Abstract Analysis of Academic Papers

Qingxin Meng

No. 2018312032

Qingxin6174@gmail.com

College of Petroleum Engineering

China University of Petroleum, Beijing

December 10, 2018

Outline

Google Scholar — Top 20 Publications

in All Categories

	Publication	h5-index	h5-median
1.	Nature	362	542
2.	The New England Journal of Medicine	358	602
3.	Science	345	497
4.	The Lancet	278	417
5.	Chemical Society reviews	256	366
6.	Cell	244	366
7.	Nature Communications	240	318
8.	Chemical Reviews	239	373
9.	Journal of the American Chemical Society	236	309
10.	Advanced Materials	235	336
11.	Proceedings of the National Academy of Sciences	226	291
12.	Angewandte Chemie International Edition	213	295
13.	JAMA	209	309
14.	Nucleic Acids Research	208	392
15.	ACS Nano	199	279
16.	Physical Review Letters	197	286
17.	Energy and Environmental Science	196	330
18.	Journal of Clinical Oncology	196	279
19.	Nano Letters	194	281
20.	IEEE Conference on Computer Vision and Pattern Recognition, CVPR	188	302

Dates and citation counts are estimated and are determined automatically by a computer program.

Google Scholar — Top 20 Publications

in Engineering & Computer Science > Computer Vision & Pattern Recognition

	Publication	h5-index	h5-median
1.	IEEE Conference on Computer Vision and Pattern Recognition, CVPR	188	302
2.	IEEE International Conference on Computer Vision	124	204
3.	IEEE Transactions on Pattern Analysis and Machine Intelligence	118	210
4.	European Conference on Computer Vision	104	180
5.	IEEE Transactions on Image Processing	101	150
6.	Pattern Recognition	74	97
7.	International Journal of Computer Vision	65	124
8.	Medical Image Analysis	57	84
9.	Pattern Recognition Letters	51	76
10.	IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops	47	83
11.	Computer Vision and Image Understanding	44	70
12.	British Machine Vision Conference (BMVC)	42	71
13.	Journal of Visual Communication and Image Representation	42	64
14.	IEEE International Conference on Image Processing (ICIP)	41	51
15.	Workshop on Applications of Computer Vision (WACV)	38	70
16.	IEEE International Conference on Automatic Face & Gesture Recognition	36	63
17.	Image and Vision Computing	36	58
18.	Machine Vision and Applications	33	50
19.	International Conference on Document Analysis and Recognition	33	47
20.	Journal of Mathematical Imaging and Vision	31	42

Dates and citation counts are estimated and are determined automatically by a computer program.

Google Scholar — Top 20 Publications

in Engineering & Computer Science > Artificial Intelligence

	Publication	h5-index	h5-median
1.	Neural Information Processing Systems (NIPS)	134	221
2.	International Conference on Machine Learning (ICML)	113	193
3.	Expert Systems with Applications	92	133
4.	IEEE Transactions On Systems, Man And Cybernetics Part B, Cybernetics	88	129
5.	IEEE Transactions on Neural Networks and Learning Systems	87	117
6.	The Journal of Machine Learning Research	79	116
7.	Applied Soft Computing	77	102
8.	IEEE Transactions on Fuzzy Systems	74	132
9.	Knowledge-Based Systems	74	110
10.	Neurocomputing	71	95
11.	AAAI Conference on Artificial Intelligence	69	101
12.	International Joint Conference on Artificial Intelligence (IJCAI)	61	98
13.	Neural Networks	56	80
14.	Neural Computing and Applications	54	77
15.	Engineering Applications of Artificial Intelligence	50	68
16.	Robotics and Autonomous Systems	49	78
17.	Artificial Intelligence	46	71
18.	Conference on Learning Theory (COLT)	44	60
19.	International Conference on Artificial Intelligence and Statistics	43	59
20.	Journal of Intelligent & Robotic Systems	42	53

Dates and citation counts are estimated and are determined automatically by a computer program.

Guide2Research — Top 25 Scholars

for Computer Science and Electronics

	Scholar	Institution	Country	Citations	H-index
1	Anil K. Jain	Michigan State University	United States	184991	179
2	Herbert Simon	Carnegie Mellon University	United States	334412	175
3	Jiawei Han	University of Illinois	United States	158874	162
4	David Haussler	University of California, Santa Cruz	United States	175427	151
5	Takeo Kanade	Carnegie Mellon University	United States	132137	151
6	Terrence Sejnowski	Salk Institute	United States	125404	151
7	Michael I. Jordan	University of California, Berkeley	United States	143534	148
8	Philip S. Yu	University of Illinois at Chicago	United States	101761	148
9	Scott Shenker	University of California, Berkeley	United States	131074	146
10	Andrew Zisserman	University of Oxford	United Kingdom	147339	144
11	Thomas S. Huang	University of Illinois	United States	105157	144
12	Guanrong Chen	City University of Hong Kong	Hong Kong	91624	143
13	Sebastian Thrun	Stanford University	United States	92971	141
14	Geoffrey Hinton	University of Toronto	Canada	248416	140
15	Wil van der Aalst	RWTH Aachen University	Germany	89287	140
16	TOMASO POGGIO	MIT	United States	104083	134
17	De Moor Bart	University of Leuven	Belgium	99805	134
18	Héctor García-Molina	Stanford University	United States	78770	133
19	Georgios B. Giannakis	University of Minnesota	United States	67999	133
20	Daphne Koller	Stanford University	United States	71985	130
21	Yoshua Bengio	University of Montreal	Canada	141755	129
22	Alex Sandy Pentland	MIT	United States	110627	129
23	Eric Horvitz	Microsoft Research	United States	61978	129
24	Ian Foster	University of Chicago	United States	113298	127
25	Jack Dongarra	University of Tennessee	United States	109606	127

onalytica — Top 25 Influencers

for Artificial intelligence

	Name	Company	Influencer Score
1	David Brin	Epocene	100
2	Bernard Marr	Advanced Performance Institute	90.85
3	Andrew Ng	Coursera & Stanford University	82.21
4	Fei-Fei- Li	Google Cloud	62.79
5	Vala Afshar	Salesforce	61.3
6	Mike Quindazzi	PwC	58.56
7	Elon Musk	SpaceX	58.35
8	Tamara McCleary	Thulium.co	49.58
9	Jack Clark	OpenAI	49
10	Marc Benioff	Salesforce	48.32
11	Evan Kirstel	Kirstel Report	47.04
12	John Hagel	Deloitte & Touche	46.55
13	Eric Horvitz	Microsoft	46.27
14	Gary Marcus	Geometric Intelligence	43.95
15	Jeremiah Owyang	Crowd Companies	41.51
16	Pedro Domingos	University of Washington	40.95
17	Sandy Carter	Amazon Web Services	38.63
18	Eric Topol	Scripps Health	35.15
19	Eric Schmidt	Google	34.63
20	Hanna Wallach	Microsoft	33
21	Spiros Margaris	Margaris Advisory	32.29
22	Peter Lee	Microsoft	32.1
23	Dharmesh Shah	HubSpot	31.54
24	Erik Brynjolfsson	MIT	30.82
25	Brian Solis	Altimeter Group	30.64

20 Samples

Title	Publication	Year
A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play	Science	2018
Mastering the game of Go without human knowledge	Nature	2017
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	arXiv	2018
Reading China: Predicting policy change with machine learning	AEI	2018
Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks	arXiv	2017
Nearly tight sample complexity bounds for learning mixtures of Gaussians via sample compression schemes	NeurIPS⋆	2018
Neural Ordinary Differential Equations	NeurIPS⋆	2018
Optimal Algorithms for Non-Smooth Distributed Optimization in Networks	NeurIPS⋆	2018
Non-delusional Q-learning and value-iteration	NeurIPS⋆	2018
Exploring Generalization in Deep Learning	NeurIPS	2018
Video-to-Video Synthesis	NeurIPS	2018
Safe and Nested Subgame Solving for Imperfect-Information Games	NIPS⋆	2017
A Linear-Time Kernel Goodness-of-Fit Test	NIPS⋆	2017
Variance-based Regularization with Convex Objectives	NIPS⋆	2017
Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images	CVPR	2015
Dance Dance Convolution	ICML	2017
Don't Decay the Learning Rate, Increase the Batch Size	ICLR	2018
Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks	ICLR	2016
U-Net: Convolutional Networks for Biomedical Image Segmentation	MICCAI	2015
Deep Learning: A Bayesian Perspective	ISBA	2017

Mastering the game of Go without human knowledge

David Silver✉, **Julian Schrittwieser**, **Karen Simonyan**, **Ioannis Antonoglou**, **Aja Huang**, **Arthur Guez**, **Thomas Hubert**, **Lucas Baker**, **Matthew Lai**, **Adrian Bolton**, **Yutian Chen**, **Timothy Lillicrap**, **Fan Hui**, **Laurent Sifre**, **George van den Driessche**, **Thore Graepel** & **Demis Hassabis**
DeepMind, 5 New Street Square, London EC4A 3TW, UK.
Nature volume 550, pages 354 359 (19 October 2017) | *doi:10.1038/nature24270*

Abstract — A long-standing goal of artificial intelligence is an algorithm that learns, *tabula rasa*, superhuman proficiency in challenging domains. Recently, AlphaGo became the first program to defeat a world champion in the game of Go. The tree search in AlphaGo evaluated positions and selected moves using deep neural networks. These neural networks were trained by supervised learning from human expert moves, and by reinforcement learning from self-play.[B] Here we introduce an algorithm based solely on reinforcement learning, without human data, guidance or domain knowledge beyond game rules.[I] AlphaGo becomes its own teacher: a neural network is trained to predict AlphaGo's own move selections and also the winner of AlphaGo's games. This neural network improves the strength of the tree search, resulting in higher quality move selection and stronger self-play in the next iteration.[M] Starting *tabula rasa*, our new program AlphaGo Zero achieved superhuman performance, winning 100 0 against the previously published, champion-defeating AlphaGo.[R]

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin **Ming-Wei Chang** **Kenton Lee** **Kristina Toutanova**
Google AI Language
Submitted on 11 Oct 2018 | *arXiv:1810.04805v1* [cs.CL]

Abstract — We introduce a new language representation model called **BERT**, which stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers.[I] Unlike recent language representation models (Peters et al., 2018; Radford et al., 2018), BERT is designed to pre-train deep bidirectional representations by jointly conditioning on both left and right context in all layers.[M] As a result, the pre-trained BERT representations can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, *without* substantial task-specific architecture modifications.[C]
BERT is conceptually simple and empirically powerful. It obtains new state-of-the-art results on eleven natural language processing tasks, including pushing the GLUE benchmark to **80.4%** (**7.6%** absolute improvement), MultiNLI accuracy to **86.7%** (**5.6%** absolute improvement) and the SQuAD v1.1 question answering Test F1 to **93.2%** (**1.5%** absolute improvement), outperforming human performance by **2.0%**.[R]

Reading China: Predicting policy change with machine learning

Julian TszKin Chan — Bates White Economic Consulting
Weifeng Zhong — American Enterprise Institute

Submitted on 22 Oct 2018, last revised 15 Nov 2018 | *AEI Economics Working Paper Series*
Project website at `https://policychangeindex.com/`

Abstract — For the first time in the literature, we develop a quantitative indicator of the Chinese government's policy priorities over a long period of time, which we call the Policy Change Index (PCI) for China. The PCI is a leading indicator of policy changes that covers the period from 1951 to the third quarter of 2018, and it can be updated in the future.[I] It is designed with two building blocks: the full text of the *People's Daily* —the official newspaper of the Communist Party of China —as input data and a set of machine learning techniques to detect changes in how this newspaper prioritizes policy issues.[M] Due to the unique role of the *People's Daily* in China's propaganda system, detecting changes in this newspaper allows us to predict changes in China's policies.[R] The construction of the PCI does not require the understanding of the Chinese text, which suggests a wide range of applications in other settings, such as predicting changes in other (ex-)Communist regimes' policies, measuring decentralization in central-local government relations, quantifying media bias in democratic countries, and predicting changes in lawmakers' voting behavior and in judges' ideological leaning.[C]

Nearly tight sample complexity bounds for learning mixtures of Gaussians via sample compression schemes

Hassan Ashtiani
Department of Computing and Software
McMaster University, and
Vector Institute, ON, Canada

Shai Ben-David
School of Computer Science
University of Waterloo
Waterloo, ON, Canada

Nicholas J. A. Harvey
Department of Computer Science
University of British Columbia
Vancouver, BC, Canada

Christopher Liaw
Department of Computer Science
University of British Columbia
Vancouver, BC, Canada

Abbas Mehrabian
School of Computer Science
McGill University
Montréal, QC, Canada

Yaniv Plan
Department of Mathematics
University of British Columbia
Vancouver, BC, Canada

32nd Conference on Neural Information Processing Systems (NeurIPS 2018 Best Paper), Montréal, Canada.

Abstract — We prove that $\tilde{\Theta}(kd^2/\varepsilon^2)$ samples are necessary and sufficient for learning a mixture of k Gaussians in $\mathbb{R}^d$, up to error ε in total variation distance. This improves both the known upper bounds and lower bounds for this problem. For mixtures of axis-aligned Gaussians, we show that $\tilde{O}(kd/\varepsilon^2)$ samples suffice, matching a known lower bound.

The upper bound is based on a novel technique for distribution learning based on a notion of *sample compression*. Any class of distributions that allows such a sample compression scheme can also be learned with few samples. Moreover, if a class of distributions has such a compression scheme, then so do the classes of *products* and *mixtures* of those distributions.[I] The core of our main result is showing that the class of Gaussians in $\mathbb{R}^d$ has an efficient sample compression.[R]

通过样本压缩方案 学习混合高斯模型的近乎紧密的 样本复杂性边界

Hassan Ashtiani
Department of Computing and Software
McMaster University, and
Vector Institute, ON, Canada

Shai Ben-David
School of Computer Science
University of Waterloo
Waterloo, ON, Canada

Nicholas J. A. Harvey
Department of Computer Science
University of British Columbia
Vancouver, BC, Canada

Christopher Liaw
Department of Computer Science
University of British Columbia
Vancouver, BC, Canada

Abbas Mehrabian
School of Computer Science
McGill University
Montréal, QC, Canada

Yaniv Plan
Department of Mathematics
University of British Columbia
Vancouver, BC, Canada

32nd Conference on Neural Information Processing Systems (NeurIPS 2018 Best Paper), Montréal, Canada.

摘要译文 —— 我们证明了 $\tilde{\Theta}(kd^2/\varepsilon^2)$ 个样本对于学习 $\mathbb{R}^d$ 中的 k 个高斯混合模型是充分必要的, 这里我们允许总变差距离误差在 ε 以内. 这一结果对该问题的已知上限和下限作出了改进. 对于轴对齐高斯分布的混合, 我们证明 $\tilde{O}(kd/\varepsilon^2)$ 个样本就足够了, 这与已知下限相符.
对于上限, 我们提出了一种基于样本压缩概念的分布式学习的新技术. 对于允许这种样本压缩方案的任何分布, 我们都可以用很少的样本来学习. 此外, 如果一类分布具有这样的压缩方案, 那么这些分布的乘积或混合也同样可以运用这种方案. I 我们主要结果是表明了 $\mathbb{R}^d$ 中的高斯类具有有效的样本压缩. R

Results

Title	Publication	B	I	M	R	C
A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play	Science	√	√		√	
Mastering the game of Go without human knowledge	Nature	√	√	√	√	
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	arXiv		√	√	√	√
Reading China: Predicting policy change with machine learning	AEI		√	√	√	√
Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks	arXiv		√	√	√	
Nearly tight sample complexity bounds for learning mixtures of Gaussians via sample compression schemes	NeurIPS⋆		√		√	
Neural Ordinary Differential Equations	NeurIPS⋆		√			
Optimal Algorithms for Non-Smooth Distributed Optimization in Networks	NeurIPS⋆		√			
Non-delusional Q-learning and value-iteration	NeurIPS⋆	√	√			
Exploring Generalization in Deep Learning	NeurIPS		√			
Video-to-Video Synthesis	NeurIPS	√	√	√	√	√
Safe and Nested Subgame Solving for Imperfect-Information Games	NIPS⋆	√	√			
A Linear-Time Kernel Goodness-of-Fit Test	NIPS⋆		√	√	√	√
Variance-based Regularization with Convex Objectives	NIPS⋆		√	√	√	
Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images	CVPR	√	√	√	√	√
Dance Dance Convolution	ICML	√	√	√		
Don't Decay the Learning Rate, Increase the Batch Size	ICLR	√	√	√	√	
Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks	ICLR	√	√		√	
U-Net: Convolutional Networks for Biomedical Image Segmentation	MICCAI	√	√	√	√	
Deep Learning: A Bayesian Perspective	ISBA	√	√		√	√

Conclusions

- 语义逻辑最重要,语言表达是锦上添花.
- BIMRC 结构适用于算法应用类文章,
  纯粹算法的构造/证明倾向以 Introduction 为主.
- 名词化结构不多.
- 从句结构很多.

A New Edge Detection Method Based on Hausdorff Distance

You Lu **Qingxin Meng** **Jiangtao Guo**
School of Computer Science and Technology, China University of Petroleum, Beijing
Date Added to IEEE *Xplore*: 29 December 2014 | *DOI: 10.1109/ICDH.2014.15*
5^{th} *International Conference on Digital Home (ICDH 2014), ©IEEE*

Abstract — Most edge detection methods are based on first-order or second-order differential. These are local methods. Using Hausdorff distance to quantify the strength of the edge is a method with a holistic property. Firstly, down sample the image, and split the image into two sets. Secondly, get the feature image by assigning a value for each point using the scalar field map constructed by Hausdorff distance. Since the edge features have local properties, in this paper, we constructed a map which can get local feature images using sub image and combined them into a feature image. Finally we present a method to get the edge image by feature image. We experimentally verify the feasibility of this method.

- 语义逻辑混乱.
- 语法有问题.

A New Edge Detection Method Based on Hausdorff Distance

You Lu **Qingxin Meng** **Jiangtao Guo**
School of Computer Science and Technology, China University of Petroleum, Beijing
Date Added to IEEE *Xplore*: 29 December 2014 | *DOI: 10.1109/ICDH.2014.15*
5^{th} International Conference on Digital Home (ICDH 2014), ©IEEE

Abstract—Most edge detection ~~methods are based on~~ algorithms, first-order or second-order differential ~~. These are local methods. Using Hausdorff distance~~ based, are localization methods. Here we introduce an algorithm based on Hausdorff distance, different from localization methods, to quantify the ~~strength of the edge is a method with a holistic property. Firstly, down sample the image, and split the image into two sets. Secondly, get the feature image by assigning a value for each point using the~~ edge intensity. We establish the algorithm in two stages: 1)we construct a scalar field map ~~constructed~~ from two down sampled images to one featured image by Hausdorff distance~~. Since the edge features have local properties, in this paper, we constructed a map which can get local feature images using sub image and combined them into a feature image. Finally~~ ; and 2)we present a method to get the edge image ~~by feature image. We experimentally verify~~ from featured image. Due to the dependency of interval starting point in down sample, the featured image is replaced by its expectation. We verified the feasibility of this ~~method~~algorithm through experiments.

A New Edge Detection Method Based on Hausdorff Distance

You Lu **Qingxin Meng** **Jiangtao Guo**
School of Computer Science and Technology, China University of Petroleum, Beijing
Date Added to IEEE *Xplore*: 29 December 2014 | *DOI: 10.1109/ICDH.2014.15*
5th International Conference on Digital Home (ICDH 2014), ©IEEE

Abstract — Most edge detection algorithms, first-order or second-order differential based, are localization methods. Here we introduce an algorithm based on Hausdorff distance, different from localization methods, to quantify the edge intensity. We establish the algorithm in two stages: 1)we construct a scalar field map from two down sampled images to one featured image by Hausdorff distance; and 2)we present a method to get the edge image from featured image. Due to the dependency of interval starting point in down sample, the featured image is replaced by its expectation. We verified the feasibility of this algorithm through experiments.

摘要译文 —— 多数边缘检测算法（基于一阶、二阶微分）是局部化方法. 这里我们介绍一种基于 Hausdorff 距离的非局部化算法以量化边缘强度. 我们分两个阶段建立算法: 1) 通过 Hausdorff 距离构造一个从两个下采样图像到一个特征图像的标量场映射; 2) 提出了一种从特征图像中获取边缘图像的方法. 由于下采样依赖于间隔起始点, 特征图像要替换为它的数学期望. 我们通过实验验证了该算法的可行性.

A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play

David Silver[1,2], **Thomas Hubert**[1], **Julian Schrittwieser**[1], **Ioannis Antonoglou**[1], **Matthew Lai**[1], **Arthur Guez**[1], **Marc Lanctot**[1], **Laurent Sifre**[1], **Dharshan Kumaran**[1], **Thore Graepel**[1], **Timothy Lillicrap**[1], **Karen Simonyan1**[1], **Demis Hassabis**[1]

[1] DeepMind, 6 Pancras Square, London N1C 4AG, UK.

[2] University College London, Gower Street, London WC1E 6BT, UK.

Science 07 Dec 2018: Vol. 362, Issue 6419, pp. 1140-1144 | *DOI: 10.1126/science.aar6404*

Abstract — The game of chess is the longest-studied domain in the history of artificial intelligence. The strongest programs are based on a combination of sophisticated search techniques, domain-specific adaptations, and handcrafted evaluation functions that have been refined by human experts over several decades. By contrast, the AlphaGo Zero program recently achieved superhuman performance in the game of Go by reinforcement learning from self-play.**B** In this paper, we generalize this approach into a single AlphaZero algorithm that can achieve superhuman performance in many challenging games.**I** Starting from random play and given no domain knowledge except the game rules, AlphaZero convincingly defeated a world champion program in the games of chess and shogi (Japanese chess), as well as Go.**R**

Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks

Pranav Rajpurkar* **Awni Y. Hannun*** **Masoumeh Haghpanahi⋆** **Codie Bourn⋆** **Andrew Y. Ng***
* Stanford University ⋆ iRhythm Technologies

Submitted on 6 Jul 2017 | *arXiv:1707.01836v1* [cs.CV]
Stanford Machine Learning Group | *Project website at* `https://stanfordmlgroup.github.io/projects/ecg/`

Abstract — We develop an algorithm which exceeds the performance of board certified cardiologists in detecting a wide range of heart arrhythmias from electrocardiograms recorded with a single-lead wearable monitor.[I] We build a dataset with more than 500 times the number of unique patients than previously studied corpora. On this dataset, we train a 34-layer convolutional neural network which maps a sequence of ECG samples to a sequence of rhythm classes. Committees of board-certified cardiologists annotate a gold standard test set on which we compare the performance of our model to that of 6 other individual cardiologists.[M] We exceed the average cardiologist performance in both recall (sensitivity) and precision (positive predictive value).[R]

Neural Ordinary Differential Equations

Ricky T. Q. Chen* **Yulia Rubanova*** **Jesse Bettencourt*** **David Duvenaud**
University of Toronto, Vector Institute, Toronto, Canada.
Submitted on 19 Jun 2018, last revised 22 Oct 2018 | *arXiv:1806.07366v3* [cs.LG]
32nd Conference on Neural Information Processing Systems (NeurIPS 2018 Best Paper), Montréal, Canada.

Abstract — We introduce a new family of deep neural network models. Instead of specifying a discrete sequence of hidden layers, we parameterize the derivative of the hidden state using a neural network. The output of the network is computed using a blackbox differential equation solver. These continuous-depth models have constant memory cost, adapt their evaluation strategy to each input, and can explicitly trade numerical precision for speed. We demonstrate these properties in continuous-depth residual networks and continuous-time latent variable models. We also construct continuous normalizing flows, a generative model that can train by maximum likelihood, without partitioning or ordering the data dimensions. For training, we show how to scalably backpropagate through any ODE solver, without access to its internal operations. This allows end-to-end training of ODEs within larger models.[1]

神经常微分方程

Ricky T. Q. Chen* **Yulia Rubanova*** **Jesse Bettencourt*** **David Duvenaud**
University of Toronto, Vector Institute, Toronto, Canada.
Submitted on 19 Jun 2018, last revised 22 Oct 2018 | *arXiv:1806.07366v3* [cs.LG]
32nd Conference on Neural Information Processing Systems (NeurIPS 2018 Best Paper), Montréal, Canada.

摘要译文 —— 我们介绍了一系列新的深度神经网络模型. 我们使用神经网络参数化隐藏状态的导数, 而不是指定隐藏层的离散序列. 我们使用黑盒微分方程求解器计算网络的输出. 这些连续深度模型具有恒定的内存消耗, 根据输入采取相应的评估策略, 并且可以直接通过降低数值精度来提升速度. 我们在连续深度残差网络和连续时间潜变量模型中证明了这些性质. 我们还构建了连续归一化流, 这是一种可以通过最大似然进行训练的生成模型, 无需对数据维度进行分区或排序. 对于模型训练, 我们展示了如何通过任何 ODE 求解器进行规模化地反向传播, 而无需知道其内部操作. 这允许我们在较大模型中对 ODE 进行端到端训练. [I]

Optimal Algorithms for Non-Smooth Distributed Optimization in Networks

Kevin Scaman **Francis Bach** **Sébastien Bubeck** **Yin Tat Lee** **Laurent Massoulié**

Submitted on 1 Jun 2018 | *arXiv:1806.00291v1* [math.OC]

32nd Conference on Neural Information Processing Systems (NeurIPS 2018 Best Paper), Montréal, Canada.

Abstract — In this work, we consider the distributed optimization of non-smooth convex functions using a network of computing units. We investigate this problem under two regularity assumptions: (1) the Lipschitz continuity of the *global* objective function, and (2) the Lipschitz continuity of *local* individual functions. Under the *local regularity* assumption, we provide the first optimal first-order decentralized algorithm called *multi-step primal-dual* (MSPD) and its corresponding optimal convergence rate. A notable aspect of this result is that, for non-smooth functions, while the dominant term of the error is in $O(1/\sqrt{t})$, the structure of the communication network only impacts a second-order term in $O(1/t)$, where t is time. In other words, the error due to limits in communication resources decreases at a fast rate even in the case of non-strongly-convex objective functions. Under the *global regularity* assumption, we provide a simple yet efficient algorithm called *distributed randomized smoothing* (DRS) based on a local smoothing of the objective function, and show that DRS is within a $d^{1/4}$ multiplicative factor of the optimal convergence rate, where d is the underlying dimension.[1]

网络中非光滑分布式优化的最优算法

Kevin Scaman **Francis Bach** **Sébastien Bubeck** **Yin Tat Lee** **Laurent Massoulié**

Submitted on 1 Jun 2018 | *arXiv:1806.00291v1* [math.OC]

32nd Conference on Neural Information Processing Systems (NeurIPS 2018 Best Paper), Montréal, Canada.

摘要译文 —— 在这项工作中, 我们使用一个计算单元网络, 对非光滑凸函数进行了分布式优化. 我们在假设下研究这个问题: (1) 全局目标函数的 Lipschitz 连续性, 以及 (2) 局部个体函数的 Lipschitz 连续性. 在局部性假设下, 我们提出了第一个最优一阶分散算法, 称为多步原始对偶 (MSPD), 及其相应的最优收敛速度. 该结果的一个值得注意的方面是, 对于非平滑函数, 当误差的主导项为 $O(1/\sqrt{t})$ 时, 通信网络结构仅影响 $O(1/t)$ 中的二阶项, 其中 t 是时间. 换句话说, 即使在非强凸目标函数的情况下, 由于通信资源中的限制而导致的误差也会以较快的速率降低. 在全局性假设下, 我们提供了一种简单而有效的算法, 称为基于目标函数局部平滑的分布式随机平滑 (DRS), 并表明 DRS 和最优收敛速度只差一个 $d^{1/4}$ 的乘法项, 其中 d 是维度. [1]

Non-delusional Q-learning and value-iteration

Tyler Lu **Dale Schuurmans** **Craig Boutilier**
Google AI

32nd Conference on Neural Information Processing Systems (NeurIPS 2018 Best Paper), Montréal, Canada.

Abstract — We identify a fundamental source of error in Q-learning and other forms of dynamic programming with function approximation. Delusional bias arises when the approximation architecture limits the class of expressible greedy policies.[I] Since standard Q-updates make globally uncoordinated action choices with respect to the expressible policy class, inconsistent or even conflicting Q-value estimates can result, leading to pathological behaviour such as over/under-estimation, instability and even divergence.[B] To solve this problem, we introduce a new notion of policy consistency and define a local backup process that ensures global consistency through the use of information sets—sets that record constraints on policies consistent with backed-up Q-values. We prove that both the model-based and model-free algorithms using this backup remove delusional bias, yielding the first known algorithms that guarantee optimal results under general conditions. These algorithms furthermore only require polynomially many information sets (from a potentially exponential support). Finally, we suggest other practical heuristics for value-iteration and Q-learning that attempt to reduce delusional bias.[I]

Exploring Generalization in Deep Learning

Behnam Neyshabur **Srinadh Bhojanapalli** **David Mcallester** **Nati Srebro**
Toyota Technological Institute at Chicago

Submitted on 16 Oct 2017, last revised 22 Feb 2018 | *arXiv:1710.05468v3* [stat.ML]
32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

Abstract — With a goal of understanding what drives generalization in deep networks, we consider several recently suggested explanations, including norm-based control, sharpness and robustness. We study how these measures can ensure generalization, highlighting the importance of scale normalization, and making a connection between sharpness and PAC-Bayes theory. We then investigate how well the measures explain different observed phenomena.[1]

Video-to-Video Synthesis

Ting-Chun Wang* **Ming-Yu Liu*** **Jun-Yan Zhu**⋆ **Nikolai Yakovenko***
Andrew Tao* **Jan Kautz*** **Bryan Catanzaro***
*NVIDIA ⋆MIT CSAIL

Submitted on 20 Aug 2018, last revised 3 Dec 2018 | *arXiv:1808.06601v2* [cs.CV]
32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

Abstract — We study the problem of video-to-video synthesis, whose goal is to learn a mapping function from an input source video (e.g., a sequence of semantic segmentation masks) to an output photorealistic video that precisely depicts the content of the source video. While its image counterpart, the image-to-image translation problem, is a popular topic, the video-to-video synthesis problem is less explored in the literature. Without modeling temporal dynamics, directly applying existing image synthesis approaches to an input video often results in temporally incoherent videos of low visual quality.[B] In this paper, we propose a video-to-video synthesis approach under the generative adversarial learning framework.[I] Through carefully-designed generators and discriminators, coupled with a spatio-temporal adversarial objective, we achieve high-resolution, photorealistic, temporally coherent video results on a diverse set of input formats including segmentation masks, sketches, and poses.[M] Experiments on multiple benchmarks show the advantage of our method compared to strong baselines. In particular, our model is capable of synthesizing 2K resolution videos of street scenes up to 30 seconds long, which significantly advances the state-of-the-art of video synthesis.[R] Finally, we apply our method to future video prediction, outperforming several competing systems.[C] Code, models, and more results are available at our `website`.

Safe and Nested Subgame Solving for Imperfect-Information Games

Noam Brown **Tuomas Sandholm**
Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15217
Submitted on 8 May 2017, last revised 16 Nov 2017 | *arXiv:1705.02955v3* [cs.AI]
31st Conference on Neural Information Processing Systems (NIPS 2017 Best Paper), Long Beach, CA, USA.

Abstract — In imperfect-information games, the optimal strategy in a subgame may depend on the strategy in other, unreached subgames. Thus a subgame cannot be solved in isolation and must instead consider the strategy for the entire game as a whole, unlike perfect-information games. Nevertheless, it is possible to first approximate a solution for the whole game and then improve it in individual subgames. This is referred to as *subgame solving*.[B] We introduce subgame-solving techniques that outperform prior methods both in theory and practice. We also show how to adapt them, and past subgame-solving techniques, to respond to opponent actions that are outside the original action abstraction; this significantly outperforms the prior state-of-the-art approach, action translation. Finally, we show that subgame solving can be repeated as the game progresses down the game tree, leading to far lower exploitability. These techniques were a key component of *Libratus*, the first AI to defeat top humans in heads-up no-limit Texas hold'em poker.[I]

A Linear-Time Kernel Goodness-of-Fit Test

Wittawat Jitkrittum Gatsby Unit, UCL
Wenkai Xu Gatsby Unit, UCL
Zoltán Szabó CMAP, École Polytechnique
Kenji Fukumizu The Institute of Statistical Mathematics
Arthur Gretton Gatsby Unit, UCL

Submitted on 22 May 2017, last revised 24 Oct 2017 | *arXiv:1705.07673v2* [stat.ML]
31st Conference on Neural Information Processing Systems (NIPS 2017 Best Paper), Long Beach, CA, USA.

Abstract — We propose a novel adaptive test of goodness-of-fit, with computational cost linear in the number of samples.**I** We learn the test features that best indicate the differences between observed samples and a reference model, by minimizing the false negative rate. These features are constructed via Stein's method, meaning that it is not necessary to compute the normalising constant of the model.**M** We analyse the asymptotic Bahadur efficiency of the new test, and prove that under a mean-shift alternative, our test always has greater relative efficiency than a previous linear-time kernel test, regardless of the choice of parameters for that test.**R** In experiments, the performance of our method exceeds that of the earlier linear-time test, and matches or exceeds the power of a quadratic-time kernel test. In high dimensions and where model structure may be exploited, our goodness of fit test performs far better than a quadratic-time two-sample test based on the Maximum Mean Discrepancy, with samples drawn from the model.**C**

Variance-based Regularization with Convex Objectives

Hongseok Namkoong **John C. Duchi**
Stanford University

Submitted on 8 Oct 2016, last revised 14 Dec 2017 | *arXiv:1610.02581v3* [stat.ML]
31st Conference on Neural Information Processing Systems (NIPS 2017 Best Paper), Long Beach, CA, USA.

Abstract — We develop an approach to risk minimization and stochastic optimization that provides a convex surrogate for variance, allowing near-optimal and computationally efficient trading between approximation and estimation error.[I] Our approach builds off of techniques for distributionally robust optimization and Owen's empirical likelihood, and we provide a number of finite-sample and asymptotic results characterizing the theoretical performance of the estimator. In particular, we show that our procedure comes with certificates of optimality, achieving (in some scenarios) faster rates of convergence than empirical risk minimization by virtue of automatically balancing bias and variance.[M] We give corroborating empirical evidence showing that in practice, the estimator indeed trades between variance and absolute performance on a training sample, improving out-of-sample (test) performance over standard empirical risk minimization for a number of classification problems.[R]

Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images

Anh Nguyen University of Wyoming
Jason Yosinski Cornell University
Jeff Clune University of Wyoming

Submitted on 5 Dec 2014, last revised 2 Apr 2015 | *arXiv:1412.1897v4* [cs.CV]
Date Added to IEEE *Xplore*: 15 October 2015 | *DOI: 10.1109/CVPR.2015.7298640*
2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), ©IEEE

Abstract — Deep neural networks (DNNs) have recently been achieving state-of-the-art performance on a variety of pattern-recognition tasks, most notably visual classification problems. Given that DNNs are now able to classify objects in images with near-human-level performance, questions naturally arise as to what differences remain between computer and human vision. A recent study[30] revealed that changing an image (e.g. of a lion) in a way imperceptible to humans can cause a DNN to label the image as something else entirely (e.g. mislabeling a lion a library).**[B]** Here we show a related result: it is easy to produce images that are completely unrecognizable to humans, but that state-of-the-art DNNs believe to be recognizable objects with 99.99% confidence (e.g. labeling with certainty that white noise static is a lion).**[I]** Specifically, we take convolutional neural networks trained to perform well on either the ImageNet or MNIST datasets and then find images with evolutionary algorithms or gradient ascent that DNNs label with high confidence as belonging to each dataset class.**[M]** It is possible to produce images totally unrecognizable to human eyes that DNNs believe with near certainty are familiar objects, which we call "fooling images" (more generally, fooling examples).**[R]** Our results shed light on interesting differences between human vision and current DNNs, and raise questions about the generality of DNN computer vision.**[C]**

Dance Dance Convolution

Chris Donahue* **Zachary C. Lipton**★ **Julian McAuley**★
* UCSD Department of Music, San Diego, CA ★ UCSD Department of Computer Science, San Diego, CA

Submitted on 20 Mar 2017, last revised 21 Jun 2017 | *arXiv:1703.06891v3* [cs.LG]
Proceedings of the 34th International Conference on Machine Learning (ICML 2017), Sydney, Australia, PMLR 70, 2017.

Abstract — Dance Dance Revolution (DDR) is a popular rhythm-based video game. Players perform steps on a dance platform in synchronization with music as directed by on-screen step charts. While many step charts are available in standardized packs, players may grow tired of existing charts, or wish to dance to a song for which no chart exists.**[B]** We introduce the task of learning to choreograph. Given a raw audio track, the goal is to produce a new step chart.**[I]** This task decomposes naturally into two subtasks: deciding when to place steps and deciding which steps to select. For the step placement task, we combine recurrent and convolutional neural networks to ingest spectrograms of low-level audio features to predict steps, conditioned on chart difficulty. For step selection, we present a conditional LSTM generative model that substantially outperforms n-gram and fixed-window approaches.**[M]**

Don't Decay the Learning Rate, Increase the Batch Size

Samuel L. Smith **Pieter-Jan Kindermans** **Chris Ying** **Quoc V. Le**
Google Brain

Submitted on 1 Nov 2017, last revised 24 Feb 2018 | *arXiv:1711.00489v2* [cs.LG]
6th International Conference on Learning Representations (ICLR 2018), Vancouver, BC, Canada.

Abstract — It is common practice to decay the learning rate.**B** 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. This procedure is successful for stochastic gradient descent (SGD), SGD with momentum, Nesterov momentum, and Adam. It reaches equivalent test accuracies after the same number of training epochs, but with fewer parameter updates, leading to greater parallelism and shorter training times.**I** We can further reduce the number of parameter updates by increasing the learning rate ϵ and scaling the batch size $B \propto \epsilon$. Finally, one can increase the momentum coefficient m and scale $B \propto 1/(1-m)$, although this tends to slightly reduce the test accuracy.**M** Crucially, our techniques allow us to repurpose existing training schedules for large batch training with no hyper-parameter tuning. We train ResNet-50 on ImageNet to 76.1% validation accuracy in under 30 minutes.**R**

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks

Alec Radford
indico Research, Boston, MA

Luke Metz
indico Research, Boston, MA

Soumith Chintala
Facebook AI Research, New York, NY

Submitted on 19 Nov 2015, last revised 7 Jan 2016 | *arXiv:1511.06434v2* [cs.LG]
4th International Conference on Learning Representations (ICLR 2016), San Juan, Puerto Rico.

Abstract — In recent years, supervised learning with convolutional networks (CNNs) has seen huge adoption in computer vision applications. Comparatively, unsupervised learning with CNNs has received less attention.**B** In this work we hope to help bridge the gap between the success of CNNs for supervised learning and unsupervised learning. We introduce a class of CNNs called deep convolutional generative adversarial networks (DCGANs), that have certain architectural constraints, and demonstrate that they are a strong candidate for unsupervised learning.**I** Training on various image datasets, we show convincing evidence that our deep convolutional adversarial pair learns a hierarchy of representations from object parts to scenes in both the generator and discriminator. Additionally, we use the learned features for novel tasks - demonstrating their applicability as general image representations.**R**

U-Net: Convolutional Networks for Biomedical Image Segmentation

Olaf Ronneberger ✉ **Philipp Fischer** **Thomas Brox**
Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany

Submitted on 18 May 2015 | *arXiv:1505.04597v1* [cs.CV]
18^{th} *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, Germany.*

Abstract — There is large consent that successful training of deep networks requires many thousand annotated training samples.[B] In this paper, we present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently.[I] The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization.[M] We show that such a network can be trained end-to-end from very few images and outperforms the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks. Using the same network trained on transmitted light microscopy images (phase contrast and DIC) we won the ISBI cell tracking challenge 2015 in these categories by a large margin. Moreover, the network is fast. Segmentation of a 512×512 image takes less than a second on a recent GPU.[R] The full implementation (based on Caffe) and the trained networks are available at

`http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net.`

Deep Learning: A Bayesian Perspective

Nicholas Polson
Booth School of Business,
University of Chicago

Vadim Sokolov
Volgenau School of Engineering,
George Mason University

Submitted on 1 Jun 2017, last revised 14 Nov 2017 | *arXiv:1706.00473v4* [stat.ML]
Bayesian Anal. Volume 12, Number 4 (2017), 1275-1304. | *DOI: 10.1214/17-BA1082*

Abstract — Deep learning is a form of machine learning for nonlinear high dimensional pattern matching and prediction.[B] By taking a Bayesian probabilistic perspective, we provide a number of insights into more efficient algorithms for optimisation and hyper-parameter tuning.[I] Traditional high-dimensional data reduction techniques, such as principal component analysis (PCA), partial least squares (PLS), reduced rank regression (RRR), projection pursuit regression (PPR) are all shown to be shallow learners. Their deep learning counterparts exploit multiple deep layers of data reduction which provide predictive performance gains. Stochastic gradient descent (SGD) training optimisation and Dropout (DO) regularization provide estimation and variable selection. Bayesian regularization is central to finding weights and connections in networks to optimize the predictive bias-variance trade-off. To illustrate our methodology, we provide an analysis of international bookings on Airbnb.[R] Finally, we conclude with directions for future research.[C]