latent points into the input space.

15, TITLE: Imitation Learning from Visual Data with Multiple Intentions
https://openreview.net/forum?id=Hk3ddfWRW
AUTHORS: Aviv Tamar, Khashayar Rohanimanesh, Yinlam Chow, Chris Vigorito, Ben Goodrich, Michael Kahane, Derik Pridmore
TL;DR: multi-modal imitation learning from unstructured demonstrations using stochastic neural network modeling intention.
HIGHLIGHT: In this paper we present an LfD approach for learning multiple modes of behavior from visual data.

16, TITLE: Hyperparameter optimization: a spectral approach
https://openreview.net/forum?id=H1zriGeCZ
AUTHORS: Elad Hazan, Adam Klivans, Yang Yuan
TL;DR: A hyperparameter tuning algorithm using discrete Fourier analysis and compressed sensing
HIGHLIGHT: We give a simple, fast algorithm for hyperparameter optimization inspired by techniques from the analysis of Boolean functions.

17, TITLE: Leveraging Grammar and Reinforcement Learning for Neural Program Synthesis
https://openreview.net/forum?id=H1Xw62kRZ
AUTHORS: Rudy Bunel, Matthew Hausknecht, Jacob Devlin, Rishabh Singh, Pushmeet Kohli
TL;DR: Using the DSL grammar and reinforcement learning to improve synthesis of programs with complex control flow.
HIGHLIGHT: For addressing thesecond limitation, we introduce a training procedure that directly maximizes theprobability of generating syntactically correct programs that fulfill the specification.We show that our contributions lead to improved accuracy of the models, especiallyin cases where the training data is limited.

18, TITLE: Efficient Sparse-Winograd Convolutional Neural Networks
https://openreview.net/forum?id=HJzgZ3JCW
AUTHORS: Xingyu Liu, Jeff Pool, Song Han, William J. Dally
TL;DR: Prune and ReLU in Winograd domain for efficient convolutional neural network
HIGHLIGHT: We propose two modifications to Winograd-based CNNs to enable these methods to exploit sparsity.

19, TITLE: Espresso: Efficient Forward Propagation for Binary Deep Neural Networks
https://openreview.net/forum?id=Sk6fD5yCb
AUTHORS: Fabrizio Pedersoli, George Tzanetakis, Andrea Tagliasacchi
TL;DR: state-of-the-art computational performance implementation of binary neural networks
HIGHLIGHT: In this paper, we show how Convolutional NeuralNetworks (CNNs) can be implemented using binaryrepresentations.

20, TITLE: Auto-Conditioned Recurrent Networks for Extended Complex Human Motion Synthesis
https://openreview.net/forum?id=r11Q2SlRW
AUTHORS: Yi Zhou, Zimo Li, Shuangjiu Xiao, Chong He, Zeng Huang, Hao Li
TL;DR: Synthesize complex and extended human motions using an auto-conditioned LSTM network
HIGHLIGHT: We present a real-time method for synthesizing highly complex human motions using a novel training regime we call the auto-conditioned Recurrent Neural Network (acRNN).

21, TITLE: Decoupling the Layers in Residual Networks
https://openreview.net/forum?id=SyMvJrdaW
AUTHORS: Ricky Fok, Aijun An, Zana Rashidi, Xiaogang Wang
TL;DR: We propose the Warped Residual Network using a parallelizable warp operator for forward and backward propagation to distant layers that trains faster than the original residual neural network.
HIGHLIGHT: We propose a Warped Residual Network (WarpNet) using a parallelizable warp operator for forward and backward propagation to distant layers that trains faster than the original residual neural network.

22, TITLE: Polar Transformer Networks
https://openreview.net/forum?id=HktRlUlAZ

AUTHORS: Carlos Esteves, Christine Allen-Blanchette, Xiaowei Zhou, Kostas Daniilidis
TL;DR: We learn feature maps invariant to translation, and equivariant to rotation and scale.
HIGHLIGHT: The ideas of PTN are extensible to 3D which we demonstrate through the Cylindrical Transformer Network.

23, TITLE: Enhancing The Reliability of Out-of-distribution Image Detection in Neural Networks
https://openreview.net/forum?id=H1VGkIxRZ
AUTHORS: Shiyu Liang, Yixuan Li, R. Srikant
TL;DR: n/a
HIGHLIGHT: We propose ODIN, a simple and effective method that does not require any change to a pre-trained neural network.

24, TITLE: Stabilizing Adversarial Nets with Prediction Methods
https://openreview.net/forum?id=Skj8Kag0Z
AUTHORS: Abhay Yadav, Sohil Shah, Zheng Xu, David Jacobs, Tom Goldstein
TL;DR: We present a simple modification to the alternating SGD method, called a prediction step, that improves the stability of adversarial networks.
HIGHLIGHT: We propose a simple modification of stochastic gradient descent that stabilizes adversarial networks.

25, TITLE: Graph Attention Networks
https://openreview.net/forum?id=rJXMpikCZ
AUTHORS: Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Li?, Yoshua Bengio
TL;DR: A novel approach to processing graph-structured data by neural networks, leveraging attention over a node's neighborhood. Achieves state-of-the-art results on transductive citation network tasks and an inductive protein-protein interaction task.
HIGHLIGHT: We present graph attention networks (GATs), novel neural network architectures that operate on graph-structured data, leveraging masked self-attentional layers to address the shortcomings of prior methods based on graph convolutions or their approximations.

26, TITLE: Minimax Curriculum Learning: Machine Teaching with Desirable Difficulties and Scheduled Diversity
https://openreview.net/forum?id=BywyFQlAW
AUTHORS: Tianyi Zhou, Jeff Bilmes
TL;DR: Minimax Curriculum Learning is a machine teaching method involving increasing desirable hardness and scheduled reducing diversity.
HIGHLIGHT: We introduce and study minimax curriculum learning (MCL), a new method for adaptively selecting a sequence of training subsets for a succession of stages in machine learning.

27, TITLE: Generalizing Hamiltonian Monte Carlo with Neural Networks
https://openreview.net/forum?id=B1n8LexRZ
AUTHORS: Daniel Levy, Matt D. Hoffman, Jascha Sohl-Dickstein
TL;DR: General method to train expressive MCMC kernels parameterized with deep neural networks. Given a target distribution p, our method provides a fast-mixing sampler, able to efficiently explore the state space.
HIGHLIGHT: We present a general-purpose method to train Markov chain Monte Carlo kernels, parameterized by deep neural networks, that converge and mix quickly to their target distribution.

28, TITLE: An Online Learning Approach to Generative Adversarial Networks
https://openreview.net/forum?id=H1Yp-j1Cb
AUTHORS: Paulina Grnarova, Kfir Y Levy, Aurelien Lucchi, Thomas Hofmann, Andreas Krause
TL;DR: n/a
HIGHLIGHT: In this paper, we view the problem of training GANs as finding a mixed strategy in a zero-sum game.

29, TITLE: Improving GANs Using Optimal Transport
https://openreview.net/forum?id=rkQkBnJAb
AUTHORS: Tim Salimans, Han Zhang, Alec Radford, Dimitris Metaxas
TL;DR: An extension of GANs combining optimal transport in primal form with an energy distance defined in an adversarially learned feature space.
HIGHLIGHT: We present Optimal Transport GAN (OT-GAN), a variant of generative adversarial nets minimizing a new metric measuring the distance between the generator distribution and the data distribution.

30, TITLE: The Kanerva Machine: A Generative Distributed Memory
https://openreview.net/forum?id=S1HlA-ZAZ
AUTHORS: Yan Wu, Greg Wayne, Alex Graves, Timothy Lillicrap
TL;DR: A generative memory model that combines slow-learning neural networks and a fast-adapting linear Gaussian model as memory.
HIGHLIGHT: We present an end-to-end trained memory system that quickly adapts to new data and generates samples like them.

31, TITLE: Mixed Precision Training
https://openreview.net/forum?id=r1gs9JgRZ
AUTHORS: Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, Hao Wu
TL;DR: n/a
HIGHLIGHT: Since this format has a narrower range than single-precision we propose three techniques for preventing the loss of critical information.

32, TITLE: Latent Constraints: Learning to Generate Conditionally from Unconditional Generative Models
https://openreview.net/forum?id=Sy8XvGb0-
AUTHORS: Jesse Engel, Matthew Hoffman, Adam Roberts
TL;DR: A new approach to conditional generation by constraining the latent space of an unconditional generative model.
HIGHLIGHT: In this paper, we develop a method to condition generation without retraining the model.

33, TITLE: MaskGAN: Better Text Generation via Filling in the _____
https://openreview.net/forum?id=ByOExmWAb
AUTHORS: William Fedus, Ian Goodfellow, Andrew M. Dai
TL;DR: Natural language GAN for filling in the blank
HIGHLIGHT: We propose to improve sample quality using Generative Adversarial Network (GANs), which explicitly train the generator to produce high quality samples and have shown a lot of success in image generation.

34, TITLE: Divide and Conquer Networks
https://openreview.net/forum?id=B1jscMbAW
AUTHORS: Alex Nowak, David Folqu?, Joan Bruna
TL;DR: Dynamic model that learns divide and conquer strategies by weak supervision.
HIGHLIGHT: We consider the learning of algorithmic tasks by mere observation of input-outputpairs.

35, TITLE: Meta-Learning and Universality: Deep Representations and Gradient Descent can Approximate any Learning Algorithm
https://openreview.net/forum?id=HyjC5yWCW
AUTHORS: Chelsea Finn, Sergey Levine
TL;DR: Deep representations combined with gradient descent can approximate any learning algorithm.
HIGHLIGHT: In this paper, we consider the meta-learning problem from the perspective of universality, formalizing the notion of learning algorithm approximation and comparing the expressive power of the aforementioned recurrent models to the more recent approaches that embed gradient descent into the meta-learner.

36, TITLE: Maximum a Posteriori Policy Optimisation
https://openreview.net/forum?id=S1ANxQW0b

AUTHORS: Abbas Abdolmaleki, Jost Tobias Springenberg, Yuval Tassa, Remi Munos, Nicolas Heess
TL;DR: n/a
HIGHLIGHT: We introduce a new algorithm for reinforcement learning called Maximum a-posteriori Policy Optimisation (MPO) based on coordinate ascent on a relative-entropy objective.

37, TITLE: META LEARNING SHARED HIERARCHIES
https://openreview.net/forum?id=SyX0IeWAW
AUTHORS: Kevin Frans, Jonathan Ho, Xi Chen, Pieter Abbeel, John Schulman
TL;DR: learn hierarchal sub-policies through end-to-end training over a distribution of tasks
HIGHLIGHT: We develop a metalearning approach for learning hierarchically structured poli- cies, improving sample efficiency on unseen tasks through the use of shared primitives?policies that are executed for large numbers of timesteps.

38, TITLE: Deep Neural Networks as Gaussian Processes
https://openreview.net/forum?id=B1EA-M-0Z
AUTHORS: Jaehoon Lee, Yasaman Bahri, Roman Novak, Samuel S. Schoenholz, Jeffrey Pennington, Jascha Sohl-Dickstein
TL;DR: We show how to make predictions using deep networks, without training deep networks.
HIGHLIGHT: We observe that the trained neural network accuracy approaches that of the corresponding GP with increasing layer width, and that the GP uncertainty is strongly correlated with trained network prediction error.

39, TITLE: Syntax-Directed Variational Autoencoder for Structured Data
https://openreview.net/forum?id=SyqShMZRb
AUTHORS: Hanjun Dai, Yingtao Tian, Bo Dai, Steven Skiena, Le Song
TL;DR: A new generative model for discrete structured data. The proposed stochastic lazy attribute converts the offline semantic check into online guidance for stochastic decoding, which effectively addresses the constraints in syntax and semantics, and also achieves superior performance
HIGHLIGHT: Inspired by the theory of compiler where syntax and semantics check is done via syntax-directed translation (SDT), we propose a novel syntax-directed variational autoencoder (SD-VAE) by introducing stochastic lazy attributes.

40, TITLE: Neural-Guided Deductive Search for Real-Time Program Synthesis from Examples
https://openreview.net/forum?id=rywDjg-RW
AUTHORS: Ashwin Kalyan, Abhishek Mohta, Oleksandr Polozov, Dhruv Batra, Prateek Jain, Sumit Gulwani
TL;DR: We integrate symbolic (deductive) and statistical (neural-based) methods to enable real-time program synthesis with almost perfect generalization from 1 input-output example.
HIGHLIGHT: In this work,we propose Neural Guided Deductive Search (NGDS), a hybrid synthesis techniquethat combines the best of both symbolic logic techniques and statistical models.Thus, it produces programs that satisfy the provided specifications by constructionand generalize well on unseen examples, similar to data-driven systems.

41, TITLE: Evidence Aggregation for Answer Re-Ranking in Open-Domain Question Answering
https://openreview.net/forum?id=rJl3yM-Ab
AUTHORS: Shuohang Wang, Mo Yu, Jing Jiang, Wei Zhang, Xiaoxiao Guo, Shiyu Chang, Zhiguo Wang, Tim Klinger, Gerald Tesauro, Murray Campbell
TL;DR: We propose a method that can make use of the multiple passages information for open-domain QA.
HIGHLIGHT: In this paper, we deal with this problem as answer re-ranking.

42, TITLE: WRPN: Wide Reduced-Precision Networks
https://openreview.net/forum?id=B1ZvaaeAZ
AUTHORS: Asit Mishra, Eriko Nurvitadhi, Jeffrey J Cook, Debbie Marr
TL;DR: Lowering precision (to 4-bits, 2-bits and even binary) and widening the filter banks gives as accurate network as those obtained with FP32 weights and activations.
HIGHLIGHT: For computer vision applications, prior works have shown the efficacy of reducing numeric precision of model parameters (network weights) in deep neural networks.

43, TITLE: MGAN: Training Generative Adversarial Nets with Multiple Generators
https://openreview.net/forum?id=rkmu5b0a-
AUTHORS: Quan Hoang, Tu Dinh Nguyen, Trung Le, Dinh Phung
TL;DR: We propose a new approach to train GANs with a mixture of generators to overcome the mode collapsing problem.
HIGHLIGHT: We propose in this paper a new approach to train the Generative Adversarial Nets (GANs) with a mixture of generators to overcome the mode collapsing problem.

44, TITLE: The Reactor: A fast and sample-efficient Actor-Critic agent for Reinforcement Learning
https://openreview.net/forum?id=rkHVZWZAZ
AUTHORS: Audrunas Gruslys, Will Dabney, Mohammad Gheshlaghi Azar, Bilal Piot, Marc Bellemare, Remi Munos
TL;DR: Reactor combines multiple algorithmic and architectural contributions to produce an agent with higher sample-efficiency than Prioritized Dueling DQN while giving better run-time performance than A3C.
HIGHLIGHT: In this work we present a new agent architecture, called Reactor, which combines multiple algorithmic and architectural contributions to produce an agent with higher sample-efficiency than Prioritized Dueling DQN (Wang et al., 2016) and Categorical DQN (Bellemare et al., 2017), while giving better run-time performance than A3C (Mnih et al., 2016).

45, TITLE: SEARNN: Training RNNs with global-local losses
https://openreview.net/forum?id=HkUR_y-RZ
AUTHORS: R?mi Leblond, Jean-Baptiste Alayrac, Anton Osokin, Simon Lacoste-Julien
TL;DR: We introduce SeaRNN, a novel algorithm for RNN training, inspired by the learning to search approach to structured prediction, in order to avoid the limitations of MLE training.
HIGHLIGHT: We propose SEARNN, a novel training algorithm for recurrent neural networks (RNNs) inspired by the "learning to search" (L2S) approach to structured prediction.

46, TITLE: Distributed Distributional Deterministic Policy Gradients
https://openreview.net/forum?id=SyZipzbCb
AUTHORS: Gabriel Barth-Maron, Matthew W. Hoffman, David Budden, Will Dabney, Dan Horgan, Dhruva TB, Alistair Muldal, Nicolas Heess, Timothy Lillicrap
TL;DR: We develop an agent that we call the Distributional Deterministic Deep Policy Gradient algorithm, which achieves state of the art performance on a number of challenging continuous control problems.
HIGHLIGHT: We combine this within a distributed framework for off-policy learning in order to develop what we call the Distributed Distributional Deep Deterministic Policy Gradient algorithm, D4PG.

47, TITLE: Hierarchical Subtask Discovery with Non-Negative Matrix Factorization
https://openreview.net/forum?id=ry80wMW0W
AUTHORS: Adam C. Earle, Andrew M. Saxe, Benjamin Rosman
TL;DR: We present a novel algorithm for hierarchical subtask discovery which leverages the multitask linear Markov decision process framework.
HIGHLIGHT: We present a novel algorithm for subtask discovery, based on the recently introduced multitask linearly-solvable Markov decision process (MLMDP) framework.

48, TITLE: Parametrized Hierarchical Procedures for Neural Programming
https://openreview.net/forum?id=rJl63fZRb
AUTHORS: Roy Fox, Richard Shin, Sanjay Krishnan, Ken Goldberg, Dawn Song, Ion Stoica
TL;DR: We introduce the PHP model for hierarchical representation of neural programs, and an algorithm for learning PHPs from a mixture of strong and weak supervision.
HIGHLIGHT: To address these challenges, we propose to model programs as Parametrized Hierarchical Procedures (PHPs).

49, TITLE: Viterbi-based Pruning for Sparse Matrix with Fixed and High Index Compression Ratio
https://openreview.net/forum?id=S1D8MPxA-
AUTHORS: Dongsoo Lee, Daehyun Ahn, Taesu Kim, Pierce I. Chuang, Jae-Joon Kim

TL;DR: We present a new pruning method and sparse matrix format to enable high index compression ratio and parallel index decoding process.
HIGHLIGHT: In this paper, a new class of sparse matrix representation utilizing Viterbi algorithm that has a high, and more importantly, fixed index compression ratio regardless of the pruning rate, is proposed.

50, TITLE: cGANs with Projection Discriminator
https://openreview.net/forum?id=ByS1VpgRZ
AUTHORS: Takeru Miyato, Masanori Koyama
TL;DR: We propose a novel, projection based way to incorporate the conditional information into the discriminator of GANs that respects the role of the conditional information in the underlining probabilistic model.
HIGHLIGHT: We propose a novel, projection based way to incorporate the conditional information into the discriminator of GANs that respects the role of the conditional information in the underlining probabilistic model.This approach is in contrast with most frameworks of conditional GANs used in application today, which use the conditional information by concatenating the (embedded) conditional vector to the feature vectors.With this modification, we were able to significantly improve the quality of the class conditional image generation on ILSVRC2012 (ImageNet) dataset from the current state-of-the-art result, and we achieved this with a single pair of a discriminator and a generator.We were also able to extend the application to super-resolution and succeeded in producing highly discriminative super-resolution images.This new structure also enabled high quality category transformation based on parametric functional transformation of conditional batch normalization layers in the generator.

51, TITLE: Unsupervised Representation Learning by Predicting Image Rotations
https://openreview.net/forum?id=S1v4N2l0-
AUTHORS: Spyros Gidaris, Praveer Singh, Nikos Komodakis
TL;DR: n/a
HIGHLIGHT: In our work we propose to learn image features by training ConvNets to recognize the 2d rotation that is applied to the image that it gets as input.

52, TITLE: Emergent Communication in a Multi-Modal, Multi-Step Referential Game
https://openreview.net/forum?id=rJGZq6g0-
AUTHORS: Katrina Evtimova, Andrew Drozdov, Douwe Kiela, Kyunghyun Cho
TL;DR: n/a
HIGHLIGHT: Inspired by previous work on emergent communication in referential games, we propose a novel multi-modal, multi-step referential game, where the sender and receiver have access to distinct modalities of an object, and their information exchange is bidirectional and of arbitrary duration.

53, TITLE: FastGCN: Fast Learning with Graph Convolutional Networks via Importance Sampling
https://openreview.net/forum?id=rytstxWAW
AUTHORS: Jie Chen, Tengfei Ma, Cao Xiao
TL;DR: n/a
HIGHLIGHT: Such an interpretation allows for the use of Monte Carlo approaches to consistently estimate the integrals, which in turn leads to a batched training scheme as we propose in this work---FastGCN.

54, TITLE: Emergent Translation in Multi-Agent Communication
https://openreview.net/forum?id=H1vEXaxA-
AUTHORS: Jason Lee, Kyunghyun Cho, Jason Weston, Douwe Kiela
TL;DR: n/a
HIGHLIGHT: In this work, we propose a communication game where two agents, native speakers of their own respective languages, jointly learn to solve a visual referential task.

55, TITLE: An efficient framework for learning sentence representations
https://openreview.net/forum?id=rJvJXZb0W
AUTHORS: Lajanugen Logeswaran, Honglak Lee
TL;DR: A framework for learning high-quality sentence representations efficiently.
HIGHLIGHT: In this work we propose a simple and efficient framework for learning sentence representations from unlabelled data.

56, TITLE: NerveNet: Learning Structured Policy with Graph Neural Networks
https://openreview.net/forum?id=S1sqHMZCb
AUTHORS: Tingwu Wang, Renjie Liao, Jimmy Ba, Sanja Fidler
TL;DR: using graph neural network to model structural information of the agents to improve policy and transferability
HIGHLIGHT: In this work, we propose NerveNet to explicitly model the structure of an agent, which naturally takes the form of a graph.

57, TITLE: Learning Latent Representations in Neural Networks for Clustering through Pseudo Supervision and Graph-based Activity Regularization
https://openreview.net/forum?id=HkMvEOlAb
AUTHORS: Ozsel Kilinc, Ismail Uysal
TL;DR: n/a
HIGHLIGHT: In this paper, we propose a novel unsupervised clustering approach exploiting the hidden information that is indirectly introduced through a pseudo classification objective.

58, TITLE: Adversarial Dropout Regularization
https://openreview.net/forum?id=HJIoJWZCZ
AUTHORS: Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, Kate Saenko
TL;DR: We present a new adversarial method for adapting neural representations based on a critic that detects non-discriminative features.
HIGHLIGHT: We present a domain adaptation method for transferring neural representations from label-rich source domains to unlabeled target domains.

59, TITLE: Demystifying MMD GANs
https://openreview.net/forum?id=r1lUOzWCW
AUTHORS: Mikolaj Binkowski, Dougal J. Sutherland, Michael Arbel, Arthur Gretton
TL;DR: Explain bias situation with MMD GANs; MMD GANs work with smaller critic networks than WGAN-GPs; new GAN evaluation metric.
HIGHLIGHT: As our main theoretical contribution, we clarify the situation with bias in GAN loss functions raised by recent work: we show that gradient estimators used in the optimization process for both MMD GANs and Wasserstein GANs are unbiased, but learning a discriminator based on samples leads to biased gradients for the generator parameters.

60, TITLE: Smooth Loss Functions for Deep Top-k Classification
https://openreview.net/forum?id=Hk5elxbRW
AUTHORS: Leonard Berrada, Andrew Zisserman, M. Pawan Kumar
TL;DR: Smooth Loss Function for Top-k Error Minimization
HIGHLIGHT: Consequently, we introduce a family of smoothed loss functions that are suited to top-k optimization via deep learning.

61, TITLE: Deep Learning as a Mixed Convex-Combinatorial Optimization Problem
https://openreview.net/forum?id=B1Lc-Gb0Z
AUTHORS: Abram L. Friesen, Pedro Domingos
TL;DR: We learn deep networks of hard-threshold units by setting hidden-unit targets using combinatorial optimization and weights by convex optimization, resulting in improved performance on ImageNet.
HIGHLIGHT: We address this problem by observing that setting targets for hard-threshold hidden units in order to minimize loss is a discrete optimization problem, and can be solved as such.

62, TITLE: Learning Approximate Inference Networks for Structured Prediction
https://openreview.net/forum?id=H1WgVz-AZ
AUTHORS: Lifu Tu, Kevin Gimpel
TL;DR: n/a
HIGHLIGHT: This?inference network? outputs continuous values that we treat as the output structure.

63, TITLE: LEARNING TO SHARE: SIMULTANEOUS PARAMETER TYING AND SPARSIFICATION IN DEEP LEARNING
https://openreview.net/forum?id=rypT3fb0b
AUTHORS: Dejiao Zhang, Haozhu Wang, Mario Figueiredo, Laura Balzano
TL;DR: We have proposed using the recent GrOWL regularizer for simultaneous parameter sparsity and tying in DNN learning.
HIGHLIGHT: In this paper, we adopt a recently proposed sparsity-inducing regularizer, named GrOWL (group ordered weighted l1), which encourages sparsity and, simulta- neously, learns which groups of parameters should share a common value.

64, TITLE: Model compression via distillation and quantization
https://openreview.net/forum?id=S1XolQbRW
AUTHORS: Antonio Polino, Razvan Pascanu, Dan Alistarh
TL;DR: Obtains state-of-the-art accuracy for quantized, shallow nets by leveraging distillation.
HIGHLIGHT: This paper focuses on this problem, and proposes two new compression methods, which jointly leverage weight quantization and distillation of larger teacher networks into smaller student networks.

65, TITLE: Variational Message Passing with Structured Inference Networks
https://openreview.net/forum?id=HyH9lbZAW
AUTHORS: Wu Lin, Nicolas Hubacher, Mohammad Emtiyaz Khan
TL;DR: We propose a variational message-passing algorithm for models that contain both the deep model and probabilistic graphical model.
HIGHLIGHT: We propose a variational message-passing algorithm for variational inference in such models.

66, TITLE: Action-dependent Control Variates for Policy Optimization via Stein Identity
https://openreview.net/forum?id=H1mCp-ZRZ
AUTHORS: Hao Liu*, Yihao Feng*, Yi Mao, Dengyong Zhou, Jian Peng, Qiang Liu
TL;DR: n/a
HIGHLIGHT: In this work, we propose a control variate method to effectively reduce variance for policy gradient methods.

67, TITLE: Variational image compression with a scale hyperprior
https://openreview.net/forum?id=rkcQFMZRb
AUTHORS: Johannes Ball?, David Minnen, Saurabh Singh, Sung Jin Hwang, Nick Johnston
TL;DR: n/a
HIGHLIGHT: We describe an end-to-end trainable model for image compression based on variational autoencoders.

68, TITLE: Variational Inference of Disentangled Latent Concepts from Unlabeled Observations
https://openreview.net/forum?id=H1kG7GZAW
AUTHORS: Abhishek Kumar, Prasanna Sattigeri, Avinash Balakrishnan
TL;DR: We propose a variational inference based approach for encouraging the inference of disentangled latents. We also propose a new metric for quantifying disentanglement.
HIGHLIGHT: We consider the problem of unsupervised learning of disentangled representations from large pool of unlabeled observations, and propose a variational inference based approach to infer disentangled latent factors.

69, TITLE: Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches
https://openreview.net/forum?id=rJNpifWAb
AUTHORS: Yeming Wen, Paul Vicol, Jimmy Ba, Dustin Tran, Roger Grosse
TL;DR: We introduce flipout, an efficient method for decorrelating the gradients computed by stochastic neural net weights within a mini-batch by implicitly sampling pseudo-independent weight perturbations for each example.
HIGHLIGHT: We introduce flipout, an efficient method for decorrelating the gradients within a mini-batch by implicitly sampling pseudo-independent weight perturbations for each example.

70, TITLE: Kernel Implicit Variational Inference
https://openreview.net/forum?id=r1l4eQW0Z
AUTHORS: Jiaxin Shi, Shengyang Sun, Jun Zhu
TL;DR: n/a
HIGHLIGHT: In this paper, we present a new approach named Kernel Implicit Variational Inference that addresses these challenges.

71, TITLE: A Scalable Laplace Approximation for Neural Networks
https://openreview.net/forum?id=Skdvd2xAZ
AUTHORS: Hippolyt Ritter, Aleksandar Botev, David Barber
TL;DR: We construct a Kronecker factored Laplace approximation for neural networks that leads to an efficient matrix normal distribution over the weights.
HIGHLIGHT: We extensively compare our method to using Dropout and a diagonal Laplace approximation for estimating the uncertainty of a network.

72, TITLE: The High-Dimensional Geometry of Binary Neural Networks
https://openreview.net/forum?id=B1IDRdeCW
AUTHORS: Alexander G. Anderson, Cory P. Berg
TL;DR: Recent successes of Binary Neural Networks can be understood based on the geometry of high-dimensional binary vectors
HIGHLIGHT: Recent research has shown that one can train a neural network with binary weights and activations at train time by augmenting the weights with a high-precision continuous latent variable that accumulates small changes from stochastic gradient descent.

73, TITLE: Apprentice: Using Knowledge Distillation Techniques To Improve Low-Precision Network Accuracy
https://openreview.net/forum?id=B1ae1lZRb
AUTHORS: Asit Mishra, Debbie Marr
TL;DR: We show that knowledge transfer techniques can improve the accuracy of low precision networks and set new state-of-the-art accuracy for ternary and 4-bits precision.
HIGHLIGHT: In this paper, we study the combination of these two techniques and show that the performance of low precision networks can be significantly improved by using knowledge distillation techniques.

74, TITLE: Distributed Prioritized Experience Replay
https://openreview.net/forum?id=H1Dy---0Z
AUTHORS: Dan Horgan, John Quan, David Budden, Gabriel Barth-Maron, Matteo Hessel, Hado van Hasselt, David Silver
TL;DR: A distributed architecture for deep reinforcement learning at scale, using parallel data-generation to improve the state of the art on the Arcade Learning Environment benchmark in a fraction of the wall-clock training time of previous approaches.
HIGHLIGHT: We propose a distributed architecture for deep reinforcement learning at scale, that enables agents to learn effectively from orders of magnitude more data than previously possible.

75, TITLE: Learning from Between-class Examples for Deep Sound Recognition
https://openreview.net/forum?id=B1Gi6LeRZ
AUTHORS: Yuji Tokozume, Yoshitaka Ushiku, Tatsuya Harada
TL;DR: We propose an novel learning method for deep sound recognition named BC learning.
HIGHLIGHT: We propose a novel learning method for deep sound recognition: Between-Class learning (BC learning).

76, TITLE: Training Confidence-calibrated Classifiers for Detecting Out-of-Distribution Samples
https://openreview.net/forum?id=ryiAv2xAZ
AUTHORS: Kimin Lee, Honglak Lee, Kibok Lee, Jinwoo Shin
TL;DR: n/a
HIGHLIGHT: In this paper, we develop a novel training method for classifiers so that such inference algorithms can work better.

77, TITLE: VoiceLoop: Voice Fitting and Synthesis via a Phonological Loop
https://openreview.net/forum?id=SkFAWax0-
AUTHORS: Yaniv Taigman, Lior Wolf, Adam Polyak, Eliya Nachmani
TL;DR: n/a
HIGHLIGHT: We present a new neural text to speech (TTS) method that is able to transform text to speech in voices that are sampled in the wild.

78, TITLE: Large scale distributed neural network training through online distillation
https://openreview.net/forum?id=rkr1UDeC-
AUTHORS: Rohan Anil, Gabriel Pereyra, Alexandre Passos, Robert Ormandi, George E. Dahl, Geoffrey E. Hinton
TL;DR: We perform large scale experiments to show that a simple online variant of distillation can help us scale distributed neural network training to more machines.
HIGHLIGHT: In this paper we explore a variant of distillation which is relatively straightforward to use as it does not require a complicated multi-stage setup or many new hyperparameters.

79, TITLE: Learning Differentially Private Recurrent Language Models
https://openreview.net/forum?id=BJ0hF1Z0b
AUTHORS: H. Brendan McMahan, Daniel Ramage, Kunal Talwar, Li Zhang
TL;DR: User-level differential privacy for recurrent neural network language models is possible with a sufficiently large dataset.
HIGHLIGHT: We demonstrate that it is possible to train large recurrent language models with user-level differential privacy guarantees with only a negligible cost in predictive accuracy.

80, TITLE: Mastering the Dungeon: Grounded Language Learning by Mechanical Turker Descent
https://openreview.net/forum?id=SJ-C6JbRW
AUTHORS: Zhilin Yang, Saizheng Zhang, Jack Urbanek, Will Feng, Alexander Miller, Arthur Szlam, Douwe Kiela, Jason Weston
TL;DR: n/a
HIGHLIGHT: In this work we propose an interactive learning procedure called Mechanical Turker Descent (MTD) that trains agents to execute natural language commands grounded in a fantasy text adventure game.

81, TITLE: Generating Wikipedia by Summarizing Long Sequences
https://openreview.net/forum?id=Hyg0vbWC-
AUTHORS: Peter J. Liu*, Mohammad Saleh*, Etienne Pot, Ben Goodrich, Ryan Sepassi, Lukasz Kaiser, Noam Shazeer
TL;DR: We generate Wikipedia articles abstractively conditioned on source document text.
HIGHLIGHT: For the abstractive model, we introduce a decoder-only architecturethat can scalably attend to very long sequences, much longer than typical encoder-decoder architectures used in sequence transduction.

82, TITLE: Unsupervised Machine Translation Using Monolingual Corpora Only
https://openreview.net/forum?id=rkYTTf-AZ
AUTHORS: Guillaume Lample, Alexis Conneau, Ludovic Denoyer, Marc'Aurelio Ranzato
TL;DR: We propose a new unsupervised machine translation model that can learn without using parallel corpora; experimental results show impressive performance on multiple corpora and pairs of languages.
HIGHLIGHT: In this work, we take this research direction to the extreme and investigate whether it is possible to learn to translate even without any parallel data.

83, TITLE: A Deep Reinforced Model for Abstractive Summarization
https://openreview.net/forum?id=HkAClQgA-
AUTHORS: Romain Paulus, Caiming Xiong, Richard Socher
TL;DR: A summarization model combining a new intra-attention and reinforcement learning method to increase summary ROUGE scores and quality for long sequences.

HIGHLIGHT: We introduce a neural network model with a novel intra-attention that attends over the input and continuously generated output separately, and a new training method that combines standard supervised word prediction and reinforcement learning (RL).

84, TITLE: Compressing Word Embeddings via Deep Compositional Code Learning
https://openreview.net/forum?id=BJRZzFlRb
AUTHORS: Raphael Shu, Hideki Nakayama
TL;DR: Compressing the word embeddings over 94% without hurting the performance.
HIGHLIGHT: We propose to directly learn the discrete codes in an end-to-end neural network by applying the Gumbel-softmax trick.

85, TITLE: Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training
https://openreview.net/forum?id=SkhQHMW0W
AUTHORS: Yujun Lin, Song Han, Huizi Mao, Yu Wang, Bill Dally
TL;DR: we find 99.9% of the gradient exchange in distributed SGD is redundant; we reduce the communication bandwidth by two orders of magnitude without losing accuracy.
HIGHLIGHT: In this paper, we find 99.9% of the gradient exchange in distributed SGD is redundant, and propose Deep Gradient Compression (DGC) to greatly reduce the communication bandwidth.

86, TITLE: QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension
https://openreview.net/forum?id=B14TlG-RW
AUTHORS: Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, Quoc V. Le
TL;DR: A simple architecture consisting of convolutions and attention achieves results on par with the best documented recurrent models.
HIGHLIGHT: We propose a new Q\&A architecture called QANet, which does not require recurrent networks: Its encoder consists exclusively of convolution and self-attention, where convolution models local interactions and self-attention models global interactions.

87, TITLE: Unsupervised Neural Machine Translation
https://openreview.net/forum?id=Sy2ogebAW
AUTHORS: Mikel Artetxe, Gorka Labaka, Eneko Agirre, Kyunghyun Cho
TL;DR: We introduce the first successful method to train neural machine translation in an unsupervised manner, using nothing but monolingual corpora
HIGHLIGHT: In this work, we completely remove the need of parallel data and propose a novel method to train an NMT system in a completely unsupervised manner, relying on nothing but monolingual corpora.

88, TITLE: Learning One-hidden-layer Neural Networks with Landscape Design
https://openreview.net/forum?id=BkwHObbRZ
AUTHORS: Rong Ge, Jason D. Lee, Tengyu Ma
TL;DR: The paper analyzes the optimization landscape of one-hidden-layer neural nets and designs a new objective that provably has no spurious local minimum.
HIGHLIGHT: We consider the problem of learning a one-hidden-layer neural network: we assume the input x is from Gaussian distribution and the label $y = a \sigma(Bx) + \xi$, where a is a nonnegative vector and B is a full-rank weight matrix, and ξ is a noise vector.

89, TITLE: Critical Points of Linear Neural Networks: Analytical Forms and Landscape Properties
https://openreview.net/forum?id=SysEexbRb
AUTHORS: Yi Zhou, Yingbin Liang
TL;DR: We provide necessary and sufficient analytical forms for the critical points of the square loss functions for various neural networks, and exploit the analytical forms to characterize the landscape properties for the loss functions of these neural networks.
HIGHLIGHT: In this paper, we provide a necessary and sufficient characterization of the analytical forms for the critical points (as well as global minimizers) of the square loss functions for linear neural networks.

90, TITLE: Learning Parametric Closed-Loop Policies for Markov Potential Games
https://openreview.net/forum?id=rJm7VfZA-
AUTHORS: Sergio Valcarcel Macua, Javier Zazo, Santiago Zazo
TL;DR: We present general closed loop analysis for Markov potential games and show that deep reinforcement learning can be used for learning approximate closed-loop Nash equilibrium.
HIGHLIGHT: We present a closed-loop (CL) analysis for MPGs and consider parametric policies that depend on the current state and where agents adapt to stochastic transitions.

91, TITLE: The power of deeper networks for expressing natural functions
https://openreview.net/forum?id=SyProzZAW
AUTHORS: David Rolnick, Max Tegmark
TL;DR: We prove that deep neural networks are exponentially more efficient than shallow ones at approximating sparse multivariate polynomials.
HIGHLIGHT: We shed light on this by proving that the total number of neurons m required to approximate natural classes of multivariate polynomials of n variables grows only linearly with n for deep neural networks, but grows exponentially when merely a single hidden layer is allowed.

92, TITLE: Empirical Risk Landscape Analysis for Understanding Deep Neural Networks
https://openreview.net/forum?id=B1QgVti6Z
AUTHORS: Pan Zhou, Jiashi Feng
TL;DR: n/a
HIGHLIGHT: This work aims to provide comprehensive landscape analysis of empirical risk in deep neural networks (DNNs), including the convergence behavior of its gradient, its stationary points and the empirical risk itself to their corresponding population counterparts, which reveals how various network parameters determine the convergence performance.

93, TITLE: On the Discrimination-Generalization Tradeoff in GANs
https://openreview.net/forum?id=Hk9Xc_lR-
AUTHORS: Pengchuan Zhang, Qiang Liu, Dengyong Zhou, Tao Xu, Xiaodong He
TL;DR: This paper studies the discrimination and generalization properties of GANs when the discriminator set is a restricted function class like neural networks.
HIGHLIGHT: In this paper, we show that a discriminator set is guaranteed to be discriminative whenever its linear span is dense in the set of bounded continuous functions.

94, TITLE: Decision-Based Adversarial Attacks: Reliable Attacks Against Black-Box Machine Learning Models
https://openreview.net/forum?id=SyZI0GWCZ
AUTHORS: Wieland Brendel *, Jonas Rauber *, Matthias Bethge
TL;DR: A novel adversarial attack that can directly attack real-world black-box machine learning models without transfer.
HIGHLIGHT: Here we introduce the Boundary Attack, a decision-based attack that starts from a large adversarial perturbation and then seeks to reduce the perturbation while staying adversarial.

95, TITLE: Unbiased Online Recurrent Optimization
https://openreview.net/forum?id=rJQDjk-0b
AUTHORS: Corentin Tallec, Yann Ollivier
TL;DR: Introduces an online, unbiased and easily implementable gradient estimate for recurrent models.
HIGHLIGHT: Unbiased Online Recurrent Optimization

96, TITLE: Measuring the Intrinsic Dimension of Objective Landscapes
https://openreview.net/forum?id=ryup8-WCW
AUTHORS: Chunyuan Li, Heerad Farkhoor, Rosanne Liu, Jason Yosinski
TL;DR: We train in random subspaces of parameter space to measure how many dimensions are really needed to find a solution.
HIGHLIGHT: In this paper we attempt to answer this question by training networks not in their native parameter space, but instead in a smaller, randomly oriented subspace.

97, TITLE: Memorization Precedes Generation: Learning Unsupervised GANs with Memory Networks
https://openreview.net/forum?id=rkO3uTkAZ
AUTHORS: Youngjin Kim, Minjung Kim, Gunhee Kim
TL;DR: n/a
HIGHLIGHT: We propose an approach to address two issues that commonly occur during training of unsupervised GANs.

98, TITLE: Stochastic Activation Pruning for Robust Adversarial Defense
https://openreview.net/forum?id=H1uR4GZRZ
AUTHORS: Guneet S. Dhillon, Kamyar Azizzadenesheli, Zachary C. Lipton, Jeremy D. Bernstein, Jean Kossaifi, Aran Khanna, Animashree Anandkumar
TL;DR: n/a
HIGHLIGHT: In this light, we propose Stochastic Activation Pruning (SAP), a mixed strategy for adversarial defense.

99, TITLE: Sparse Persistent RNNs: Squeezing Large Recurrent Networks On-Chip
https://openreview.net/forum?id=HkxF5RgC-
AUTHORS: Feiwen Zhu, Jeff Pool, Michael Andersch, Jeremy Appleyard, Fung Xie
TL;DR: Combining network pruning and persistent kernels into a practical, fast, and accurate network implementation.
HIGHLIGHT: With these optimizations, we achieve speedups of over 6x over the next best algorithm for a hidden layer of size 2304, batch size of 4, and a density of 30%.

100, TITLE: GANITE: Estimation of Individualized Treatment Effects using Generative Adversarial Nets
https://openreview.net/forum?id=ByKWUeWA-
AUTHORS: Jinsung Yoon, James Jordon, Mihaela van der Schaar
TL;DR: n/a
HIGHLIGHT: We propose a novel method for inferring ITE based on the Generative Adversarial Nets (GANs) framework.

101, TITLE: Thermometer Encoding: One Hot Way To Resist Adversarial Examples
https://openreview.net/forum?id=S18Su--CW
AUTHORS: Jacob Buckman, Aurko Roy, Colin Raffel, Ian Goodfellow
TL;DR: Input discretization leads to robustness against adversarial examples
HIGHLIGHT: We propose a simplemodification to standard neural network architectures, thermometerencoding, which significantly increases the robustness of the network toadversarial examples.

102, TITLE: Trust-PCL: An Off-Policy Trust Region Method for Continuous Control
https://openreview.net/forum?id=HyrCWeWCb
AUTHORS: Ofir Nachum, Mohammad Norouzi, Kelvin Xu, Dale Schuurmans
TL;DR: We extend recent insights related to softmax consistency to achieve state-of-the-art results in continuous control.
HIGHLIGHT: To address this problem, we propose an off-policy trust region method, Trust-PCL, which exploits an observation that the optimal policy and state values of a maximum reward objective with a relative-entropy regularizer satisfy a set of multi-step pathwise consistencies along any path.

103, TITLE: Stochastic Variational Video Prediction
https://openreview.net/forum?id=rk49Mg-CW
AUTHORS: Mohammad Babaeizadeh, Chelsea Finn, Dumitru Erhan, Roy H. Campbell, Sergey Levine
TL;DR: Stochastic variational video prediction in real-world settings.
HIGHLIGHT: In this paper, we develop a stochastic variational video prediction (SV2P) method that predicts a different possible future for each sample of its latent variables.

104, TITLE: Towards Image Understanding from Deep Compression Without Decoding
https://openreview.net/forum?id=HkXWCMbRW
AUTHORS: Robert Torfason, Fabian Mentzer, Eirikur Agustsson, Michael Tschannen, Radu Timofte, Luc Van Gool
TL;DR: n/a
HIGHLIGHT: Motivated by recent work on deep neural network (DNN)-based image compression methods showing potential improvements in image quality, savings in storage, and bandwidth reduction, we propose to perform image understanding tasks such as classification and segmentation directly on the compressed representations produced by these compression methods.

105, TITLE: Automatically Inferring Data Quality for Spatiotemporal Forecasting
https://openreview.net/forum?id=ByJIWUnpW
AUTHORS: Sungyong Seo, Arash Mohegh, George Ban-Weiss, Yan Liu
TL;DR: We propose a method that infers the time-varying data quality level for spatiotemporal forecasting without explicitly assigned labels.
HIGHLIGHT: In this paper, we propose a novel solution that can automatically infer data quality levels of different sources through local variations of spatiotemporal signals without explicit labels.

106, TITLE: Towards better understanding of gradient-based attribution methods for Deep Neural Networks
https://openreview.net/forum?id=Sy21R9JAW
AUTHORS: Marco Ancona, Enea Ceolini, Cengiz ?ztireli, Markus Gross
TL;DR: Four existing backpropagation-based attribution methods are fundamentally similar. How to assess it?
HIGHLIGHT: In this work we analyze four gradient-based attribution methods and formally prove conditions of equivalence and approximation between them.

107, TITLE: Countering Adversarial Images using Input Transformations
https://openreview.net/forum?id=SyJ7ClWCb
AUTHORS: Chuan Guo, Mayank Rana, Moustapha Cisse, Laurens van der Maaten
TL;DR: We apply a model-agnostic defense strategy against adversarial examples and achieve 60% white-box accuracy and 90% black-box accuracy against major attack algorithms.
HIGHLIGHT: This paper investigates strategies that defend against adversarial-example attacks on image-classification systems by transforming the inputs before feeding them to the system.

108, TITLE: Skip RNN: Learning to Skip State Updates in Recurrent Neural Networks
https://openreview.net/forum?id=HkwVAXyCW
AUTHORS: V?ctor Campos, Brendan Jou, Xavier Gir?-i-Nieto, Jordi Torres, Shih-Fu Chang
TL;DR: A modification for existing RNN architectures which allows them to skip state updates while preserving the performance of the original architectures.
HIGHLIGHT: We introduce the Skip RNN model which extends existing RNN models by learning to skip state updates and shortens the effective size of the computational graph.

109, TITLE: Modular Continual Learning in a Unified Visual Environment
https://openreview.net/forum?id=rkPLzgZAZ
AUTHORS: Kevin T. Feigelis, Blue Sheffer, Daniel L. K. Yamins
TL;DR: We propose a neural module approach to continual learning using a unified visual environment with a large action space.
HIGHLIGHT: Here, we describe a modular continual reinforcement learning paradigm inspired by these abilities.

110, TITLE: Twin Networks: Matching the Future for Sequence Generation
https://openreview.net/forum?id=BydLzGb0Z
AUTHORS: Dmitriy Serdyuk, Nan Rosemary Ke, Alessandro Sordoni, Adam Trischler, Chris Pal, Yoshua Bengio
TL;DR: The paper introduces a method of training generative recurrent networks that helps to plan ahead. We run a second RNN in a reverse direction and make a soft constraint between cotemporal forward and backward states.
HIGHLIGHT: We propose a simple technique for encouraging generative RNNs to plan ahead.

111, TITLE: Interpretable Counting for Visual Question Answering
https://openreview.net/forum?id=S1J2ZyZ0Z
AUTHORS: Alexander Trott, Caiming Xiong, Richard Socher
TL;DR: We perform counting for visual question answering; our model produces interpretable outputs by counting directly from detected objects.
HIGHLIGHT: In contrast, we treat counting as a sequential decision process and force our model to make discrete choices of what to count.

112, TITLE: Interactive Grounded Language Acquisition and Generalization in a 2D World
https://openreview.net/forum?id=H1UOm4gA-
AUTHORS: Haonan Yu, Haichao Zhang, Wei Xu
TL;DR: Training an agent in a 2D virtual world for grounded language acquisition and generalization.
HIGHLIGHT: We build a virtual agent for learning language in a 2D maze-like world.

113, TITLE: Emergent Complexity via Multi-Agent Competition
https://openreview.net/forum?id=Sy0GnUxCb
AUTHORS: Trapit Bansal, Jakub Pachocki, Szymon Sidor, Ilya Sutskever, Igor Mordatch
TL;DR: n/a
HIGHLIGHT: In this paper, we point out that a competitive multi-agent environment trained with self-play can produce behaviors that are far more complex than the environment itself.

114, TITLE: Universal Agent for Disentangling Environments and Tasks
https://openreview.net/forum?id=B1mvVm-C-
AUTHORS: Jiayuan Mao, Honghua Dong, Joseph J. Lim
TL;DR: We propose a DRL framework that disentangles task and environment specific knowledge.
HIGHLIGHT: Hence, borrowing the idea from hierarchical reinforcement learning, we propose a framework that disentangles task and environment specific knowledge by separating them into two units.

115, TITLE: Residual Connections Encourage Iterative Inference
https://openreview.net/forum?id=SJa9iHgAZ
AUTHORS: Stanislaw Jastrzebski, Devansh Arpit, Nicolas Ballas, Vikas Verma, Tong Che, Yoshua Bengio
TL;DR: Residual connections really perform iterative inference
HIGHLIGHT: We attempt to further expose properties of this aspect.

116, TITLE: Emergent Communication through Negotiation
https://openreview.net/forum?id=Hk6WhagRW
AUTHORS: Kris Cao, Angeliki Lazaridou, Marc Lanctot, Joel Z Leibo, Karl Tuyls, Stephen Clark
TL;DR: We teach agents to negotiate using only reinforcement learning; selfish agents can do so, but only using a trustworthy communication channel, and prosocial agents can negotiate using cheap talk.
HIGHLIGHT: In this paper, we study the emergence of communication in the negotiation environment, a semi-cooperative model of agent interaction.

117, TITLE: Semi-parametric topological memory for navigation
https://openreview.net/forum?id=SygwwGbRW
AUTHORS: Nikolay Savinov, Alexey Dosovitskiy, Vladlen Koltun
TL;DR: We introduce a new memory architecture for navigation in previously unseen environments, inspired by landmark-based navigation in animals.
HIGHLIGHT: We introduce a new memory architecture for navigation in previously unseen environments, inspired by landmark-based navigation in animals.

118, TITLE: Learning to Count Objects in Natural Images for Visual Question Answering
https://openreview.net/forum?id=B12Js_yRb
AUTHORS: Yan Zhang, Jonathon Hare, Adam Pr?gel-Bennett
TL;DR: Enabling Visual Question Answering models to count by handling overlapping object proposals.

HIGHLIGHT: To circumvent this problem, we propose a neural network component that allows robust counting from object proposals.

119, TITLE: i-RevNet: Deep Invertible Networks
https://openreview.net/forum?id=HJsjkMb0Z
AUTHORS: J?rn-Henrik Jacobsen, Arnold W.M. Smeulders, Edouard Oyallon
TL;DR: n/a
HIGHLIGHT: In this paper we show via a one-to-one mapping that this loss of information is not a necessary condition to learn representations that generalize well on complicated problems, such as ImageNet.

120, TITLE: Evaluating the Robustness of Neural Networks: An Extreme Value Theory Approach
https://openreview.net/forum?id=BkUHlMZ0b
AUTHORS: Tsui-Wei Weng*, Huan Zhang*, Pin-Yu Chen, Jinfeng Yi, Dong Su, Yupeng Gao, Cho-Jui Hsieh, Luca Daniel
TL;DR: We propose the first attack-independent robustness metric, a.k.a CLEVER, that can be applied to any neural network classifier.
HIGHLIGHT: In this paper, we provide theoretical justification for converting robustness analysis into a local Lipschitz constant estimation problem, and propose to use the Extreme Value Theory for efficient evaluation.

121, TITLE: HexaConv
https://openreview.net/forum?id=r1vuQG-CW
AUTHORS: Emiel Hoogeboom, Jorn W.T. Peters, Taco S. Cohen, Max Welling
TL;DR: We introduce G-HexaConv, a group equivariant convolutional neural network on hexagonal lattices.
HIGHLIGHT: In this paper we show how one can efficiently implement planar convolution and group convolution over hexagonal lattices, by re-using existing highly optimized convolution routines.

122, TITLE: Towards Deep Learning Models Resistant to Adversarial Attacks
https://openreview.net/forum?id=rJzIBfZAb
AUTHORS: Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, Adrian Vladu
TL;DR: We provide a principled, optimization-based re-look at the notion of adversarial examples, and develop methods that produce models that are adversarially robust against a wide range of adversaries.
HIGHLIGHT: Recent work has demonstrated that neural networks are vulnerable to adversarial examples, i.e., inputs that are almost indistinguishable from natural data and yet classified incorrectly by the network.

123, TITLE: Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge
https://openreview.net/forum?id=By4HsfWAZ
AUTHORS: Emmanuel de Bezenac, Arthur Pajot, Patrick Gallinari
TL;DR: n/a
HIGHLIGHT: We consider the use of Deep Learning methods for modeling complex phenomena like those occurring in natural physical processes.

124, TITLE: Communication Algorithms via Deep Learning
https://openreview.net/forum?id=ryazCMbR-
AUTHORS: Hyeji Kim, Yihan Jiang, Ranvir B. Rana, Sreeram Kannan, Sewoong Oh, Pramod Viswanath
TL;DR: We show that creatively designed and trained RNN architectures can decode well known sequential codes and achieve close to optimal performances.
HIGHLIGHT: In this paper we study whether it is possible to automate the discovery of decoding algorithms via deep learning.

125, TITLE: Simulating Action Dynamics with Neural Process Networks
https://openreview.net/forum?id=rJYFzMZC-
AUTHORS: Antoine Bosselut, Omer Levy, Ari Holtzman, Corin Ennis, Dieter Fox, Yejin Choi
TL;DR: We propose a new recurrent memory architecture that can track common sense state changes of entities by simulating the causal effects of actions.

HIGHLIGHT: In this work, we introduce Neural Process Networks to understand procedural text through (neural) simulation of action dynamics.

126, TITLE: Unsupervised Cipher Cracking Using Discrete GANs
https://openreview.net/forum?id=BkeqO7x0-
AUTHORS: Aidan N. Gomez, Sicong Huang, Ivan Zhang, Bryan M. Li, Muhammad Osama, Lukasz Kaiser
TL;DR: n/a
HIGHLIGHT: We present how CycleGAN can be made compatible with discrete data and train in a stable way.

127, TITLE: Neural Speed Reading via Skim-RNN
https://openreview.net/forum?id=Sy-dQG-Rb
AUTHORS: Minjoon Seo, Sewon Min, Ali Farhadi, Hannaneh Hajishirzi
TL;DR: n/a
HIGHLIGHT: Inspired by the principles of speed reading, we introduce Skim-RNN, a recurrent neural network (RNN) that dynamically decides to update only a small fraction of the hidden state for relatively unimportant input tokens.

128, TITLE: Multi-level Residual Networks from Dynamical Systems View
https://openreview.net/forum?id=SyJS-OgR-
AUTHORS: Bo Chang, Lili Meng, Eldad Haber, Frederick Tung, David Begert
TL;DR: n/a
HIGHLIGHT: In this paper, we adopt the dynamical systems point of view, and analyze the lesioning properties of ResNet both theoretically and experimentally.

129, TITLE: Towards Neural Phrase-based Machine Translation
https://openreview.net/forum?id=HktJec1RZ
AUTHORS: Po-Sen Huang, Chong Wang, Sitao Huang, Dengyong Zhou, Li Deng
TL;DR: Neural phrase-based machine translation with linear decoding time
HIGHLIGHT: In this paper, we present Neural Phrase-based Machine Translation (NPMT).

130, TITLE: On the State of the Art of Evaluation in Neural Language Models
https://openreview.net/forum?id=ByJHuTgA-
AUTHORS: G?bor Melis, Chris Dyer, Phil Blunsom
TL;DR: Show that LSTMs are as good or better than recent innovations for LM and that model evaluation is often unreliable.
HIGHLIGHT: We establish a new state of the art on the Penn Treebank and Wikitext-2 corpora, as well as strong baselines on the Hutter Prize dataset.

131, TITLE: Memory-based Parameter Adaptation
https://openreview.net/forum?id=rkfOvGbCW
AUTHORS: Pablo Sprechmann, Siddhant M. Jayakumar, Jack W. Rae, Alexander Pritzel, Adria Puigdomenech Badia, Benigno Uria, Oriol Vinyals, Demis Hassabis, Razvan Pascanu, Charles Blundell
TL;DR: n/a
HIGHLIGHT: We demonstrate this on a range of supervised tasks: large-scale image classification and language modelling.

132, TITLE: Initialization matters: Orthogonal Predictive State Recurrent Neural Networks
https://openreview.net/forum?id=HJJ23bW0b
AUTHORS: Krzysztof Choromanski, Carlton Downey, Byron Boots
TL;DR: Improving Predictive State Recurrent Neural Networks via Orthogonal Random Features
HIGHLIGHT: In this paper, we extend the theory of ORFs to Kernel Ridge Regression and show that ORFs can be used to obtain Orthogonal PSRNNs (OPSRNNs), which are smaller and faster than PSRNNs.

133, TITLE: PixelDefend: Leveraging Generative Models to Understand and Defend against Adversarial Examples
https://openreview.net/forum?id=rJUYGxbCW
AUTHORS: Yang Song, Taesup Kim, Sebastian Nowozin, Stefano Ermon, Nate Kushman
TL;DR: n/a
HIGHLIGHT: In this paper, we show empirically that adversarial examples mainly lie in the low probability regions of the training distribution, regardless of attack types and targeted models.

134, TITLE: Certified Defenses against Adversarial Examples
https://openreview.net/forum?id=Bys4ob-Rb
AUTHORS: Aditi Raghunathan, Jacob Steinhardt, Percy Liang
TL;DR: We demonstrate a certifiable, trainable, and scalable method for defending against adversarial examples.
HIGHLIGHT: In this work, we study this problem for neural networks with one hidden layer.

135, TITLE: Defense-GAN: Protecting Classifiers Against Adversarial Attacks Using Generative Models
https://openreview.net/forum?id=BkJ3ibb0-
AUTHORS: Pouya Samangouei, Maya Kabkab, Rama Chellappa
TL;DR: Defense-GAN uses a Generative Adversarial Network to defend against white-box and black-box attacks in classification models.
HIGHLIGHT: We propose Defense-GAN, a new framework leveraging the expressive capability of generative models to defend deep neural networks against such attacks.

136, TITLE: Ensemble Adversarial Training: Attacks and Defenses
https://openreview.net/forum?id=rkZvSe-RZ
AUTHORS: Florian Tram?r, Alexey Kurakin, Nicolas Papernot, Ian Goodfellow, Dan Boneh, Patrick McDaniel
TL;DR: Adversarial training with single-step methods overfits, and remains vulnerable to simple black-box and white-box attacks. We show that including adversarial examples from multiple sources helps defend against black-box attacks.
HIGHLIGHT: As a result, we find that adversarial training remains vulnerable to black-box attacks, where we transfer perturbations computed on undefended models, as well as to a powerful novel single-step attack that escapes the non-smooth vicinity of the input data via a small random step.We further introduce Ensemble Adversarial Training, a technique that augments training data with perturbations transferred from other models.

137, TITLE: Fraternal Dropout
https://openreview.net/forum?id=SJyVzQ-C-
AUTHORS: Konrad Zolna, Devansh Arpit, Dendi Suhubdy, Yoshua Bengio
TL;DR: We propose to train two identical copies of an recurrent neural network (that share parameters) with different dropout masks while minimizing the difference between their (pre-softmax) predictions.
HIGHLIGHT: In this paper we propose a simple technique called fraternal dropout that takes advantage of dropout to achieve this goal.

138, TITLE: Can recurrent neural networks warp time?
https://openreview.net/forum?id=SJcKhk-Ab
AUTHORS: Corentin Tallec, Yann Ollivier
TL;DR: Proves that gating mechanisms provide invariance to time transformations. Introduces and tests a new initialization for LSTMs from this insight.
HIGHLIGHT: Successful recurrent models such as long short-term memories (LSTMs) and gated recurrent units (GRUs) use \emph{ad hoc} gating mechanisms.

139, TITLE: Parallelizing Linear Recurrent Neural Nets Over Sequence Length
https://openreview.net/forum?id=HyUNwulC-
AUTHORS: Eric Martin, Chris Cundy
TL;DR: use parallel scan to parallelize linear recurrent neural nets. train model on length 1 million dependency
HIGHLIGHT: We abstract recent workon linear RNNs into a new framework of linear surrogate RNNs and develop alinear surrogate model for the long short-term memory unit, the GILR-LSTM, thatutilizes parallel linear recurrence.

140, TITLE: Attacking Binarized Neural Networks
https://openreview.net/forum?id=HkTEFfZRb
AUTHORS: Angus Galloway, Graham W. Taylor, Medhat Moussa
TL;DR: We conduct adversarial attacks against binarized neural networks and show that we reduce the impact of the strongest attacks, while maintaining comparable accuracy in a black-box setting
HIGHLIGHT: We propose a third benefit of very low-precision neural networks:improved robustness against some adversarial attacks, and in the worst case,performance that is on par with full-precision models.

141, TITLE: Depthwise Separable Convolutions for Neural Machine Translation
https://openreview.net/forum?id=S1jBcueAb
AUTHORS: Lukasz Kaiser, Aidan N. Gomez, Francois Chollet
TL;DR: Depthwise separable convolutions improve neural machine translation: the more separable the better.
HIGHLIGHT: In this work, we study how depthwise separable convolutions can be applied to neural machine translation.

142, TITLE: Noisy Networks For Exploration
https://openreview.net/forum?id=rywHCPkAW
AUTHORS: Meire Fortunato, Mohammad Gheshlaghi Azar, Bilal Piot, Jacob Menick, Matteo Hessel, Ian Osband, Alex Graves, Volodymyr Mnih, Remi Munos, Demis Hassabis, Olivier Pietquin, Charles Blundell, Shane Legg
TL;DR: A deep reinforcement learning agent with parametric noise added to its weights can be used to aid efficient exploration.
HIGHLIGHT: We introduce NoisyNet, a deep reinforcement learning agent with parametric noise added to its weights, and show that the induced stochasticity of the agent?s policy can be used to aid efficient exploration.

143, TITLE: A Hierarchical Model for Device Placement
https://openreview.net/forum?id=Hkc-TeZ0W
AUTHORS: Azalia Mirhoseini, Anna Goldie, Hieu Pham, Benoit Steiner, Quoc V. Le, Jeff Dean
TL;DR: We introduce a hierarchical model for efficient, end-to-end placement of computational graphs onto hardware devices.
HIGHLIGHT: We introduce a hierarchical model for efficient placement of computational graphs onto hardware devices, especially in heterogeneous environments with a mixture of CPUs, GPUs, and other computational devices.

144, TITLE: Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection
https://openreview.net/forum?id=BJJLHbb0-
AUTHORS: Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, Haifeng Chen
TL;DR: An end-to-end trained deep neural network that leverages Gaussian Mixture Modeling to perform density estimation and unsupervised anomaly detection in a low-dimensional space learned by deep autoencoder.
HIGHLIGHT: In this paper, we present a Deep Autoencoding Gaussian Mixture Model (DAGMM) for unsupervised anomaly detection.

145, TITLE: Learning Discrete Weights Using the Local Reparameterization Trick
https://openreview.net/forum?id=BySRH6CpW
AUTHORS: Oran Shayer, Dan Levi, Ethan Fetaya
TL;DR: Training binary/ternary networks using local reparameterization with the CLT approximation
HIGHLIGHT: In this work, we introduce LR-nets (Local reparameterization networks), a new method for training neural networks with discrete weights using stochastic parameters.

146, TITLE: Deep Rewiring: Training very sparse deep networks
https://openreview.net/forum?id=BJ_wN01C-
AUTHORS: Guillaume Bellec, David Kappel, Wolfgang Maass, Robert Legenstein
TL;DR: The paper presents Deep Rewiring, an algorithm that can be used to train deep neural networks when the network connectivity is severely constrained during training.

HIGHLIGHT: We present an algorithm, DEEP R, that enables us to train directly a sparsely connected neural network.

147, TITLE: Quantitatively Evaluating GANs With Divergences Proposed for Training
https://openreview.net/forum?id=SJQHjzZ0-
AUTHORS: Daniel Jiwoong Im, He Ma, Graham W. Taylor, Kristin Branson
TL;DR: An empirical evaluation on generative adversarial networks
HIGHLIGHT: In this paper, we evaluate the performance of various types of GANs using divergence and distance functions typically used only for training.

148, TITLE: Improving GAN Training via Binarized Representation Entropy (BRE) Regularization
https://openreview.net/forum?id=BkLhaGZRW
AUTHORS: Yanshuai Cao, Gavin Weiguang Ding, Kry Yik-Chau Lui, Ruitong Huang
TL;DR: n/a
HIGHLIGHT: We propose a novel regularizer to improve the training of Generative Adversarial Networks (GANs).

149, TITLE: Generative networks as inverse problems with Scattering transforms
https://openreview.net/forum?id=r1NYjfbR-
AUTHORS: Tom?s Angles, St?phane Mallat
TL;DR: We introduce generative networks that do not require to be learned with a discriminator or an encoder; they are obtained by inverting a special embedding operator defined by a wavelet Scattering transform.
HIGHLIGHT: Generative networks as inverse problems with Scattering transforms

150, TITLE: Critical Percolation as a Framework to Analyze the Training of Deep Networks
https://openreview.net/forum?id=BJGWO9k0Z
AUTHORS: Zohar Ringel, Rodrigo Andrade de Bem
TL;DR: A toy dataset based on critical percolation in a planar graph provides an analytical window to the training dynamics of deep neural networks
HIGHLIGHT: In this paper we approach two relevant deep learning topics: i) tackling of graph structured input data and ii) a better understanding and analysis of deep networks and related learning algorithms.

151, TITLE: On the Expressive Power of Overlapping Architectures of Deep Learning
https://openreview.net/forum?id=HkNGsseC-
AUTHORS: Or Sharir, Amnon Shashua
TL;DR: We analyze how the degree of overlaps between the receptive fields of a convolutional network affects its expressive power.
HIGHLIGHT: In this work, we extend the study of expressive efficiency to the attribute of network connectivity and in particular to the effect of "overlaps" in the convolutional process, i.e., when the stride of the convolution is smaller than its filter size (receptive field).

152, TITLE: Rethinking the Smaller-Norm-Less-Informative Assumption in Channel Pruning of Convolution Layers
https://openreview.net/forum?id=HJ94fqApW
AUTHORS: Jianbo Ye, Xin Lu, Zhe Lin, James Z. Wang
TL;DR: A CNN model pruning method using ISTA and rescaling trick to enforce sparsity of scaling parameters in batch normalization.
HIGHLIGHT: In this paper, we propose a channel pruning technique for accelerating the computations of deep convolutional neural networks (CNNs) that does not critically rely on this assumption.

153, TITLE: Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting
https://openreview.net/forum?id=SJiHXGWAZ
AUTHORS: Yaguang Li, Rose Yu, Cyrus Shahabi, Yan Liu
TL;DR: A neural sequence model that learns to forecast on a directed graph.

HIGHLIGHT: To address these challenges, we propose to model the traffic flow as a diffusion process on a directed graph and introduce Diffusion Convolutional Recurrent Neural Network (DCRNN), a deep learning framework for traffic forecasting that incorporates both spatial and temporal dependency in the traffic flow.

154, TITLE: Simulated+Unsupervised Learning With Adaptive Data Generation and Bidirectional Mappings
https://openreview.net/forum?id=SkHDoG-Cb
AUTHORS: Kangwook Lee, Hoon Kim, Changho Suh
TL;DR: n/a
HIGHLIGHT: Inspired by this limitation, we propose a new S+U learning algorithm, which fully leverage the flexibility of data simulators and bidirectional mappings between synthetic data and real data.

155, TITLE: Relational Neural Expectation Maximization: Unsupervised Discovery of Objects and their Interactions
https://openreview.net/forum?id=ryH20GbRW
AUTHORS: Sjoerd van Steenkiste, Michael Chang, Klaus Greff, J?rgen Schmidhuber
TL;DR: We introduce a novel approach to common-sense physical reasoning that learns to discover objects and model their physical interactions from raw visual images in a purely unsupervised fashion
HIGHLIGHT: To address this problem we present a novel method that learns to discover objects and model their physical interactions from raw visual images in a purely unsupervised fashion.

156, TITLE: Generative Models of Visually Grounded Imagination
https://openreview.net/forum?id=HkCsm6lRb
AUTHORS: Ramakrishna Vedantam, Ian Fischer, Jonathan Huang, Kevin Murphy
TL;DR: A VAE-variant which can create diverse images corresponding to novel concrete or abstract "concepts" described using attribute vectors.
HIGHLIGHT: In this paper, we show how we can modify variational auto-encoders to perform this task.
We also propose a set of easy-to-compute evaluation metrics that capture our intuitive notions of what it means to have good visual imagination, namely correctness, coverage, and compositionality (the 3 C?s).

157, TITLE: Few-shot Autoregressive Density Estimation: Towards Learning to Learn Distributions
https://openreview.net/forum?id=r1wEFyWCW
AUTHORS: Scott Reed, Yutian Chen, Thomas Paine, A?ron van den Oord, S. M. Ali Eslami, Danilo Rezende, Oriol Vinyals, Nando de Freitas
TL;DR: Few-shot learning PixelCNN
HIGHLIGHT: In this paper, we show how 1) neural attention and 2) meta learning techniques can be used in combination with autoregressive models to enable effective few-shot density estimation.

158, TITLE: Compositional Obverter Communication Learning from Raw Visual Input
https://openreview.net/forum?id=rknt2Be0-
AUTHORS: Edward Choi, Angeliki Lazaridou, Nando de Freitas
TL;DR: We train neural network agents to develop a language with compositional properties from raw pixel input.
HIGHLIGHT: In this work, we train neural agents to simultaneously develop visual perception from raw image pixels, and learn to communicate with a sequence of discrete symbols.

159, TITLE: SCAN: Learning Hierarchical Compositional Visual Concepts
https://openreview.net/forum?id=rkN2Il-RZ
AUTHORS: Irina Higgins, Nicolas Sonnerat, Loic Matthey, Arka Pal, Christopher P Burgess, Matko Bo?njak, Murray Shanahan, Matthew Botvinick, Demis Hassabis, Alexander Lerchner
TL;DR: We present a neural variational model for learning language-guided compositional visual concepts.
HIGHLIGHT: This paper describes SCAN (Symbol-Concept Association Network), a new framework for learning such abstractions in the visual domain.

160, TITLE: Hierarchical Density Order Embeddings
https://openreview.net/forum?id=HJCXZQbAZ
AUTHORS: Ben Athiwaratkun, Andrew Gordon Wilson

TL;DR: n/a
HIGHLIGHT: We introduce density order embeddings, which learn hierarchical representations through encapsulation of probability distributions.

161, TITLE: Identifying Analogies Across Domains
https://openreview.net/forum?id=BkN_r2lR-
AUTHORS: Yedid Hoshen, Lior Wolf
TL;DR: Finding correspondences between domains by performing matching/mapping iterations
HIGHLIGHT: In this paper, we tackle this very task of finding exact analogies between datasets i.e. for every image from domain A find an analogous image in domain B.

162, TITLE: Emergence of grid-like representations by training recurrent neural networks to perform spatial localization
https://openreview.net/forum?id=B17JTOe0-
AUTHORS: Christopher J. Cueva, Xue-Xin Wei
TL;DR: To our knowledge, this is the first study to show how neural representations of space, including grid-like cells and border cells as observed in the brain, could emerge from training a recurrent neural network to perform navigation tasks.
HIGHLIGHT: As a new way to understand these neural representations, we trained recurrent neural networks (RNNs) to perform navigation tasks in 2D arenas based on velocity inputs.

163, TITLE: Learning a neural response metric for retinal prosthesis
https://openreview.net/forum?id=HJhIM0xAW
AUTHORS: Nishal P Shah, Sasidhar Madugula, EJ Chichilnisky, Yoram Singer, Jonathon Shlens
TL;DR: Using triplets to learn a metric for comparing neural responses and improve the performance of a prosthesis.
HIGHLIGHT: Here we propose a method to learn such a metric on neural responses, directly from recorded light responses of a population of retinal ganglion cells (RGCs) in the primate retina.

164, TITLE: Few-Shot Learning with Graph Neural Networks
https://openreview.net/forum?id=BJj6qGbRW
AUTHORS: Victor Garcia Satorras, Joan Bruna Estrach
TL;DR: n/a
HIGHLIGHT: We propose to study the problem of few-shot learning with the prism of inference on a partially observed graphical model, constructed from a collection of input images whose label can be either observed or not.

165, TITLE: Semantically Decomposing the Latent Spaces of Generative Adversarial Networks
https://openreview.net/forum?id=S1nQvfgA-
AUTHORS: Chris Donahue, Zachary C. Lipton, Akshay Balsubramani, Julian McAuley
TL;DR: SD-GANs disentangle latent codes according to known commonalities in a dataset (e.g. photographs depicting the same person).
HIGHLIGHT: We propose a new algorithm for training generative adversarial networks to jointly learn latent codes for both identities (e.g. individual humans) and observations (e.g. specific photographs).

166, TITLE: A Framework for the Quantitative Evaluation of Disentangled Representations
https://openreview.net/forum?id=By-7dz-AZ
AUTHORS: Cian Eastwood, Christopher K. I. Williams
TL;DR: n/a
HIGHLIGHT: In this work we propose a framework for the quantitative evaluation of disentangled representations when the ground-truth latent structure is available.

167, TITLE: Meta-Learning for Semi-Supervised Few-Shot Classification
https://openreview.net/forum?id=HJcSzz-CZ
AUTHORS: Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B. Tenenbaum, Hugo Larochelle, Richard S. Zemel

TL;DR: We propose novel extensions of Prototypical Networks that are augmented with the ability to use unlabeled examples when producing prototypes.
HIGHLIGHT: In this work, we advance this few-shot classification paradigm towards a scenario where unlabeled examples are also available within each episode.

168, TITLE: A DIRT-T Approach to Unsupervised Domain Adaptation
https://openreview.net/forum?id=H1q-TM-AW
AUTHORS: Rui Shu, Hung Bui, Hirokazu Narui, Stefano Ermon
TL;DR: SOTA on unsupervised domain adaptation by leveraging the cluster assumption.
HIGHLIGHT: We propose two novel and related models: 1) the Virtual Adversarial Domain Adaptation (VADA) model, which combines domain adversarial training with a penalty term that punishes the violation the cluster assumption; 2) the Decision-boundary Iterative Refinement Training with a Teacher (DIRT-T) model, which takes the VADA model as initialization and employs natural gradient steps to further minimize the cluster assumption violation.

169, TITLE: Generalizing Across Domains via Cross-Gradient Training
https://openreview.net/forum?id=r1Dx7fbCW
AUTHORS: Shiv Shankar*, Vihari Piratla*, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, Sunita Sarawagi
TL;DR: Domain guided augmentation of data provides a robust and stable method of domain generalization
HIGHLIGHT: We present CROSSGRAD , a method to use multi-domain training data to learn a classifier that generalizes to new domains.

170, TITLE: Learning to cluster in order to transfer across domains and tasks
https://openreview.net/forum?id=ByRWCqvT-
AUTHORS: Yen-Chang Hsu, Zhaoyang Lv, Zsolt Kira
TL;DR: A learnable clustering objective to facilitate transfer learning across domains and tasks
HIGHLIGHT: This paper introduces a novel method to perform transfer learning across domains and tasks, formulating it as a problem of learning to cluster.

171, TITLE: Deep Complex Networks
https://openreview.net/forum?id=H1T2hmZAb
AUTHORS: Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, Christopher J Pal
TL;DR: n/a
HIGHLIGHT: In this work, we provide the key atomic components for complex-valued deep neural networks and apply them to convolutional feed-forward networks.

172, TITLE: Skip Connections Eliminate Singularities
https://openreview.net/forum?id=HkwBEMWCZ
AUTHORS: Emin Orhan, Xaq Pitkow
TL;DR: Degenerate manifolds arising from the non-identifiability of the model slow down learning in deep networks; skip connections help by breaking degeneracies.
HIGHLIGHT: Here, we present a novel explanation for the benefits of skip connections in training very deep networks.

173, TITLE: Bi-Directional Block Self-Attention for Fast and Memory-Efficient Sequence Modeling
https://openreview.net/forum?id=H1cWzoxA-
AUTHORS: Tao Shen, Tianyi Zhou, Guodong Long, Jing Jiang, Chengqi Zhang
TL;DR: A self-attention network for RNN/CNN-free sequence encoding with small memory consumption, highly parallelizable computation and state-of-the-art performance on several NLP tasks
HIGHLIGHT: In this paper, we propose a model, called "bi-directional block self-attention network (Bi-BloSAN)", for RNN/CNN-free sequence encoding.

174, TITLE: Routing Networks: Adaptive Selection of Non-Linear Functions for Multi-Task Learning
https://openreview.net/forum?id=ry8dvM-R-

AUTHORS: Clemens Rosenbaum, Tim Klinger, Matthew Riemer
TL;DR: routing networks: a new kind of neural network which learns to adaptively route its input for multi-task learning
HIGHLIGHT: To address this issue we introduce the routing network paradigm, a novel neural network and training algorithm.

175, TITLE: Wavelet Pooling for Convolutional Neural Networks
https://openreview.net/forum?id=rkhlb8lCZ
AUTHORS: Travis Williams, Robert Li
TL;DR: Pooling is achieved using wavelets instead of traditional neighborhood approaches (max, average, etc).
HIGHLIGHT: We introduce Wavelet Pooling as another alternative to traditional neighborhood pooling.

176, TITLE: FearNet: Brain-Inspired Model for Incremental Learning
https://openreview.net/forum?id=SJ1Xmf-Rb
AUTHORS: Ronald Kemker, Christopher Kanan
TL;DR: FearNet is a memory efficient neural-network, inspired by memory formation in the mammalian brain, that is capable of incremental class learning without catastrophic forgetting.
HIGHLIGHT: Here, we propose FearNet for incremental class learning.

177, TITLE: Do GANs learn the distribution? Some Theory and Empirics
https://openreview.net/forum?id=BJehNfW0-
AUTHORS: Sanjeev Arora, Andrej Risteski, Yi Zhang
TL;DR: We propose a support size estimator of GANs's learned distribution to show they indeed suffer from mode collapse, and we prove that encoder-decoder GANs do not avoid the issue as well.
HIGHLIGHT: The foundational paper of Goodfellow et al. (2014) suggested they do, if they were given sufficiently large deep nets, sample size, and computation time.

178, TITLE: Towards Reverse-Engineering Black-Box Neural Networks
https://openreview.net/forum?id=BydjJte0-
AUTHORS: Seong Joon Oh, Max Augustin, Mario Fritz, Bernt Schiele
TL;DR: Querying a black-box neural network reveals a lot of information about it; we propose novel "metamodels" for effectively extracting information from a black box.
HIGHLIGHT: This work shows that such attributes of neural networks can be exposed from a sequence of queries.

179, TITLE: Understanding Deep Neural Networks with Rectified Linear Units
https://openreview.net/forum?id=B1J_rgWRW
AUTHORS: Raman Arora, Amitabh Basu, Poorya Mianjy, Anirbit Mukherjee
TL;DR: This paper 1) characterizes functions representable by ReLU DNNs, 2) formally studies the benefit of depth in such architectures, 3) gives an algorithm to implement empirical risk minimization to global optimality for two layer ReLU nets.
HIGHLIGHT: In this paper we investigate the family of functions representable by deep neural networks (DNN) with rectified linear units (ReLU).

180, TITLE: Training wide residual networks for deployment using a single bit for each weight
https://openreview.net/forum?id=rytNfI1AZ
AUTHORS: Mark D. McDonnell
TL;DR: We train wide residual networks that can be immediately deployed using only a single bit for each convolutional weight, with signficantly better accuracy than past methods.
HIGHLIGHT: Here, we report large improvements in error rates on multiple datasets, for deep convolutional neural networks deployed with 1-bit-per-weight.

181, TITLE: Learn to Pay Attention
https://openreview.net/forum?id=HyzbhfWRW
AUTHORS: Saumya Jetley, Nicholas A. Lord, Namhoon Lee, Philip H. S. Torr

TL;DR: The paper proposes a method for forcing CNNs to leverage spatial attention in learning more object-centric representations that perform better in various respects.
HIGHLIGHT: We propose an end-to-end-trainable attention module for convolutional neural network (CNN) architectures built for image classification.

182, TITLE: Monotonic Chunkwise Attention
https://openreview.net/forum?id=Hko85plCW
AUTHORS: Chung-Cheng Chiu*, Colin Raffel*
TL;DR: An online and linear-time attention mechanism that performs soft attention over adaptively-located chunks of the input sequence.
HIGHLIGHT: To address these issues, we propose Monotonic Chunkwise Attention (MoChA), which adaptively splits the input sequence into small chunks over which soft attention is computed.

183, TITLE: Recasting Gradient-Based Meta-Learning as Hierarchical Bayes
https://openreview.net/forum?id=BJ_UL-k0b
AUTHORS: Erin Grant, Chelsea Finn, Sergey Levine, Trevor Darrell, Thomas Griffiths
TL;DR: A specific gradient-based meta-learning algorithm, MAML, is equivalent to an inference procedure in a hierarchical Bayesian model. We use this connection to improve MAML via methods from approximate inference and curvature estimation.
HIGHLIGHT: We use this opportunity to propose an improvement to the MAML algorithm that makes use of techniques from approximate inference and curvature estimation.

184, TITLE: Don't Decay the Learning Rate, Increase the Batch Size
https://openreview.net/forum?id=B1Yy1BxCZ
AUTHORS: Samuel L. Smith, Pieter-Jan Kindermans, Chris Ying, Quoc V. Le
TL;DR: Decaying the learning rate and increasing the batch size during training are equivalent.
HIGHLIGHT: Here we show one can usually obtain the same learning curve on both training and test sets by instead increasing the batch size during training.

185, TITLE: Kronecker-factored Curvature Approximations for Recurrent Neural Networks
https://openreview.net/forum?id=HyMTkQZAb
AUTHORS: James Martens, Jimmy Ba, Matt Johnson
TL;DR: We extend the K-FAC method to RNNs by developing a new family of Fisher approximations.
HIGHLIGHT: In this work we extend the method to handle RNNs by introducing a novel approximation to the FIM for RNNs.

186, TITLE: Proximal Backpropagation
https://openreview.net/forum?id=ByeqORgAW
AUTHORS: Thomas Frerix, Thomas M?llenhoff, Michael Moeller, Daniel Cremers
TL;DR: n/a
HIGHLIGHT: We propose proximal backpropagation (ProxProp) as a novel algorithm that takes implicit instead of explicit gradient steps to update the network parameters during neural network training.

187, TITLE: Neumann Optimizer: A Practical Optimization Algorithm for Deep Neural Networks
https://openreview.net/forum?id=rkLyJl-0-
AUTHORS: Shankar Krishnan, Ying Xiao, Rif. A. Saurous
TL;DR: We describe a practical optimization algorithm for deep neural networks that works faster and generates better models compared to widely used algorithms.
HIGHLIGHT: In this paper, we present a large batch, stochastic optimization algorithm that is both faster than widely used algorithms for fixed amounts of computation, and also scales up substantially better as more computational resources become available.

188, TITLE: SGD Learns Over-parameterized Networks that Provably Generalize on Linearly Separable Data
https://openreview.net/forum?id=rJ33wwxRb
AUTHORS: Alon Brutzkus, Amir Globerson, Eran Malach, Shai Shalev-Shwartz

TL;DR: We show that SGD learns two-layer over-parameterized neural networks with Leaky ReLU activations that provably generalize on linearly separable data.
HIGHLIGHT: In an attempt to bridge this gap, we study the problem oflearning a two-layer over-parameterized neural network, when the data is generated by a linearly separable function.

189, TITLE: A PAC-Bayesian Approach to Spectrally-Normalized Margin Bounds for Neural Networks
https://openreview.net/forum?id=Skz_WfbCZ
AUTHORS: Behnam Neyshabur, Srinadh Bhojanapalli, Nathan Srebro
TL;DR: n/a
HIGHLIGHT: We present a generalization bound for feedforward neural networks in terms of the product of the spectral norm of the layers and the Frobenius norm of the weights.

190, TITLE: On the importance of single directions for generalization
https://openreview.net/forum?id=r1iuQjxCZ
AUTHORS: Ari S. Morcos, David G.T. Barrett, Neil C. Rabinowitz, Matthew Botvinick
TL;DR: We find that deep networks which generalize poorly are more reliant on single directions than those that generalize well, and evaluate the impact of dropout and batch normalization, as well as class selectivity on single direction reliance.
HIGHLIGHT: Here, we connect these lines of inquiry to demonstrate that a network?s reliance on single directions is a good predictor of its generalization performance, across networks trained on datasets with different fractions of corrupted labels, across ensembles of networks trained on datasets with unmodified labels, across different hyper- parameters, and over the course of training.

191, TITLE: The Implicit Bias of Gradient Descent on Separable Data
https://openreview.net/forum?id=r1q7n9gAb
AUTHORS: Daniel Soudry, Elad Hoffer, Mor Shpigel Nacson, Nathan Srebro
TL;DR: The normalized solution of gradient descent on logistic regression (or a similarly decaying loss) slowly converges to the L2 max margin solution on separable data.
HIGHLIGHT: We show that gradient descent on an unregularized logistic regressionproblem, for almost all separable datasets, converges to the same direction as the max-margin solution.

192, TITLE: Many Paths to Equilibrium: GANs Do Not Need to Decrease a Divergence At Every Step
https://openreview.net/forum?id=ByQpn1ZA-
AUTHORS: William Fedus*, Mihaela Rosca*, Balaji Lakshminarayanan, Andrew M. Dai, Shakir Mohamed, Ian Goodfellow
TL;DR: We find evidence that divergence minimization may not be an accurate characterization of GAN training.
HIGHLIGHT: We show that this view is overly restrictive.

193, TITLE: Adaptive Dropout with Rademacher Complexity Regularization
https://openreview.net/forum?id=S1uxsye0Z
AUTHORS: Ke Zhai, Huan Wang
TL;DR: We propose a novel framework to adaptively adjust the dropout rates for the deep neural network based on a Rademacher complexity bound.
HIGHLIGHT: We propose a novel framework to adaptively adjust the dropout rates for the deep neural network based on a Rademacher complexity bound.

194, TITLE: A Bayesian Perspective on Generalization and Stochastic Gradient Descent
https://openreview.net/forum?id=BJij4yg0Z
AUTHORS: Samuel L. Smith and Quoc V. Le
TL;DR: Generalization is strongly correlated with the Bayesian evidence, and gradient noise drives SGD towards minima whose evidence is large.
HIGHLIGHT: We propose that the noise introduced by small mini-batches drives the parameters towards minima whose evidence is large.

195, TITLE: Implicit Causal Models for Genome-wide Association Studies

https://openreview.net/forum?id=SyELrEeAb
AUTHORS: Dustin Tran, David M. Blei
TL;DR: Implicit models applied to causality and genetics
HIGHLIGHT: In this work, we focus on two challenges in particular: How do we build richer causal models, which can capture highly nonlinear relationships and interactions between multiple causes?

196, TITLE: Sensitivity and Generalization in Neural Networks: an Empirical Study
https://openreview.net/forum?id=HJC2SzZCW
AUTHORS: Roman Novak, Yasaman Bahri, Daniel A. Abolafia, Jeffrey Pennington, Jascha Sohl-Dickstein
TL;DR: We perform massive experimental studies characterizing the relationships between Jacobian norms, linear regions, and generalization.
HIGHLIGHT: In this work, we investigate this tension between complexity and generalization through an extensive empirical exploration of two natural metrics of complexity related to sensitivity to input perturbations.

197, TITLE: Regularizing and Optimizing LSTM Language Models
https://openreview.net/forum?id=SyyGPP0TZ
AUTHORS: Stephen Merity, Nitish Shirish Keskar, Richard Socher
TL;DR: Effective regularization and optimization strategies for LSTM-based language models achieves SOTA on PTB and WT2.
HIGHLIGHT: In this paper, we consider the specific problem of word-level language modeling and investigate strategies for regularizing and optimizing LSTM-based models.

198, TITLE: DCN+: Mixed Objective And Deep Residual Coattention for Question Answering
https://openreview.net/forum?id=H1meywxRW
AUTHORS: Caiming Xiong, Victor Zhong, Richard Socher
TL;DR: We introduce the DCN+ with deep residual coattention and mixed-objective RL, which achieves state of the art performance on the Stanford Question Answering Dataset.
HIGHLIGHT: We propose a mixed objective that combines cross entropy loss with self-critical policy learning, using rewards derived from word overlap to solve the misalignment between evaluation metric and optimization objective.

199, TITLE: Word translation without parallel data
https://openreview.net/forum?id=H196sainb
AUTHORS: Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, Herv? J?gou
TL;DR: Aligning languages without the Rosetta Stone: with no parallel data, we construct bilingual dictionaries using adversarial training, cross-domain local scaling, and an accurate proxy criterion for cross-validation.
HIGHLIGHT: In this work, we show that we can build a bilingual dictionary between two languages without using any parallel corpora, by aligning monolingual word embedding spaces in an unsupervised way.

200, TITLE: All-but-the-Top: Simple and Effective Postprocessing for Word Representations
https://openreview.net/forum?id=HkuGJ3kCb
AUTHORS: Jiaqi Mu, Pramod Viswanath
TL;DR: n/a
HIGHLIGHT: In this paper, we demonstrate a {\em very simple}, and yet counter-intuitive, postprocessing technique -- eliminate the common mean vector and a few top dominating directions from the word vectors -- that renders off-the-shelf representations {\em even stronger}.

201, TITLE: Learning General Purpose Distributed Sentence Representations via Large Scale Multi-task Learning
https://openreview.net/forum?id=B18WgG-CZ
AUTHORS: Sandeep Subramanian, Adam Trischler, Yoshua Bengio, Christopher J Pal
TL;DR: A large-scale multi-task learning framework with diverse training objectives to learn fixed-length sentence representations
HIGHLIGHT: In this work, we present a simple, effective multi-task learning framework for sentence representations that combines the inductive biases of diverse training objectives in a single model.We train this model on several data sources with multiple training objectives on over 100 million sentences.

202, TITLE: Natural Language Inference over Interaction Space
https://openreview.net/forum?id=r1dHXnH6-
AUTHORS: Yichen Gong, Heng Luo, Jian Zhang
TL;DR: show multi-channel attention weight contains semantic feature to solve natural language inference task.
HIGHLIGHT: We introduce Interactive Inference Network (IIN), a novel class of neural network architectures that is able to achieve high-level understanding of the sentence pair by hierarchically extracting semantic features from interaction space.

203, TITLE: Multi-Task Learning for Document Ranking and Query Suggestion
https://openreview.net/forum?id=SJ1nzBeA-
AUTHORS: Wasi Uddin Ahmad, Kai-Wei Chang, Hongning Wang
TL;DR: n/a
HIGHLIGHT: We propose a multi-task learning framework to jointly learn document ranking and query suggestion for web search.

204, TITLE: Distributed Fine-tuning of Language Models on Private Data
https://openreview.net/forum?id=HkgNdt26Z
AUTHORS: Vadim Popov, Mikhail Kudinov, Irina Piontkovskaya, Petr Vytovtov, Alex Nevidomsky
TL;DR: We propose a method of distributed fine-tuning of language models on user devices without collection of private data
HIGHLIGHT: We propose a novel technique that significantly improves prediction quality on users? language compared to a general model and outperforms gradient compression methods in terms of communication efficiency.

205, TITLE: Intrinsic Motivation and Automatic Curricula via Asymmetric Self-Play
https://openreview.net/forum?id=SkT5Yg-RZ
AUTHORS: Sainbayar Sukhbaatar, Zeming Lin, Ilya Kostrikov, Gabriel Synnaeve, Arthur Szlam, Rob Fergus
TL;DR: Unsupervised learning for reinforcement learning using an automatic curriculum of self-play
HIGHLIGHT: In this work we will focus on two kinds of environments: (nearly) reversible environments and environments that can be reset.

206, TITLE: Reinforcement Learning Algorithm Selection
https://openreview.net/forum?id=SyoDInJ0-
AUTHORS: Romain Laroche, Raphael Feraud
TL;DR: This paper formalises the problem of online algorithm selection in the context of Reinforcement Learning.
HIGHLIGHT: The article presents a novel meta-algorithm, called Epochal Stochastic Bandit Algorithm Selection (ESBAS).

207, TITLE: Leave no Trace: Learning to Reset for Safe and Autonomous Reinforcement Learning
https://openreview.net/forum?id=S1vuO-bCW
AUTHORS: Benjamin Eysenbach, Shixiang Gu, Julian Ibarz, Sergey Levine
TL;DR: We propose an autonomous method for safe and efficient reinforcement learning that simultaneously learns a forward and backward policy, with the backward policy resetting the environment for a subsequent attempt.
HIGHLIGHT: In this work, we propose an autonomous method for safe and efficient reinforcement learning that simultaneously learns a forward and backward policy, with the backward policy resetting the environment for a subsequent attempt.

208, TITLE: Consequentialist conditional cooperation in social dilemmas with imperfect information
https://openreview.net/forum?id=BkabRiQpb
AUTHORS: Alexander Peysakhovich, Adam Lerer
TL;DR: We show how to use deep RL to construct agents that can solve social dilemmas beyond matrix games.
HIGHLIGHT: We wish to construct agents that cooperate with pure cooperators, avoid exploitation by pure defectors, and incentivize cooperation from the rest.

209, TITLE: Can Neural Networks Understand Logical Entailment?
https://openreview.net/forum?id=SkZxCk-0Z
AUTHORS: Richard Evans, David Saxton, David Amos, Pushmeet Kohli, Edward Grefenstette
TL;DR: We introduce a new dataset of logical entailments for the purpose of measuring models' ability to capture and exploit the structure of logical expressions against an entailment prediction task.
HIGHLIGHT: We introduce a new dataset of logical entailments for the purpose of measuring models' ability to capture and exploit the structure of logical expressions against an entailment prediction task.

210, TITLE: Cascade Adversarial Machine Learning Regularized with a Unified Embedding
https://openreview.net/forum?id=HyRVBzap-
AUTHORS: Taesik Na, Jong Hwan Ko, Saibal Mukhopadhyay
TL;DR: Cascade adversarial training + low level similarity learning improve robustness against both white box and black box attacks.
HIGHLIGHT: Inspired by this observation, we propose cascade adversarial training, which transfers the knowledge of the end results of adversarial training.

211, TITLE: Mitigating Adversarial Effects Through Randomization
https://openreview.net/forum?id=Sk9yuql0Z
AUTHORS: Cihang Xie, Jianyu Wang, Zhishuai Zhang, Zhou Ren, Alan Yuille
TL;DR: n/a
HIGHLIGHT: In this paper, we propose to utilize randomization at inference time to mitigate adversarial effects.

212, TITLE: Decision Boundary Analysis of Adversarial Examples
https://openreview.net/forum?id=BkpiPMbA-
AUTHORS: Warren He, Bo Li, Dawn Song
TL;DR: Looking at decision boundaries around an input gives you more information than a fixed small neighborhood
HIGHLIGHT: In this paper, we argue that information from larger neighborhoods, such as from more directions and from greater distances, will better characterize the relationship between adversarial examples and the DNN models.

213, TITLE: Matrix capsules with EM routing
https://openreview.net/forum?id=HJWLfGWRb
AUTHORS: Geoffrey E Hinton, Sara Sabour, Nicholas Frosst
TL;DR: Capsule networks with learned pose matrices and EM routing improves state of the art classification on smallNORB, improves generalizability to new view points, and white box adversarial robustness.
HIGHLIGHT: We describe a version of capsules in which each capsule has a logistic unit to represent the presence of an entity and a 4x4 matrix which could learn to represent the relationship between that entity and the viewer (the pose).

214, TITLE: CausalGAN: Learning Causal Implicit Generative Models with Adversarial Training
https://openreview.net/forum?id=BJE-4xW0W
AUTHORS: Murat Kocaoglu, Christopher Snyder, Alexandros G. Dimakis, Sriram Vishwanath
TL;DR: We introduce causal implicit generative models, which can sample from conditional and interventional distributions and also propose two new conditional GANs which we use for training them.
HIGHLIGHT: We introduce causal implicit generative models (CiGMs): models that allow sampling from not only the true observational but also the true interventional distributions.

215, TITLE: Learning Wasserstein Embeddings
https://openreview.net/forum?id=SJyEH91A-
AUTHORS: Nicolas Courty, R?mi Flamary, M?lanie Ducoffe
TL;DR: We show that it is possible to fastly approximate Wasserstein distances computation by finding an appropriate embedding where Euclidean distance emulates the Wasserstein distance
HIGHLIGHT: Our goal is to alleviate this problem by providing an approximation mechanism that allows to break its inherent complexity.

216, TITLE: TRAINING GENERATIVE ADVERSARIAL NETWORKS VIA PRIMAL-DUAL SUBGRADIENT METHODS: A LAGRANGIAN PERSPECTIVE ON GAN
https://openreview.net/forum?id=BJNRFNlRW
AUTHORS: Xu Chen, Jiang Wang, Hao Ge
TL;DR: We propose a primal-dual subgradient method for training GANs and this method effectively alleviates mode collapse.
HIGHLIGHT: We relate the minimax game of generative adversarial networks (GANs) to finding the saddle points of the Lagrangian function for a convex optimization problem, where the discriminator outputs and the distribution of generator outputs play the roles of primal variables and dual variables, respectively.

217, TITLE: Activation Maximization Generative Adversarial Nets
https://openreview.net/forum?id=HyyP33gAZ
AUTHORS: Zhiming Zhou, Han Cai, Shu Rong, Yuxuan Song, Kan Ren, Weinan Zhang, Jun Wang, Yong Yu
TL;DR: Understand how class labels help GAN training. Propose a new evaluation metric for generative models.
HIGHLIGHT: Based on that, we propose Activation Maximization Generative Adversarial Networks (AM-GAN) as an advanced solution.

218, TITLE: Coulomb GANs: Provably Optimal Nash Equilibria via Potential Fields
https://openreview.net/forum?id=SkVqXOxCb
AUTHORS: Thomas Unterthiner, Bernhard Nessler, Calvin Seward, G?nter Klambauer, Martin Heusel, Hubert Ramsauer, Sepp Hochreiter
TL;DR: Coulomb GANs can optimally learn a distribution by posing the distribution learning problem as optimizing a potential field
HIGHLIGHT: We introduce Coulomb GANs, which pose the GAN learning problem as a potential field, where generated samples are attracted to training set samples but repel each other.

219, TITLE: Improving the Improved Training of Wasserstein GANs: A Consistency Term and Its Dual Effect
https://openreview.net/forum?id=SJx9GQb0-
AUTHORS: Xiang Wei, Boqing Gong, Zixia Liu, Wei Lu, Liqiang Wang
TL;DR: n/a
HIGHLIGHT: In this paper, we propose a novel approach for enforcing the Lipschitz continuity in the training procedure of WGANs.

220, TITLE: FusionNet: Fusing via Fully-aware Attention with Application to Machine Comprehension
https://openreview.net/forum?id=BJIgi_eCZ
AUTHORS: Hsin-Yuan Huang, Chenguang Zhu, Yelong Shen, Weizhu Chen
TL;DR: We propose a light-weight enhancement for attention and a neural architecture, FusionNet, to achieve SotA on SQuAD and adversarial SQuAD.
HIGHLIGHT: This paper introduces a new neural structure called FusionNet, which extends existing attention approaches from three perspectives.

221, TITLE: Neural Language Modeling by Jointly Learning Syntax and Lexicon
https://openreview.net/forum?id=rkgOLb-0W
AUTHORS: Yikang Shen, Zhouhan Lin, Chin-wei Huang, Aaron Courville
TL;DR: In this paper, We propose a novel neural language model, called the Parsing-Reading-Predict Networks (PRPN), that can simultaneously induce the syntactic structure from unannotated sentences and leverage the inferred structure to learn a better language model.
HIGHLIGHT: We propose a neural language model capable of unsupervised syntactic structure induction.

222, TITLE: Learning Intrinsic Sparse Structures within Long Short-Term Memory
https://openreview.net/forum?id=rk6cfpRjZ
AUTHORS: Wei Wen, Yuxiong He, Samyam Rajbhandari, Minjia Zhang, Wenhan Wang, Fang Liu, Bin Hu, Yiran Chen, Hai Li
TL;DR: n/a
HIGHLIGHT: This work aims to learn structurally-sparse Long Short-Term Memory (LSTM) by reducing the sizes of basic structures within LSTM units, including input updates, gates, hidden states, cell states and outputs.

223, TITLE: Deep Active Learning for Named Entity Recognition
https://openreview.net/forum?id=ry018WZAZ
AUTHORS: Yanyao Shen, Hyokun Yun, Zachary C. Lipton, Yakov Kronrod, Animashree Anandkumar
TL;DR: We introduce a lightweight architecture for named entity recognition and carry out incremental active learning, which is able to match state-of-the-art performance with just 25% of the original training data.
HIGHLIGHT: In this work, we demonstrate that the amount of labeled training data can be drastically reduced when deep learning is combined with active learning.

224, TITLE: Go for a Walk and Arrive at the Answer: Reasoning Over Paths in Knowledge Bases using Reinforcement Learning
https://openreview.net/forum?id=Syg-YfWCW
AUTHORS: Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Luke Vilnis, Ishan Durugkar, Akshay Krishnamurthy, Alex Smola, Andrew McCallum
TL;DR: We present a RL agent MINERVA which learns to walk on a knowledge graph and answer queries
HIGHLIGHT: We propose a new algorithm, MINERVA, which addresses the much more difficult and practical task of answering questions where the relation is known, but only one entity.

225, TITLE: No Title
https://openreview.net/forum?id=
AUTHORS:
TL;DR: n/a
HIGHLIGHT: No Title

226, TITLE: Lifelong Learning with Dynamically Expandable Networks
https://openreview.net/forum?id=Sk7KsfW0-
AUTHORS: Jaehong Yoon, Eunho Yang, Jeongtae Lee, Sung Ju Hwang
TL;DR: We propose a novel deep network architecture that can dynamically decide its network capacity as it trains on a lifelong learning scenario.
HIGHLIGHT: We propose a novel deep network architecture for lifelong learning which we refer to as Dynamically Expandable Network (DEN), that can dynamically decide its network capacity as it trains on a sequence of tasks, to learn a compact overlapping knowledge sharing structure among tasks.

227, TITLE: The Role of Minimal Complexity Functions in Unsupervised Learning of Semantic Mappings
https://openreview.net/forum?id=H1VjBebR-
AUTHORS: Tomer Galanti, Lior Wolf, Sagie Benaim
TL;DR: Our hypothesis is that given two domains, the lowest complexity mapping that has a low discrepancy approximates the target mapping.
HIGHLIGHT: We discuss the feasibility of the following learning problem: given unmatched samples from two domains and nothing else, learn a mapping between the two, which preserves semantics.

228, TITLE: Dynamic Neural Program Embeddings for Program Repair
https://openreview.net/forum?id=BJuWrGW0Z
AUTHORS: Ke Wang, Rishabh Singh, Zhendong Su
TL;DR: A new way of learning semantic program embedding
HIGHLIGHT: We propose a novel semantic program embedding that is learned from program execution traces.

229, TITLE: Compositional Attention Networks for Machine Reasoning
https://openreview.net/forum?id=S1Euwz-Rb
AUTHORS: Drew A. Hudson, Christopher D. Manning
TL;DR: We present a novel architecture, based on dynamic memory, attention and composition for the task of machine reasoning.
HIGHLIGHT: We present Compositional Attention Networks, a novel fully differentiable neural network architecture, designed to facilitate explicit and expressive reasoning.

230, TITLE: Beyond Shared Hierarchies: Deep Multitask Learning through Soft Layer Ordering
https://openreview.net/forum?id=BkXmYfbAZ
AUTHORS: Elliot Meyerson, Risto Miikkulainen
TL;DR: Relaxing the constraint of shared hierarchies enables more effective deep multitask learning.
HIGHLIGHT: Beyond Shared Hierarchies: Deep Multitask Learning through Soft Layer Ordering

231, TITLE: Hierarchical Representations for Efficient Architecture Search
https://openreview.net/forum?id=BJQRKzbA-
AUTHORS: Hanxiao Liu, Karen Simonyan, Oriol Vinyals, Chrisantha Fernando, Koray Kavukcuoglu
TL;DR: In this paper we propose a hierarchical architecture representation in which doing random or evolutionary architecture search yields highly competitive results using fewer computational resources than the prior art.
HIGHLIGHT: We explore efficient neural architecture search methods and show that a simple yet powerful evolutionary algorithm can discover new architectures with excellent performance.

232, TITLE: Reinforcement Learning on Web Interfaces using Workflow-Guided Exploration
https://openreview.net/forum?id=ryTp3f-0-
AUTHORS: Evan Zheran Liu, Kelvin Guu, Panupong Pasupat, Tianlin Shi, Percy Liang
TL;DR: We solve the sparse rewards problem on web UI tasks using exploration guided by demonstrations
HIGHLIGHT: Instead, we propose to constrain exploration using demonstrations.

233, TITLE: Combining Symbolic Expressions and Black-box Function Evaluations in Neural Programs
https://openreview.net/forum?id=Hksj2WWAW
AUTHORS: Forough Arabshahi, Sameer Singh, Animashree Anandkumar
TL;DR: n/a
HIGHLIGHT: We present a novel framework that utilizes black-box function evaluations, in conjunction with symbolic expressions that define relationships between the given functions.
We present an evaluation benchmark for this task to demonstrate our proposed model combines symbolic reasoning and function evaluation in a fruitful manner, obtaining high accuracies in our experiments.

234, TITLE: Scalable Private Learning with PATE
https://openreview.net/forum?id=rkZB1XbRZ
AUTHORS: Nicolas Papernot, Shuang Song, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, Ulfar Erlingsson
TL;DR: n/a
HIGHLIGHT: For this, we introduce new noisy aggregation mechanisms for teacher ensembles that are more selective and add less noise, and prove their tighter differential-privacy guarantees.

235, TITLE: Active Learning for Convolutional Neural Networks: A Core-Set Approach
https://openreview.net/forum?id=H1aIuk-RW
AUTHORS: Ozan Sener, Silvio Savarese
TL;DR: We approach to the problem of active learning as a core-set selection problem and show that this approach is especially useful in the batch active learning setting which is crucial when training CNNs.
HIGHLIGHT: Inspired by these limitations, we define the problem of active learning as core-set selection, i.e. choosing set of points such that a model learned over the selected subset is competitive for the remaining data points.

236, TITLE: Loss-aware Weight Quantization of Deep Networks
https://openreview.net/forum?id=BkrSv0lA-
AUTHORS: Lu Hou, James T. Kwok
TL;DR: A loss-aware weight quantization algorithm that directly considers its effect on the loss is proposed.
HIGHLIGHT: In this paper, we consider compressing the network by weight quantization.

237, TITLE: Global Optimality Conditions for Deep Neural Networks
https://openreview.net/forum?id=BJk7Gf-CZ
AUTHORS: Chulhee Yun, Suvrit Sra, Ali Jadbabaie

TL;DR: We provide efficiently checkable necessary and sufficient conditions for global optimality in deep linear neural networks, with some initial extensions to nonlinear settings.
HIGHLIGHT: For deep linear networks, we present necessary and sufficient conditions for a critical point of the risk function to be a global minimum.

238, TITLE: SpectralNet: Spectral Clustering using Deep Neural Networks
https://openreview.net/forum?id=HJ_aoCyRZ
AUTHORS: Uri Shaham, Kelly Stanton, Henry Li, Ronen Basri, Boaz Nadler, Yuval Kluger
TL;DR: Unsupervised spectral clustering using deep neural networks
HIGHLIGHT: In this paper we introduce a deep learning approach to spectral clustering that overcomes the above shortcomings.

239, TITLE: Not-So-Random Features
https://openreview.net/forum?id=Hk8XMWgRb
AUTHORS: Brian Bullins, Cyril Zhang, Yi Zhang
TL;DR: A simple and practical algorithm for learning a margin-maximizing translation-invariant or spherically symmetric kernel from training data, using tools from Fourier analysis and regret minimization.
HIGHLIGHT: We propose a principled method for kernel learning, which relies on a Fourier-analytic characterization of translation-invariant or rotation-invariant kernels.

240, TITLE: Learning how to explain neural networks: PatternNet and PatternAttribution
https://openreview.net/forum?id=Hkn7CBaTW
AUTHORS: Pieter-Jan Kindermans, Kristof T. Sch?tt, Maximilian Alber, Klaus-Robert M?ller, Dumitru Erhan, Been Kim, Sven D?hne
TL;DR: Without learning, it is impossible to explain a machine learning model's decisions.
HIGHLIGHT: Based on our analysis of linear models we propose a generalization that yields two explanation techniques (PatternNet and PatternAttribution) that are theoretically sound for linear models and produce improved explanations for deep networks.

241, TITLE: Detecting Statistical Interactions from Neural Network Weights
https://openreview.net/forum?id=ByOfBggRZ
AUTHORS: Michael Tsang, Dehua Cheng, Yan Liu
TL;DR: We detect statistical interactions captured by a feedforward multilayer neural network by directly interpreting its learned weights.
HIGHLIGHT: In this paper, we develop a novel framework for detecting statistical interactions captured by a feedforward multilayer neural network by directly interpreting its learned weights.

242, TITLE: Deep Gaussian Embedding of Graphs: Unsupervised Inductive Learning via Ranking
https://openreview.net/forum?id=r1ZdKJ-0W
AUTHORS: Aleksandar Bojchevski, Stephan G?nnemann
TL;DR: We embed nodes in a graph as Gaussian distributions allowing us to capture uncertainty about their representation.
HIGHLIGHT: We propose Graph2Gauss - an approach that can efficiently learn versatile node embeddings on large scale (attributed) graphs that show strong performance on tasks such as link prediction and node classification.

243, TITLE: Generating Natural Adversarial Examples
https://openreview.net/forum?id=H1BLjgZCb
AUTHORS: Zhengli Zhao, Dheeru Dua, Sameer Singh
TL;DR: We propose a framework to generate ?natural? adversaries against black-box classifiers for both visual and textual domains, by doing the search for adversaries in the latent semantic space.
HIGHLIGHT: In this paper, we propose a framework to generate natural and legible adversarial examples that lie on the data manifold, by searching in semantic space of dense and continuous data representation, utilizing the recent advances in generative adversarial networks.

244, TITLE: Spatially Transformed Adversarial Examples

https://openreview.net/forum?id=HyydRMZC-
AUTHORS: Chaowei Xiao, Jun-Yan Zhu, Bo Li, Warren He, Mingyan Liu, Dawn Song
TL;DR: We propose a new approach for generating adversarial examples based on spatial transformation, which produces perceptually realistic examples compared to existing attacks.
HIGHLIGHT: Recent studies show that widely used Deep neural networks (DNNs) are vulnerable to the carefully crafted adversarial examples.Many advanced algorithms have been proposed to generate adversarial examples by leveraging the L_p distance for penalizing perturbations.Different defense methods have also been explored to defend against such adversarial attacks.While the effectiveness of L_p distance as a metric of perceptual quality remains an active research area, in this paper we will instead focus on a different type of perturbation, namely spatial transformation, as opposed to manipulating the pixel values directly as in prior works.Perturbations generated through spatial transformation could result in large L_p distance measures, but our extensive experiments show that such spatially transformed adversarial examples are perceptually realistic and more difficult to defend against with existing defense systems.

245, TITLE: Predicting Floor-Level for 911 Calls with Neural Networks and Smartphone Sensor Data
https://openreview.net/forum?id=ryBnUWb0b
AUTHORS: William Falcon, Henning Schulzrinne
TL;DR: We used an LSTM to detect when a smartphone walks into a building. Then we predict the device's floor level using data from sensors aboard the smartphone.
HIGHLIGHT: We introduce a system to estimate a victim's floor level via their mobile device's sensor data in a two-step process.

246, TITLE: Understanding image motion with group representations
https://openreview.net/forum?id=SJLlmG-AZ
AUTHORS: Andrew Jaegle, Stephen Phillips, Daphne Ippolito, Kostas Daniilidis
TL;DR: We propose of method of using group properties to learn a representation of motion without labels and demonstrate the use of this method for representing 2D and 3D motion.
HIGHLIGHT: We propose a model of motion based on elementary group properties of transformations and use it to train a representation of image motion.

247, TITLE: Learning Awareness Models
https://openreview.net/forum?id=r1HhRfWRZ
AUTHORS: Brandon Amos, Laurent Dinh, Serkan Cabi, Thomas Roth?rl, Sergio G?mez Colmenarejo, Alistair Muldal, Tom Erez, Yuval Tassa, Nando de Freitas, Misha Denil
TL;DR: We train predictive models on proprioceptive information and show they represent properties of external objects.
HIGHLIGHT: We show that active data collection by maximizing the entropy of predictions about the body---touch sensors, proprioception and vestibular information---leads to learning of dynamic models that show superior performance when used for control.
We also collect data from a real robotic hand and show that the same models can be used to answer questions about properties of objects in the real world.

248, TITLE: Backpropagation through the Void: Optimizing control variates for black-box gradient estimation
https://openreview.net/forum?id=SyzKd1bCW
AUTHORS: Will Grathwohl, Dami Choi, Yuhuai Wu, Geoff Roeder, David Duvenaud
TL;DR: We present a general method for unbiased estimation of gradients of black-box functions of random variables. We apply this method to discrete variational inference and reinforcement learning.
HIGHLIGHT: We introduce a general framework for learning low-variance, unbiased gradient estimators for black-box functions of random variables, based on gradients of a learned function.These estimators can be jointly trained with model parameters or policies, and are applicable in both discrete and continuous settings.

249, TITLE: On Unifying Deep Generative Models
https://openreview.net/forum?id=rylSzl-R-
AUTHORS: Zhiting Hu, Zichao Yang, Ruslan Salakhutdinov, Eric P. Xing
TL;DR: A unified statistical view of the broad class of deep generative models
HIGHLIGHT: This paper aims to establish formal connections between GANs and VAEs through a new formulation of them.

250, TITLE: Debiasing Evidence Approximations: On Importance-weighted Autoencoders and Jackknife Variational Inference
https://openreview.net/forum?id=HyZoi-WRb
AUTHORS: Sebastian Nowozin
TL;DR: Variational inference is biased, let's debias it.
HIGHLIGHT: In this work, we provide yet another perspective on the IWAE bounds.

251, TITLE: Learning a Generative Model for Validity in Complex Discrete Structures
https://openreview.net/forum?id=rkrC3GbRW
AUTHORS: Dave Janz, Jos van der Westhuizen, Brooks Paige, Matt Kusner, Jos? Miguel Hern?ndez-Lobato
TL;DR: n/a
HIGHLIGHT: As a step towards solving this problem, we propose to learn a deep recurrent validator model, which can estimate whether a partial sequence can function as the beginning of a full, valid sequence.

252, TITLE: Boundary Seeking GANs
https://openreview.net/forum?id=rkTS8lZAb
AUTHORS: R Devon Hjelm, Athul Paul Jacob, Adam Trischler, Gerry Che, Kyunghyun Cho, Yoshua Bengio
TL;DR: We address training GANs with discrete data by formulating a policy gradient that generalizes across f-divergences
HIGHLIGHT: We introduce a method for training GANs with discrete data that uses the estimated difference measure from the discriminator to compute importance weights for generated samples, thus providing a policy gradient for training the generator.

253, TITLE: Learning Sparse Latent Representations with the Deep Copula Information Bottleneck
https://openreview.net/forum?id=Hk0wHx-RW
AUTHORS: Aleksander Wieczorek*, Mario Wieser*, Damian Murezzan, Volker Roth
TL;DR: We apply the copula transformation to the Deep Information Bottleneck which leads to restored invariance properties and a disentangled latent space with superior predictive capabilities.
HIGHLIGHT: In this paper, we adopt the deep information bottleneck model, identify its shortcomings and propose a model that circumvents them.

254, TITLE: WHAI: Weibull Hybrid Autoencoding Inference for Deep Topic Modeling
https://openreview.net/forum?id=S1cZsf-RW
AUTHORS: Hao Zhang, Bo Chen, Dandan Guo, Mingyuan Zhou
TL;DR: n/a
HIGHLIGHT: To train an inference network jointly with a deep generative topic model, making it both scalable to big corpora and fast in out-of-sample prediction, we develop Weibull hybrid autoencoding inference (WHAI) for deep latent Dirichlet allocation, which infers posterior samples via a hybrid of stochastic-gradient MCMC and autoencoding variational Bayes.

255, TITLE: Understanding Short-Horizon Bias in Stochastic Meta-Optimization
https://openreview.net/forum?id=H1MczcgR-
AUTHORS: Yuhuai Wu, Mengye Ren, Renjie Liao, Roger Grosse.
TL;DR: We investigate the bias in the short-horizon meta-optimization objective.
HIGHLIGHT: We introduce a toy problem, a noisy quadratic cost function, on which we analyze short-horizon bias by deriving and comparing the optimal schedules for short and long time horizons.

256, TITLE: Self-ensembling for visual domain adaptation
https://openreview.net/forum?id=rkpoTaxA-
AUTHORS: Geoff French, Michal Mackiewicz, Mark Fisher
TL;DR: Self-ensembling based algorithm for visual domain adaptation, state of the art results, won VisDA-2017 image classification domain adaptation challenge.
HIGHLIGHT: We introduce a number of modifications to their approach for challenging domain adaptation scenarios and evaluate its effectiveness.

257, TITLE: Gradient Estimators for Implicit Models
https://openreview.net/forum?id=SJi9WOeRb
AUTHORS: Yingzhen Li, Richard E. Turner
TL;DR: We introduced a novel gradient estimator using Stein's method, and compared with other methods on learning implicit models for approximate inference and image generation.
HIGHLIGHT: Some examples include data simulators that are widely used in engineering and scientific research, generative adversarial networks (GANs) for image synthesis, and hot-off-the-press approximate inference techniques relying on implicit distributions.

258, TITLE: Learning to Multi-Task by Active Sampling
https://openreview.net/forum?id=B1nZ1weCZ
AUTHORS: Sahil Sharma*, Ashutosh Kumar Jha*, Parikshit S Hegde, Balaraman Ravindran
TL;DR: Letting a meta-learner decide the task to train on for an agent in a multi-task setting improves multi-tasking ability substantially
HIGHLIGHT: We propose three distinct models under our active sampling framework.

259, TITLE: Learning Robust Rewards with Adverserial Inverse Reinforcement Learning
https://openreview.net/forum?id=rkHywl-A-
AUTHORS: Justin Fu, Katie Luo, Sergey Levine
TL;DR: We propose an adversarial inverse reinforcement learning algorithm capable of learning reward functions which can transfer to new, unseen environments.
HIGHLIGHT: Inthis work, we propose AIRL, a practical and scalable inverse reinforcement learningalgorithm based on an adversarial reward learning formulation that is competitivewith direct imitation learning algorithms.

260, TITLE: A Simple Neural Attentive Meta-Learner
https://openreview.net/forum?id=B1DmUzWAW
AUTHORS: Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, Pieter Abbeel
TL;DR: a simple RNN-based meta-learner that achieves SOTA performance on popular benchmarks
HIGHLIGHT: In response, recent work in meta-learning proposes training a meta-learner on a distribution of similar tasks, in the hopes of generalization to novel but related tasks by learning a high-level strategy that captures the essence of the problem it is asked to solve.

261, TITLE: Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design
https://openreview.net/forum?id=SywXXwJAb
AUTHORS: Yoav Levine, David Yakira, Nadav Cohen, Amnon Shashua
TL;DR: Employing quantum entanglement measures for quantifying correlations in deep learning, and using the connection to fit the deep network's architecture to correlations in the data.
HIGHLIGHT: In this work, we establish a fundamental connection between the fields of quantum physics and deep learning, and use it for obtaining novel theoretical observations regarding the inductive bias of convolutional networks.

262, TITLE: Towards Synthesizing Complex Programs From Input-Output Examples
https://openreview.net/forum?id=Skp1ESxRZ
AUTHORS: Xinyun Chen, Chang Liu, Dawn Song
TL;DR: n/a
HIGHLIGHT: In this work, we move a significant step forward along this direction by proposing a new class of challenging tasks in the domain of program synthesis from input-output examples: learning a context-free parser from pairs of input programs and their parse trees.

263, TITLE: Expressive power of recurrent neural networks
https://openreview.net/forum?id=S1WRibb0Z
AUTHORS: Valentin Khrulkov, Alexander Novikov, Ivan Oseledets
TL;DR: We prove the exponential efficiency of recurrent-type neural networks over shallow networks.

HIGHLIGHT: In this paper,we prove the expressive power theorem (an exponential lower bound onthe width of the equivalent shallow network) for a class of recurrent neuralnetworks ? ones that correspond to the Tensor Train (TT) decomposition.This means that even processing an image patch by patch with an RNNcan be exponentially more efficient than a (shallow) convolutional networkwith one hidden layer.

264, TITLE: Improving the Universality and Learnability of Neural Programmer-Interpreters with Combinator Abstraction
https://openreview.net/forum?id=rJlMAAeC-
AUTHORS: Da Xiao, Jo-Yu Liao, Xingyuan Yuan
TL;DR: n/a
HIGHLIGHT: To overcome the limitations of Neural Programmer-Interpreters (NPI) in its universality and learnability, we propose the incorporation of combinator abstraction into neural programing and a new NPI architecture to support this abstraction, which we call Combinatory Neural Programmer-Interpreter (CNPI).
We propose a small set of four combinators to capture the most pervasive programming patterns.

265, TITLE: An image representation based convolutional network for DNA classification
https://openreview.net/forum?id=HJvvRoe0W
AUTHORS: Bojian Yin, Marleen Balvert, Davide Zambrano, Alexander Schoenhuth, Sander Bohte
TL;DR: A method to transform DNA sequences into 2D images using space-filling Hilbert Curves to enhance the strengths of CNNs
HIGHLIGHT: In this paper we develop a convolutional neural network that takes an image-representation of primary DNA sequence as its input, and predicts key determinants of chromatin structure.

266, TITLE: SMASH: One-Shot Model Architecture Search through HyperNetworks
https://openreview.net/forum?id=rydeCEhs-
AUTHORS: Andrew Brock, Theo Lim, J.M. Ritchie, Nick Weston
TL;DR: A technique for accelerating neural architecture selection by approximating the weights of each candidate architecture instead of training them individually.
HIGHLIGHT: We propose a technique to accelerate architecture selection by learning an auxiliary HyperNet that generates the weights of a main model conditioned on that model's architecture.

267, TITLE: Parameter Space Noise for Exploration
https://openreview.net/forum?id=ByBAl2eAZ
AUTHORS: Matthias Plappert, Rein Houthooft, Prafulla Dhariwal, Szymon Sidor, Richard Y. Chen, Xi Chen, Tamim Asfour, Pieter Abbeel, Marcin Andrychowicz
TL;DR: Parameter space noise allows reinforcement learning algorithms to explore by perturbing parameters instead of actions, often leading to significantly improved exploration performance.
HIGHLIGHT: Methods such as evolutionary strategies use parameter perturbations, but discard all temporal structure in the process and require significantly more samples.

268, TITLE: Synthesizing realistic neural population activity patterns using Generative Adversarial Networks
https://openreview.net/forum?id=r1VVsebAZ
AUTHORS: Manuel Molano-Mazon, Arno Onken, Eugenio Piasini*, Stefano Panzeri*
TL;DR: Using Wasserstein-GANs to generate realistic neural activity and to detect the most relevant features present in neural population patterns.
HIGHLIGHT: Here we used the Generative Adversarial Networks (GANs) framework to simulate the concerted activity of a population of neurons.We adapted the Wasserstein-GAN variant to facilitate the generation of unconstrained neural population activity patterns while still benefiting from parameter sharing in the temporal domain.We demonstrate that our proposed GAN, which we termed Spike-GAN, generates spike trains that match accurately the first- and second-order statistics of datasets of tens of neurons and also approximates well their higher-order statistics.

269, TITLE: Auto-Encoding Sequential Monte Carlo
https://openreview.net/forum?id=BJ8c3f-0b
AUTHORS: Tuan Anh Le, Maximilian Igl, Tom Rainforth, Tom Jin, Frank Wood

TL;DR: We build on auto-encoding sequential Monte Carlo, gain new theoretical insights and develop an improved training procedure based on those insights.
HIGHLIGHT: We develop additional theoretical insights and introduce a new training procedure which improves both model and proposal learning.

270, TITLE: Learning to Teach
https://openreview.net/forum?id=HJewuJWCZ
AUTHORS: Yang Fan, Fei Tian, Tao Qin, Xiang-Yang Li
TL;DR: We propose and verify the effectiveness of learning to teach, a new framework to automatically guide machine learning process.
HIGHLIGHT: We call this approach ``learning to teach".

271, TITLE: PixelNN: Example-based Image Synthesis
https://openreview.net/forum?id=Syhr6pxCW
AUTHORS: Aayush Bansal, Yaser Sheikh, Deva Ramanan
TL;DR: Pixel-wise nearest neighbors used for generating multiple images from incomplete priors such as a low-res images, surface normals, edges etc.
HIGHLIGHT: We present a simple nearest-neighbor (NN) approach that synthesizes high-frequency photorealistic images from an ``incomplete" signal such as a low-resolution image, a surface normal map, or edges.

272, TITLE: Non-Autoregressive Neural Machine Translation
https://openreview.net/forum?id=B1l8BtlCb
AUTHORS: Jiatao Gu, James Bradbury, Caiming Xiong, Victor O.K. Li, Richard Socher
TL;DR: We introduce the first NMT model with fully parallel decoding, reducing inference latency by 10x.
HIGHLIGHT: We introduce a model that avoids this autoregressive property and produces its outputs in parallel, allowing an order of magnitude lower latency during inference.

273, TITLE: Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning
https://openreview.net/forum?id=HJtEm4p6Z
AUTHORS: Wei Ping, Kainan Peng, Andrew Gibiansky, Sercan O. Arik, Ajay Kannan, Sharan Narang, Jonathan Raiman, John Miller
TL;DR: n/a
HIGHLIGHT: We present Deep Voice 3, a fully-convolutional attention-based neural text-to-speech (TTS) system.

274, TITLE: mixup: Beyond Empirical Risk Minimization
https://openreview.net/forum?id=r1Ddp1-Rb
AUTHORS: Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz
TL;DR: Training on convex combinations between random training examples and their labels improves generalization in deep neural networks
HIGHLIGHT: In this work, we propose mixup, a simple learning principle to alleviate these issues.

275, TITLE: TD or not TD: Analyzing the Role of Temporal Differencing in Deep Reinforcement Learning
https://openreview.net/forum?id=HyiAuyb0b
AUTHORS: Artemij Amiranashvili, Alexey Dosovitskiy, Vladlen Koltun, Thomas Brox
TL;DR: n/a
HIGHLIGHT: In this paper, we re-examine the role of TD in modern deep RL, using specially designed environments that control for specific factors that affect performance, such as reward sparsity, reward delay, and the perceptual complexity of the task.

276, TITLE: DORA The Explorer: Directed Outreaching Reinforcement Action-Selection
https://openreview.net/forum?id=ry1arUgCW
AUTHORS: Lior Fox, Leshem Choshen, Yonatan Loewenstein
TL;DR: We propose a generalization of visit-counters that evaluate the propagating exploratory value over trajectories, enabling efficient exploration for model-free RL

HIGHLIGHT: We compare our approach to commonly used RL techniques, and show that using E-values improves learning and performance over traditional counters.

277, TITLE: Temporal Difference Models: Model-Free Deep RL for Model-Based Control
https://openreview.net/forum?id=Skw0n-W0Z
AUTHORS: Vitchyr Pong*, Shixiang Gu*, Murtaza Dalal, Sergey Levine
TL;DR: We show that a special goal-condition value function trained with model free methods can be used within model-based control, resulting in substantially better sample efficiency and performance.
HIGHLIGHT: We introduce temporal difference models (TDMs), a family of goal-conditioned value functions that can be trained with model-free learning and used for model-based control.

278, TITLE: TreeQN and ATreeC: Differentiable Tree-Structured Models for Deep Reinforcement Learning
https://openreview.net/forum?id=H1dh6Ax0Z
AUTHORS: Gregory Farquhar, Tim Rockt?schel, Maximilian Igl, Shimon Whiteson
TL;DR: We present TreeQN and ATreeC, new architectures for deep reinforcement learning in discrete-action domains that integrate differentiable on-line tree planning into the action-value function or policy.
HIGHLIGHT: To address these challenges, we propose TreeQN, a differentiable, recursive, tree-structured model that serves as a drop-in replacement for any value function network in deep RL with discrete actions.

279, TITLE: Alternating Multi-bit Quantization for Recurrent Neural Networks
https://openreview.net/forum?id=S19dR9x0b
AUTHORS: Chen Xu, Jianqiang Yao, Zhouchen Lin, Wenwu Ou, Yuanbin Cao, Zhirong Wang, Hongbin Zha
TL;DR: We propose a new quantization method and apply it to quantize RNNs for both compression and acceleration
HIGHLIGHT: In this work, we address these problems by quantizing the network, both weights and activations, into multiple binary codes {-1,+1}.

280, TITLE: Residual Loss Prediction: Reinforcement Learning With No Incremental Feedback
https://openreview.net/forum?id=HJNMYceCW
AUTHORS: Hal Daum? III, John Langford, Amr Sharaf
TL;DR: We present a novel algorithm for solving reinforcement learning and bandit structured prediction problems with very sparse loss feedback.
HIGHLIGHT: We introduce a novel algorithm, RESIDUAL LOSS PREDICTION (RESLOPE), that solves such problems by automatically learning an internal representation of a denser reward function.

281, TITLE: Adaptive Quantization of Neural Networks
https://openreview.net/forum?id=SyOK1Sg0W
AUTHORS: Soroosh Khoram, Jing Li
TL;DR: An adaptive method for fixed-point quantization of neural networks based on theoretical analysis rather than heuristics.
HIGHLIGHT: We address these issues in this paper by proposing a new method, called adaptive quantization, which simplifies a trained DNN model by finding a unique, optimal precision for each network parameter such that the increase in loss is minimized.

282, TITLE: Boosting the Actor with Dual Critic
https://openreview.net/forum?id=BkUp6GZRW
AUTHORS: Bo Dai, Albert Shaw, Niao He, Lihong Li, Le Song
TL;DR: We propose Dual Actor-Critic algorithm, which is derived in a principled way from the Lagrangian dual form of the Bellman optimality equation. The algorithm achieves the state-of-the-art performances across several benchmarks.
HIGHLIGHT: This paper proposes a new actor-critic-style algorithm called Dual Actor-Critic or Dual-AC.

283, TITLE: Guide Actor-Critic for Continuous Control
https://openreview.net/forum?id=BJk59JZ0b
AUTHORS: Voot Tangkaratt, Abbas Abdolmaleki, Masashi Sugiyama

TL;DR: This paper proposes a novel actor-critic method that uses Hessians of a critic to update an actor.
HIGHLIGHT: In this paper, we propose a novel actor-critic method called the guide actor-critic (GAC).

284, TITLE: Policy Optimization by Genetic Distillation
https://openreview.net/forum?id=ByOnmlWC-
AUTHORS: Tanmay Gangwani, Jian Peng
TL;DR: Genetic algorithms based approach for optimizing deep neural network policies
HIGHLIGHT: Genetic algorithms have been widely used in many practical optimization problems.Inspired by natural selection, operators, including mutation, crossoverand selection, provide effective heuristics for search and black-box optimization.However, they have not been shown useful for deep reinforcement learning, possiblydue to the catastrophic consequence of parameter crossovers of neural networks.Here, we present Genetic Policy Optimization (GPO), a new genetic algorithmfor sample-efficient deep policy optimization.

285, TITLE: When is a Convolutional Filter Easy to Learn?
https://openreview.net/forum?id=SkA-IE06W
AUTHORS: Simon S. Du, Jason D. Lee, Yuandong Tian
TL;DR: We prove randomly initialized (stochastic) gradient descent learns a convolutional filter in polynomial time.
HIGHLIGHT: We analyze the convergence of (stochastic) gradient descent algorithm for learning a convolutional filter with Rectified Linear Unit (ReLU) activation function.

286, TITLE: Online Learning Rate Adaptation with Hypergradient Descent
https://openreview.net/forum?id=BkrsAzWAb
AUTHORS: Atilim Gunes Baydin, Robert Cornish, David Martinez Rubio, Mark Schmidt, Frank Wood
TL;DR: n/a
HIGHLIGHT: We introduce a general method for improving the convergence rate of gradient-based optimizers that is easy to implement and works well in practice.

287, TITLE: Stochastic gradient descent performs variational inference, converges to limit cycles for deep networks
https://openreview.net/forum?id=HyWrIgW0W
AUTHORS: Pratik Chaudhari, Stefano Soatto
TL;DR: SGD implicitly performs variational inference; gradient noise is highly non-isotropic, so SGD does not even converge to critical points of the original loss
HIGHLIGHT: We prove that SGD minimizes an average potential over the posterior distribution of weights along with an entropic regularization term.

288, TITLE: Robustness of Classifiers to Universal Perturbations: A Geometric Perspective
https://openreview.net/forum?id=ByrZyglCb
AUTHORS: Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, Pascal Frossard, Stefano Soatto
TL;DR: Analysis of vulnerability of classifiers to universal perturbations and relation to the curvature of the decision boundary.
HIGHLIGHT: In this paper, we provide a quantitative analysis of the robustness of classifiers to universal perturbations, and draw a formal link between the robustness to universal perturbations, and the geometry of the decision boundary.

289, TITLE: On the regularization of Wasserstein GANs
https://openreview.net/forum?id=B1hYRMbCW
AUTHORS: Henning Petzka, Asja Fischer, Denis Lukovnikov
TL;DR: A new regularization term can improve your training of wasserstein gans
HIGHLIGHT: We present theoretical arguments why using a weaker regularization term enforcing the Lipschitz constraint is preferable.

290, TITLE: Eigenoption Discovery through the Deep Successor Representation
https://openreview.net/forum?id=Bk8ZcAxR-

AUTHORS: Marlos C. Machado, Clemens Rosenbaum, Xiaoxiao Guo, Miao Liu, Gerald Tesauro, Murray Campbell
TL;DR: We show how we can use the successor representation to discover eigenoptions in stochastic domains, from raw pixels. Eigenoptions are options learned to navigate the latent dimensions of a learned representation.
HIGHLIGHT: We propose an algorithm that discovers eigenoptions while learning non-linear state representations from raw pixels.

291, TITLE: Neural Map: Structured Memory for Deep Reinforcement Learning
https://openreview.net/forum?id=Bk9zbyZCZ
AUTHORS: Emilio Parisotto, Ruslan Salakhutdinov
TL;DR: n/a
HIGHLIGHT: In this paper, we develop a memory system with an adaptable write operator that is customized to the sorts of 3D environments that DRL agents typically interact with.

292, TITLE: Active Neural Localization
https://openreview.net/forum?id=ry6-G_66b
AUTHORS: Devendra Singh Chaplot, Emilio Parisotto, Ruslan Salakhutdinov
TL;DR: "Active Neural Localizer", a fully differentiable neural network that learns to localize efficiently using deep reinforcement learning.
HIGHLIGHT: We propose "Active Neural Localizer", a fully differentiable neural network that learns to localize efficiently.

293, TITLE: Overcoming Catastrophic Interference using Conceptor-Aided Backpropagation
https://openreview.net/forum?id=B1al7jg0b
AUTHORS: Xu He, Herbert Jaeger
TL;DR: We propose a variant of the backpropagation algorithm, in which gradients are shielded by conceptors against degradation of previously learned tasks.
HIGHLIGHT: Catastrophic interference has been a major roadblock in the research of continual learning.

294, TITLE: Memory Augmented Control Networks
https://openreview.net/forum?id=HyfHgI6aW
AUTHORS: Arbaaz Khan, Clark Zhang, Nikolay Atanasov, Konstantinos Karydis, Vijay Kumar, Daniel D. Lee
TL;DR: Memory Augmented Network to plan in partially observable environments.
HIGHLIGHT: To mitigate these challenges we propose the Memory Augmented Control Network (MACN).

295, TITLE: Progressive Reinforcement Learning with Distillation for Multi-Skilled Motion Control
https://openreview.net/forum?id=B13njo1R-
AUTHORS: Glen Berseth, Cheng Xie, Paul Cernek, Michiel Van de Panne
TL;DR: A continual learning method that uses distillation to combine expert policies and transfer learning to accelerate learning new skills.
HIGHLIGHT: Progressive Reinforcement Learning with Distillation for Multi-Skilled Motion Control

296, TITLE: N2N learning: Network to Network Compression via Policy Gradient Reinforcement Learning
https://openreview.net/forum?id=B1hcZZ-AW
AUTHORS: Anubhav Ashok, Nicholas Rhinehart, Fares Beainy, Kris M. Kitani
TL;DR: A novel reinforcement learning based approach to compress deep neural networks with knowledge distillation
HIGHLIGHT: In this paper, we tackle this issue by introducing a principled method for learning reduced network architectures in a data-driven way using reinforcement learning.

297, TITLE: Hierarchical and Interpretable Skill Acquisition in Multi-task Reinforcement Learning
https://openreview.net/forum?id=SJJQVZW0b
AUTHORS: Tianmin Shu, Caiming Xiong, Richard Socher
TL;DR: A novel hierarchical policy network which can reuse previously learned skills alongside and as subcomponents of new skills by discovering the underlying relations between skills.

HIGHLIGHT: This paper proposes a novel framework for efficient multi-task reinforcement learning.

298, TITLE: Divide-and-Conquer Reinforcement Learning
https://openreview.net/forum?id=rJwelMbR-
AUTHORS: Dibya Ghosh, Avi Singh, Aravind Rajeswaran, Vikash Kumar, Sergey Levine
TL;DR: n/a
HIGHLIGHT: In this paper, we develop a novel algorithm that instead partitions the initial state space into "slices", and optimizes an ensemble of policies, each on a different slice.

299, TITLE: A Compressed Sensing View of Unsupervised Text Embeddings, Bag-of-n-Grams, and LSTMs
https://openreview.net/forum?id=B1e5ef-C-
AUTHORS: Sanjeev Arora, Mikhail Khodak, Nikunj Saunshi, Kiran Vodrahalli
TL;DR: We use the theory of compressed sensing to prove that LSTMs can do at least as well on linear text classification as Bag-of-n-Grams.
HIGHLIGHT: Using the theory of compressed sensing we show that representations combining the constituent word vectors are essentially information-preserving linear measurements of Bag-of-n-Grams (BonG) representations of text.

300, TITLE: A New Method of Region Embedding for Text Classification
https://openreview.net/forum?id=BkSDMA36Z
AUTHORS: chao qiao, bo huang, guocheng niu, daren li, daxiang dong, wei he, dianhai yu, hua wu
TL;DR: n/a
HIGHLIGHT: In this paper, we propose a new method of learning and utilizing task-specific distributed representations of n-grams, referred to as ?region embeddings?.

301, TITLE: Fix your classifier: the marginal value of training the last weight layer
https://openreview.net/forum?id=S1Dh8Tg0-
AUTHORS: Elad Hoffer, Itay Hubara, Daniel Soudry
TL;DR: You can fix the classifier in neural networks without losing accuracy
HIGHLIGHT: This classifier can have a vast number of parameters, which grows linearly with the number of possible classes, thus requiring increasingly more resources.In this work we argue that this classifier can be fixed, up to a global scale constant, with little or no loss of accuracy for most tasks, allowing memory and computational benefits.

302, TITLE: Multi-Mention Learning for Reading Comprehension with Neural Cascades
https://openreview.net/forum?id=HyRnez-RW
AUTHORS: Swabha Swayamdipta, Ankur P. Parikh, Tom Kwiatkowski
TL;DR: We propose neural cascades, a simple and trivially parallelizable approach to reading comprehension, consisting only of feed-forward nets and attention that achieves state-of-the-art performance on the TriviaQA dataset.
HIGHLIGHT: In this work, we take a different approach by constructing lightweight models that are combined in a cascade to find the answer.

303, TITLE: Deep Sensing: Active Sensing using Multi-directional Recurrent Neural Networks
https://openreview.net/forum?id=r1SnX5xCb
AUTHORS: Jinsung Yoon, William R. Zame, Mihaela van der Schaar
TL;DR: n/a
HIGHLIGHT: To solve the problem of active sensing we develop a novel deep learning architecture: Deep Sensing.

304, TITLE: Temporally Efficient Deep Learning with Spikes
https://openreview.net/forum?id=HkZy-bW0-
AUTHORS: Peter O'Connor, Efstratios Gavves, Matthias Reisser, Max Welling
TL;DR: An algorithm for training neural networks efficiently on temporally redundant data.
HIGHLIGHT: We present a variant on backpropagation for neural networks in which computation scales with the rate of change of the data - not the rate at which we process the data.

305, TITLE: Variational Network Quantization
https://openreview.net/forum?id=ry-TW-WAb
AUTHORS: Jan Achterhold, Jan Mathias Koehler, Anke Schmeink, Tim Genewein
TL;DR: We quantize and prune neural network weights using variational Bayesian inference with a multi-modal, sparsity inducing prior.
HIGHLIGHT: In this paper, the preparation of a neural network for pruning and few-bit quantization is formulated as a variational inference problem.

306, TITLE: Training GANs with Optimism
https://openreview.net/forum?id=SJJySbbAZ
AUTHORS: Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, Haoyang Zeng
TL;DR: We propose the use of optimistic mirror decent to address cycling problems in the training of GANs. We also introduce the Optimistic Adam algorithm
HIGHLIGHT: We address the issue of limit cycling behavior in training Generative Adversarial Networks and propose the use of Optimistic Mirror Decent (OMD) for training Wasserstein GANs.

307, TITLE: Sobolev GAN
https://openreview.net/forum?id=SJA7xfb0b
AUTHORS: Youssef Mroueh, Chun-Liang Li, Tom Sercu, Anant Raj, Yu Cheng
TL;DR: We define a new Integral Probability Metric (Sobolev IPM) and show how it can be used for training GANs for text generation and semi-supervised learning.
HIGHLIGHT: We propose a new Integral Probability Metric (IPM) between distributions: the Sobolev IPM.

308, TITLE: Learning From Noisy Singly-labeled Data
https://openreview.net/forum?id=H1sUHgb0Z
AUTHORS: Ashish Khetan, Zachary C. Lipton, Animashree Anandkumar
TL;DR: A new approach for learning a model from noisy crowdsourced annotations.
HIGHLIGHT: We propose a new algorithm for jointly modeling labels and worker quality from noisy crowd-sourced data.

309, TITLE: Learning Sparse Neural Networks through L_0 Regularization
https://openreview.net/forum?id=H1Y8hhg0b
AUTHORS: Christos Louizos, Max Welling, Diederik P. Kingma
TL;DR: We show how to optimize the expected L_0 norm of parametric models with gradient descent and introduce a new distribution that facilitates hard gating.
HIGHLIGHT: We propose a practical method for L_0 norm regularization for neural networks: pruning the network during training by encouraging weights to become exactly zero.

310, TITLE: Variational Continual Learning
https://openreview.net/forum?id=BkQqq0gRb
AUTHORS: Cuong V. Nguyen, Yingzhen Li, Thang D. Bui, Richard E. Turner
TL;DR: This paper develops a principled method for continual learning in deep models.
HIGHLIGHT: This paper develops variational continual learning (VCL), a simple but general framework for continual learning that fuses online variational inference (VI) and recent advances in Monte Carlo VI for neural networks.

311, TITLE: Gaussian Process Behaviour in Wide Deep Neural Networks
https://openreview.net/forum?id=H1-nGgWC-
AUTHORS: Alexander G. de G. Matthews, Jiri Hron, Mark Rowland, Richard E. Turner, Zoubin Ghahramani
TL;DR: n/a
HIGHLIGHT: In this paper, we study the relationship between Gaussian processes with a recursive kernel definition and random wide fully connected feedforward networks with more than one hidden layer.

312, TITLE: Mixed Precision Training of Convolutional Neural Networks using Integer Operations

https://openreview.net/forum?id=H135uzZ0-
AUTHORS: Dipankar Das, Naveen Mellempudi, Dheevatsa Mudigere, Dhiraj Kalamkar, Sasikanth Avancha, Kunal Banerjee, Srinivas Sridharan, Karthik Vaidyanathan, Bharat Kaul, Evangelos Georganas, Alexander Heinecke, Pradeep Dubey, Jesus Corbal, Nikita Shustrov, Roma Dubtsov, Evarist Fomenko, Vadim Pirogov
TL;DR: Mixed precision training pipeline using 16-bit integers on general purpose HW; SOTA accuracy for ImageNet-class CNNs; Best reported accuracy for ImageNet-1K classification task with any reduced precision training;
HIGHLIGHT: In this work, we train state-of-the-art visual understanding neural networks on the ImageNet-1K dataset, with Integer operations on General Purpose (GP) hardware.

313, TITLE: Memory Architectures in Recurrent Neural Network Language Models
https://openreview.net/forum?id=SkFqf0lAZ
AUTHORS: Dani Yogatama, Yishu Miao, Gabor Melis, Wang Ling, Adhiguna Kuncoro, Chris Dyer, Phil Blunsom
TL;DR: n/a
HIGHLIGHT: We compare and analyze sequential, random access, and stack memory architectures for recurrent neural network language models.

314, TITLE: On the Information Bottleneck Theory of Deep Learning
https://openreview.net/forum?id=ry_WPG-A-
AUTHORS: Andrew Michael Saxe, Yamini Bansal, Joel Dapello, Madhu Advani, Artemy Kolchinsky, Brendan Daniel Tracey, David Daniel Cox
TL;DR: We show that several claims of the information bottleneck theory of deep learning are not true in the general case.
HIGHLIGHT: In this work, we study the information bottleneck (IB) theory of deep learning, which makes three specific claims: first, that deep networks undergo two distinct phases consisting of an initial fitting phase and a subsequent compression phase; second, that the compression phase is causally related to the excellent generalization performance of deep networks; and third, that the compression phase occurs due to the diffusion-like behavior of stochastic gradient descent.