1, TITLE: On the Convergence of Adam and Beyond
https://openreview.net/forum?id=ryQu7f-RZ
AUTHORS: Sashank J. Reddi, Satyen Kale, Sanjiv Kumar
TL;DR: We investigate the convergence of popular optimization algorithms like Adam , RMSProp and propose new variants of these methods which provably converge to optimal solution in convex settings.
HIGHLIGHT: We provide an explicit example of a simple convex optimization setting where Adam does not converge to the optimal solution, and describe the precise problems with the previous analysis of Adam algorithm.

2, TITLE: Synthetic and Natural Noise Both Break Neural Machine Translation
https://openreview.net/forum?id=BJ8vJebC-
AUTHORS: Yonatan Belinkov, Yonatan Bisk
TL;DR: CharNMT is brittle
HIGHLIGHT: In this paper, we confront NMT models with synthetic and natural sources of noise.

3, TITLE: Multi-Scale Dense Networks for Resource Efficient Image Classification
https://openreview.net/forum?id=Hk2aImxAb
AUTHORS: Gao Huang, Danlu Chen, Tianhong Li, Felix Wu, Laurens van der Maaten, Kilian Weinberger
TL;DR: n/a
HIGHLIGHT: In this paper we investigate image classification with computational resource limits at test time.

4, TITLE: Training and Inference with Integers in Deep Neural Networks
https://openreview.net/forum?id=HJGXzmspb
AUTHORS: Shuang Wu, Guoqi Li, Feng Chen, Luping Shi
TL;DR: We apply training and inference with only low-bitwidth integers in DNNs
HIGHLIGHT: In this work, we develop a new method termed as ``"WAGE" to discretize both training and inference, where weights (W), activations (A), gradients (G) and errors (E) among layers are shifted and linearly constrained to low-bitwidth integers.

5, TITLE: Emergence of Linguistic Communication from Referential Games with Symbolic and Pixel Input
https://openreview.net/forum?id=HJGv1Z-AW
AUTHORS: Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, Stephen Clark
TL;DR: A controlled study of the role of environments with respect to properties in emergent communication protocols.
HIGHLIGHT: Here we scale up this research by using contemporary deep learning methods and by training reinforcement-learning neural network agents on referential communication games.

6, TITLE: Spherical CNNs
https://openreview.net/forum?id=Hkbd5xZRb
AUTHORS: Taco S. Cohen, Mario Geiger, Jonas Kï¿½hler, Max Welling
TL;DR: We introduce Spherical CNNs, a convolutional network for spherical signals, and apply it to 3D model recognition and molecular energy regression.
HIGHLIGHT: A naive application of convolutional networks to a planar projection of the spherical signal is destined to fail, because the space-varying distortions introduced by such a projection will make translational weight sharing ineffective.In this paper we introduce the building blocks for constructing spherical CNNs.

7, TITLE: Ask the Right Questions: Active Question Reformulation with Reinforcement Learning
https://openreview.net/forum?id=S1CChZ-CZ
AUTHORS: Christian Buck, Jannis Bulian, Massimiliano Ciaramita, Wojciech Gajewski, Andrea Gesmundo, Neil Houlsby, Wei Wang.
TL;DR: We propose an agent that sits between the user and a black box question-answering system and which learns to reformulate questions to elicit the best possible answers
HIGHLIGHT: We frame Question Answering (QA) as a Reinforcement Learning task, an approach that we call Active Question Answering.We propose an agent that sits between the user and a black box QA system and learns to reformulate questions to elicit the best possible answers.

8, TITLE:	On the insufficiency of existing momentum schemes for Stochastic Optimization
https://openreview.net/forum?id=rJTutzbA-
AUTHORS:	Rahul Kidambi, Praneeth Netrapalli, Prateek Jain, Sham M. Kakade
TL;DR:	Existing momentum/acceleration schemes such as heavy ball method and Nesterov's acceleration employed with stochastic gradients do not improve over vanilla stochastic gradient descent, especially when employed with small batch sizes.
HIGHLIGHT:	Extensive empirical results in this paper show that ASGD has performance gains over HB, NAG, and SGD.

9, TITLE:	Certifying Some Distributional Robustness with Principled Adversarial Training
https://openreview.net/forum?id=Hk6kPgZA-
AUTHORS:	Aman Sinha, Hongseok Namkoong, John Duchi
TL;DR:	We provide a fast, principled adversarial training procedure with computational and statistical performance guarantees.
HIGHLIGHT:	We address this problem through the principled lens of distributionally robust optimization, which guarantees performance under adversarial input perturbations.

10, TITLE:	Learning Deep Mean Field Games for Modeling Large Population Behavior
https://openreview.net/forum?id=HktK4BeCZ
AUTHORS:	Jiachen Yang, Xiaojing Ye, Rakshit Trivedi, Huan Xu, Hongyuan Zha
TL;DR:	Inference of a mean field game (MFG) model of large population behavior via a synthesis of MFG and Markov decision processes.
HIGHLIGHT:	We consider the problem of representing collective behavior of large populations and predicting the evolution of a population distribution over a discrete state space.

11, TITLE:	Wasserstein Auto-Encoders
https://openreview.net/forum?id=HkL7n1-0b
AUTHORS:	Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, Bernhard Schoelkopf
TL;DR:	We propose a new auto-encoder based on the Wasserstein distance, which improves on the sampling properties of VAE.
HIGHLIGHT:	We propose the Wasserstein Auto-Encoder (WAE)---a new algorithm for building a generative model of the data distribution.

12, TITLE:	Spectral Normalization for Generative Adversarial Networks
https://openreview.net/forum?id=B1QRgziT-
AUTHORS:	Takeru Miyato, Toshiki Kataoka, Masanori Koyama, Yuichi Yoshida
TL;DR:	We propose a novel weight normalization technique called spectral normalization to stabilize the training of the discriminator of GANs.
HIGHLIGHT:	One of the challenges in the study of generative adversarial networks is the instability of its training.In this paper, we propose a novel weight normalization technique called spectral normalization to stabilize the training of the discriminator.Our new normalization technique is computationally light and easy to incorporate into existing implementations.We tested the efficacy of spectral normalization on CIFAR10, STL-10, and ILSVRC2012 dataset, and we experimentally confirmed that spectrally normalized GANs (SN-GANs) is capable of generating images of better or equal quality relative to the previous training stabilization techniques.

13, TITLE:	Learning to Represent Programs with Graphs
https://openreview.net/forum?id=BJOFETxR-
AUTHORS:	Miltiadis Allamanis, Marc Brockschmidt, Mahmoud Khademi
TL;DR:	Programs have structure that can be represented as graphs, and graph neural networks can learn to find bugs on such graphs
HIGHLIGHT:	We propose to use graphs to represent both the syntactic and semantic structure of code and use graph-based deep learning methods to learn to reason over program structures.In this work, we present how to construct graphs from source code and how to scale Gated Graph Neural Networks training to such large graphs.

14, TITLE:	Characterizing Adversarial Subspaces Using Local Intrinsic Dimensionality
https://openreview.net/forum?id=B1gJ1L2aW

AUTHORS: Xingjun Ma, Bo Li, Yisen Wang, Sarah M. Erfani, Sudanthi Wijewickrema, Grant Schoenebeck, Dawn Song, Michael E. Houle, James Bailey
TL;DR: We characterize the dimensional properties of adversarial subspaces in the neighborhood of adversarial examples via the use of Local Intrinsic Dimensionality (LID).
HIGHLIGHT: As a proof-of-concept, we show that a potential application of LID is to distinguish adversarial examples, and the preliminary results show that it can outperform several state-of-the-art detection measures by large margins for five attack strategies considered in this paper across three benchmark datasets.

15, TITLE: Breaking the Softmax Bottleneck: A High-Rank RNN Language Model
https://openreview.net/forum?id=HkwZSG-CZ
AUTHORS: Zhilin Yang, Zihang Dai, Ruslan Salakhutdinov, William W. Cohen
TL;DR: n/a
HIGHLIGHT: We propose a simple and effective method to address this issue, and improve the state-of-the-art perplexities on Penn Treebank and WikiText-2 to 47.69 and 40.68 respectively.

16, TITLE: Continuous Adaptation via Meta-Learning in Nonstationary and Competitive Environments
https://openreview.net/forum?id=Sk2u1g-0-
AUTHORS: Maruan Al-Shedivat, Trapit Bansal, Yura Burda, Ilya Sutskever, Igor Mordatch, Pieter Abbeel
TL;DR: n/a
HIGHLIGHT: In this paper, we cast the problem of continuous adaptation into the learning-to-learn framework.

17, TITLE: Boosting Dilated Convolutional Networks with Mixed Tensor Decompositions
https://openreview.net/forum?id=S1JHhv6TW
AUTHORS: Nadav Cohen, Ronen Tamari, Amnon Shashua
TL;DR: We introduce the notion of mixed tensor decompositions, and use it to prove that interconnecting dilated convolutional networks boosts their expressive power.
HIGHLIGHT: In this paper we study the expressive efficiency brought forth by connectivity, motivated by the observation that modern networks interconnect their layers in elaborate ways.

18, TITLE: Neural Sketch Learning for Conditional Program Generation
https://openreview.net/forum?id=HkfXMz-Ab
AUTHORS: Vijayaraghavan Murali, Letao Qi, Swarat Chaudhuri, Chris Jermaine
TL;DR: We give a method for generating type-safe programs in a Java-like language, given a small amount of syntactic information about the desired code.
HIGHLIGHT: We study the problem of generating source code in a strongly typed,Java-like programming language, given a label (for example a set ofAPI calls or types) carrying a small amount of information about thecode that is desired.

19, TITLE: Progressive Growing of GANs for Improved Quality, Stability, and Variation
https://openreview.net/forum?id=Hk99zCeAb
AUTHORS: Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen
TL;DR: We train generative adversarial networks in a progressive fashion, enabling us to generate high-resolution images with high quality.
HIGHLIGHT: We describe a new training methodology for generative adversarial networks.

20, TITLE: Variance Reduction for Policy Gradient with Action-Dependent Factorized Baselines
https://openreview.net/forum?id=H1tSsb-AW
AUTHORS: Cathy Wu, Aravind Rajeswaran, Yan Duan, Vikash Kumar, Alexandre M Bayen, Sham Kakade, Igor Mordatch, Pieter Abbeel
TL;DR: Action-dependent baselines can be bias-free and yield greater variance reduction than state-only dependent baselines for policy gradient methods.
HIGHLIGHT: To mitigate this issue, we derive a bias-free action-dependent baseline for variance reduction which fully exploits the structural form of the stochastic policy itself and does not make any additional assumptions about the MDP.

21, TITLE: Zero-Shot Visual Imitation
https://openreview.net/forum?id=BkisuzWRW
AUTHORS: Deepak Pathak, Parsa Mahmoudieh, Guanghao Luo, Pulkit Agrawal, Dian Chen, Yide Shentu, Evan Shelhamer, Jitendra Malik, Alexei A. Efros, Trevor Darrell
TL;DR: Agents can learn to imitate solely visual demonstrations (without actions) at test time after learning from their own experience without any form of supervision at training time.
HIGHLIGHT: We evaluate our zero-shot imitator in two real-world settings: complex rope manipulation with a Baxter robot and navigation in previously unseen office environments with a TurtleBot.

22, TITLE: Beyond Word Importance: Contextual Decomposition to Extract Interactions from LSTMs
https://openreview.net/forum?id=rkRwGg-0Z
AUTHORS: W. James Murdoch, Peter J. Liu, Bin Yu
TL;DR: We introduce contextual decompositions, an interpretation algorithm for LSTMs capable of extracting word, phrase and interaction-level importance score
HIGHLIGHT: To this end, we introduce contextual decomposition (CD), an interpretation algorithm for analysing individual predictions made by standard LSTMs, without any changes to the underlying model.

23, TITLE: AmbientGAN: Generative models from lossy measurements
https://openreview.net/forum?id=Hy7fDog0b
AUTHORS: Ashish Bora, Eric Price, Alexandros G. Dimakis
TL;DR: How to learn GANs from noisy, distorted, partial observations
HIGHLIGHT: Based on this, we propose a new method of training Generative Adversarial Networks (GANs) which we call AmbientGAN.