$rm K$-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation

Type	Journal Article
Author	M. Aharon
Author	M. Elad
Author	A. Bruckstein
URL	http://ieeexplore.ieee.org/document/1710377/
Volume	54
Issue	11
Pages	4311-4322
Publication	IEEE Transactions on Signal Processing
ISSN	1053-587X
Date	11/2006
Extra	00000
Journal Abbr	IEEE Trans. Signal Process.
DOI	10/dx8n4f
Accessed	6/21/2019, 11:46:07 AM
Library Catalog	DOI.org (Crossref)
Language	en
Short Title	$rm K$-SVD
Date Added	6/21/2019, 11:46:07 AM
Modified	6/21/2019, 11:46:11 AM

Attachments

- Aharon et al. - 2006 - $rm K$-SVD An Algorithm for Designing Overcomplet.pdf

A Study on Overfitting in Deep Reinforcement Learning

Type	Journal Article
Author	Chiyuan Zhang
Author	Oriol Vinyals
Author	Remi Munos
Author	Samy Bengio
URL	http://arxiv.org/abs/1804.06893
Publication	arXiv:1804.06893 [cs, stat]
Date	2018-04-18
Extra	00032 arXiv: 1804.06893

Accessed 6/21/2019, 12:46:15 PM
Library Catalog arXiv.org
Abstract Recent years have witnessed significant progresses in deep Reinforcement Learning (RL). Empowered with large scale neural networks, carefully designed architectures, novel training algorithms and massively parallel computing devices, researchers are able to attack many challenging RL problems. However, in machine learning, more training power comes with a potential risk of more overfitting. As deep RL techniques are being applied to critical problems such as healthcare and finance, it is important to understand the generalization behaviors of the trained agents. In this paper, we conduct a systematic study of standard RL agents and find that they could overfit in various ways. Moreover, overfitting could happen "robustly": commonly used techniques in RL that add stochasticity do not necessarily prevent or detect overfitting. In particular, the same agents and learning algorithms could have drastically different test performance, even when all of them achieve optimal rewards during training. The observations call for more principled and careful evaluation protocols in RL. We conclude with a general discussion on overfitting in RL and a study of the generalization behaviors from the perspective of inductive bias.
Date Added 6/21/2019, 12:46:15 PM
Modified 6/21/2019, 12:46:58 PM

Tags:

Computer Science - Machine Learning, Statistics - Machine Learning

Attachments

- arXiv.org Snapshot
- Zhang et al_2018_A Study on Overfitting in Deep Reinforcement Learning.pdf

---

A tutorial on statistically sound pattern discovery

Type Journal Article
Author Wilhelmiina Hämäläinen
Author Geoffrey I. Webb
URL http://link.springer.com/10.1007/s10618-018-0590-x
Volume 33
Issue 2
Pages 325-377
Publication Data Mining and Knowledge Discovery
ISSN 1384-5810, 1573-756X
Date 3/2019
Journal Abbr Data Min Knowl Disc

DOI 10/gf35nv
Accessed 6/21/2019, 11:46:25 AM
Library Catalog DOI.org (Crossref)
Language en
Abstract Statistically sound pattern discovery harnesses the rigour of statistical hypothesis testing to overcome many of the issues that have hampered standard data mining approaches to pattern discovery. Most importantly, application of appropriate statistical tests allows precise control over the risk of false discoveries—patterns that are found in the sample data but do not hold in the wider population from which the sample was drawn. Statistical tests can also be applied to filter out patterns that are unlikely to be useful, removing uninformative variations of the key patterns in the data. This tutorial introduces the key statistical and data mining theory and techniques that underpin this fast developing field. We concentrate on two general classes of patterns: dependency rules that express statistical dependencies between condition and consequent parts and dependency sets that express mutual dependence between set elements. We clarify alternative interpretations of statistical dependence and introduce appropriate tests for evaluating statistical significance of patterns in different situations. We also introduce special techniques for controlling the likelihood of spurious discoveries when multitudes of patterns are evaluated. The paper is aimed at a wide variety of audiences. It provides the necessary statistical background and summary of the state-of-the-art for any data mining researcher or practitioner wishing to enter or understand statistically sound pattern discovery research or practice. It can serve as a general introduction to the field of statistically sound pattern discovery for any reader with a general background in data sciences.
Date Added 6/21/2019, 11:46:25 AM
Modified 6/21/2019, 11:46:27 AM

Attachments

- Hämäläinen and Webb - 2019 - A tutorial on statistically sound pattern discover.pdf

---

An empirical analysis of dropout in piecewise linear networks

Type Journal Article
Author David Warde-Farley
Author Ian J. Goodfellow
Author Aaron Courville
Author Yoshua Bengio
URL http://arxiv.org/abs/1312.6197
Publication arXiv:1312.6197 [cs, stat]
Date 2013-12-20
Extra 00053 arXiv: 1312.6197

Accessed 6/21/2019, 11:56:11 AM
Library Catalog arXiv.org
Abstract The recently introduced dropout training criterion for neural networks has been the subject of much attention due to its simplicity and remarkable effectiveness as a regularizer, as well as its interpretation as a training procedure for an exponentially large ensemble of networks that share parameters. In this work we empirically investigate several questions related to the efficacy of dropout, specifically as it concerns networks employing the popular rectified linear activation function. We investigate the quality of the test time weight-scaling inference procedure by evaluating the geometric average exactly in small models, as well as compare the performance of the geometric mean to the arithmetic mean more commonly employed by ensemble techniques. We explore the effect of tied weights on the ensemble interpretation by training ensembles of masked networks without tied weights. Finally, we investigate an alternative criterion based on a biased estimator of the maximum likelihood ensemble gradient.
Date Added 6/21/2019, 11:56:11 AM
Modified 6/21/2019, 11:56:43 AM

Tags:

⛔ No DOI found, Computer Science - Neural and Evolutionary Computing, Computer Science - Machine Learning, Statistics - Machine Learning

Notes:

Comment: Extensive updates; 8 pages plus acknowledgements/references

---

An Introduction to Variational Autoencoders

Type Journal Article
Author Diederik P. Kingma
Author Max Welling
URL http://arxiv.org/abs/1906.02691
Publication arXiv:1906.02691 [cs, stat]
Date 2019-06-06
Extra 00000 arXiv: 1906.02691
Accessed 6/21/2019, 6:44:44 PM
Library Catalog arXiv.org
Abstract Variational autoencoders provide a principled framework for learning deep latent-variable models and corresponding inference models. In this work, we provide an introduction to variational autoencoders and some important extensions.

Date Added 6/21/2019, 6:44:44 PM
Modified 6/21/2019, 6:45:13 PM

Tags:

Computer Science - Machine Learning, Statistics - Machine Learning

Attachments

- arXiv.org Snapshot
- Kingma_Welling_2019_An Introduction to Variational Autoencoders.pdf

Contents

Attention Is All You Need

Type	Journal Article
Author	Ashish Vaswani
Author	Noam Shazeer
Author	Niki Parmar
Author	Jakob Uszkoreit
Author	Llion Jones
Author	Aidan N. Gomez
Author	Lukasz Kaiser
Author	Illia Polosukhin
URL	http://arxiv.org/abs/1706.03762
Publication	arXiv:1706.03762 [cs]
Date	2017-06-12
Extra	02136 arXiv: 1706.03762
Accessed	6/21/2019, 5:04:35 PM
Library Catalog	arXiv.org
Abstract	The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.
Date Added	6/21/2019, 5:04:35 PM
Modified	6/21/2019, 5:05:45 PM

Tags:

Computer Science - Computation and Language, Computer Science - Machine Learning

Notes:

Comment: 15 pages, 5 figures

Attachments

- arXiv.org Snapshot
- Vaswani et al_2017_Attention Is All You Need.pdf

Contents

Attentive Neural Processes

Type Journal Article
Author Hyunjik Kim
Author Andriy Mnih
Author Jonathan Schwarz
Author Marta Garnelo
Author Ali Eslami
Author Dan Rosenbaum
Author Oriol Vinyals
Author Yee Whye Teh
URL http://arxiv.org/abs/1901.05761

Publication	arXiv:1901.05761 [cs, stat]
Date	2019-01-17
Extra	00007 arXiv: 1901.05761
Accessed	6/21/2019, 12:45:52 PM
Library Catalog	arXiv.org
Abstract	Neural Processes (NPs) (Garnelo et al 2018a;b) approach regression by learning to map a context set of observed input-output pairs to a distribution over regression functions. Each function models the distribution of the output given an input, conditioned on the context. NPs have the benefit of fitting observed data efficiently with linear complexity in the number of context input-output pairs, and can learn a wide family of conditional distributions; they learn predictive distributions conditioned on context sets of arbitrary size. Nonetheless, we show that NPs suffer a fundamental drawback of underfitting, giving inaccurate predictions at the inputs of the observed data they condition on. We address this issue by incorporating attention into NPs, allowing each input location to attend to the relevant context points for the prediction. We show that this greatly improves the accuracy of predictions, results in noticeably faster training, and expands the range of functions that can be modelled.
Date Added	6/21/2019, 12:45:52 PM
Modified	6/21/2019, 12:46:44 PM

Tags:

Computer Science - Machine Learning, Statistics - Machine Learning

Attachments

- arXiv.org Snapshot
- Kim et al_2019_Attentive Neural Processes.pdf

Contents

Auto-Encoding Variational Bayes

Type	Journal Article
Author	Diederik P. Kingma
Author	Max Welling
URL	http://arxiv.org/abs/1312.6114
Publication	arXiv:1312.6114 [cs, stat]
Date	2013-12-20
Extra	05075 arXiv: 1312.6114
Accessed	6/21/2019, 5:57:38 PM
Library Catalog	arXiv.org
Abstract	How can we perform efficient inference and learning in directed probabilistic models, in the presence of continuous latent variables with intractable posterior distributions, and large datasets? We introduce a stochastic variational inference and learning algorithm that scales to large datasets and, under some mild differentiability conditions, even works in the intractable case. Our contributions is two-fold. First, we show that a reparameterization of the variational lower bound yields a lower bound estimator that can be straightforwardly optimized using standard stochastic gradient methods. Second, we show that for i.i.d. datasets with continuous latent variables per datapoint, posterior inference can be made especially efficient by fitting an approximate inference model (also called a recognition model) to the intractable posterior using the proposed lower bound estimator. Theoretical advantages are reflected in experimental results.

Date Added 6/21/2019, 5:57:38 PM
Modified 6/21/2019, 5:59:07 PM

Tags:

Computer Science - Machine Learning, Statistics - Machine Learning

Attachments

- arXiv.org Snapshot
- Kingma_Welling_2013_Auto-Encoding Variational Bayes.pdf

Contents

Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift

Type Journal Article
Author Sergey Ioffe

Author Christian Szegedy
Pages 9
Library Catalog Zotero
Language en
Abstract Training Deep Neural Networks is complicated by the fact that the distribution of each layer's inputs changes during training, as the parameters of the previous layers change. This slows down the training by requiring lower learning rates and careful parameter initialization, and makes it notoriously hard to train models with saturating nonlinearities. We refer to this phenomenon as internal covariate shift, and address the problem by normalizing layer inputs. Our method draws its strength from making normalization a part of the model architecture and performing the normalization for each training mini-batch. Batch Normalization allows us to use much higher learning rates and be less careful about initialization, and in some cases eliminates the need for Dropout. Applied to a stateof-the-art image classification model, Batch Normalization achieves the same accuracy with 14 times fewer training steps, and beats the original model by a significant margin. Using an ensemble of batch-normalized networks, we improve upon the best published result on ImageNet classification: reaching 4.82% top-5 test error, exceeding the accuracy of human raters.
Date Added 6/21/2019, 11:31:58 AM
Modified 6/21/2019, 11:32:02 AM

Tags:

⛔ No DOI found

Attachments

- Ioffe and Szegedy - Batch Normalization Accelerating Deep Network Tra.pdf

---

Convolutional Neural Networks for Sentence Classification

Type Journal Article
Author Yoon Kim
URL http://arxiv.org/abs/1408.5882
Publication arXiv:1408.5882 [cs]
Date 2014-08-25
Extra 04407 arXiv: 1408.5882
Accessed 6/21/2019, 11:30:28 AM
Library Catalog arXiv.org
Language en
Abstract We report on a series of experiments with convolutional neural networks (CNN) trained on top of pre-trained word vectors for sentence-level

	classification tasks. We show that a simple CNN with little hyperparameter tuning and static vectors achieves excellent results on multiple benchmarks. Learning task-specific vectors through fine-tuning offers further gains in performance. We additionally propose a simple modification to the architecture to allow for the use of both task-specific and static vectors. The CNN models discussed herein improve upon the state of the art on 4 out of 7 tasks, which include sentiment analysis and question classification.
Date Added	6/21/2019, 11:30:28 AM
Modified	6/21/2019, 11:30:30 AM

Tags:

⛔ No DOI found, Computer Science - Computation and Language, Computer Science - Neural and Evolutionary Computing

Notes:

Comment: To appear in EMNLP 2014

Attachments

- Kim - 2014 - Convolutional Neural Networks for Sentence Classif.pdf

---

Convolutional Sequence to Sequence Learning

Type	Journal Article
Author	Jonas Gehring
Author	Michael Auli
Author	David Grangier
Author	Denis Yarats
Author	Yann N. Dauphin
URL	http://arxiv.org/abs/1705.03122
Publication	arXiv:1705.03122 [cs]
Date	2017-05-08
Extra	00688 arXiv: 1705.03122
Accessed	6/21/2019, 5:04:24 PM
Library Catalog	arXiv.org
Abstract	The prevalent approach to sequence to sequence learning maps an input sequence to a variable length output sequence via recurrent neural networks. We introduce an architecture based entirely on convolutional neural networks. Compared to recurrent models, computations over all elements can be fully parallelized during training and optimization is easier since the number of non-linearities is fixed and independent of the input length. Our

	use of gated linear units eases gradient propagation and we equip each decoder layer with a separate attention module. We outperform the accuracy of the deep LSTM setup of Wu et al. (2016) on both WMT'14 English-German and WMT'14 English-French translation at an order of magnitude faster speed, both on GPU and CPU.
Date Added	6/21/2019, 5:04:24 PM
Modified	6/21/2019, 5:04:39 PM

Tags:

Computer Science - Computation and Language

Attachments

- arXiv.org Snapshot
- Gehring et al_2017_Convolutional Sequence to Sequence Learning.pdf

---

Deep learning

Type	Journal Article
Author	Yann LeCun
Author	Yoshua Bengio
Author	Geoffrey Hinton
URL	http://www.nature.com/articles/nature14539
Volume	521
Issue	7553
Pages	436-444
Publication	Nature
ISSN	0028-0836, 1476-4687
Date	5/2015
Extra	15695
Journal Abbr	Nature
DOI	10/bmqp
Accessed	6/21/2019, 11:43:56 AM
Library Catalog	DOI.org (Crossref)
Language	en
Date Added	6/21/2019, 11:43:56 AM
Modified	6/21/2019, 11:43:58 AM

Attachments

- LeCun et al. - 2015 - Deep learning.pdf

Design challenges and misconceptions in named entity recognition

Type	Conference Paper
Author	Lev Ratinov
Author	Dan Roth
URL	http://portal.acm.org/citation.cfm?doid=1596374.1596399
Place	Boulder, Colorado
Publisher	Association for Computational Linguistics
Pages	147
ISBN	978-1-932432-29-9
Date	2009
DOI	10/dkk8vf
Accessed	6/21/2019, 1:35:08 PM
Library Catalog	DOI.org (Crossref)
Conference Name	the Thirteenth Conference
Language	en
Abstract	We analyze some of the fundamental design challenges and misconceptions that underlie the development of an efficient and robust NER system. In particular, we address issues such as the representation of text chunks, the inference approach needed to combine local NER decisions, the sources of prior knowledge and how to use them within an NER system. In the process of comparing several solutions to these challenges we reach some surprising conclusions, as well as develop an NER system that achieves 90.8 F1 score on the CoNLL-2003 NER shared task, the best reported result for this dataset.
Proceedings Title	Proceedings of the Thirteenth Conference on Computational Natural Language Learning - CoNLL '09
Date Added	6/21/2019, 1:35:08 PM
Modified	6/21/2019, 1:35:12 PM

Attachments

- Ratinov and Roth - 2009 - Design challenges and misconceptions in named enti.pdf

Dropout: A Simple Way to Prevent Neural Networks from Overfitting

Type	Journal Article
Author	Nitish Srivastava
Author	Geoffrey Hinton
Author	Alex Krizhevsky
Author	Ilya Sutskever
Author	Ruslan Salakhutdinov

Pages	30
Extra	12530
Library Catalog	Zotero
Language	en
Abstract	Deep neural nets with a large number of parameters are very powerful machine learning systems. However, overfitting is a serious problem in such networks. Large networks are also slow to use, making it difficult to deal with overfitting by combining the predictions of many different large neural nets at test time. Dropout is a technique for addressing this problem. The key idea is to randomly drop units (along with their connections) from the neural network during training. This prevents units from co-adapting too much. During training, dropout samples from an exponential number of different “thinned” networks. At test time, it is easy to approximate the effect of averaging the predictions of all these thinned networks by simply using a single unthinned network that has smaller weights. This significantly reduces overfitting and gives major improvements over other regularization methods. We show that dropout improves the performance of neural networks on supervised learning tasks in vision, speech recognition, document classification and computational biology, obtaining state-of-the-art results on many benchmark data sets.
Date Added	6/21/2019, 11:44:59 AM
Modified	6/21/2019, 11:45:03 AM

Tags:

⛔ No DOI found

Attachments

- Srivastava et al. - Dropout A Simple Way to Prevent Neural Networks f.pdf

---

Efficient Object Localization Using Convolutional Networks

Type	Journal Article
Author	Jonathan Tompson
Author	Ross Goroshin
Author	Arjun Jain
Author	Yann LeCun
Author	Christopher Bregler
URL	http://arxiv.org/abs/1411.4280
Publication	arXiv:1411.4280 [cs]
Date	2014-11-16
Extra	00392 arXiv: 1411.4280
Accessed	6/21/2019, 11:32:25 AM

Library Catalog arXiv.org
Language en
Abstract Recent state-of-the-art performance on human-body pose estimation has been achieved with Deep Convolutional Networks (ConvNets). Traditional ConvNet architectures include pooling and sub-sampling layers which reduce computational requirements, introduce invariance and prevent over-training. These benefits of pooling come at the cost of reduced localization accuracy. We introduce a novel architecture which includes an efficient 'position refinement' model that is trained to estimate the joint offset location within a small region of the image. This refinement model is jointly trained in cascade with a state-of-the-art ConvNet model [21] to achieve improved accuracy in human joint location estimation. We show that the variance of our detector approaches the variance of human annotations on the FLIC [20] dataset and outperforms all existing approaches on the MPII-human-pose dataset [1].
Date Added 6/21/2019, 11:32:25 AM
Modified 6/21/2019, 11:32:30 AM

Tags:

⛔ No DOI found, Computer Science - Computer Vision and Pattern Recognition

Notes:

Comment: 8 pages with 1 page of citations

Attachments

- Tompson et al. - 2014 - Efficient Object Localization Using Convolutional .pdf

---

Explaining Landscape Connectivity of Low-cost Solutions for Multilayer Nets

Type Journal Article
Author Rohith Kuditipudi
Author Xiang Wang
Author Holden Lee
Author Yi Zhang
Author Zhiyuan Li
Author Wei Hu
Author Sanjeev Arora
Author Rong Ge
URL http://arxiv.org/abs/1906.06247

Publication	arXiv:1906.06247 [cs, stat]
Date	2019-06-14
Extra	00000 arXiv: 1906.06247
Accessed	6/21/2019, 5:06:10 PM
Library Catalog	arXiv.org
Abstract	Mode connectivity is a surprising phenomenon in the loss landscape of deep nets. Optima---at least those discovered by gradient-based optimization---turn out to be connected by simple paths on which the loss function is almost constant. Often, these paths can be chosen to be piece-wise linear, with as few as two segments. We give mathematical explanations for this phenomenon, assuming generic properties (such as dropout stability and noise stability) of well-trained deep nets, which have previously been identified as part of understanding the generalization properties of deep nets. Our explanation holds for realistic multilayer nets, and experiments are presented to verify the theory.
Date Added	6/21/2019, 5:06:10 PM
Modified	6/21/2019, 5:06:30 PM

Tags:

Computer Science - Machine Learning, Statistics - Machine Learning

Attachments

- arXiv.org Snapshot
- Kuditipudi et al_2019_Explaining Landscape Connectivity of Low-cost Solutions for Multilayer Nets.pdf

Contents

Expressiveness of Rectifier Networks

Type	Journal Article
Author	Xingyuan Pan
Author	Vivek Srikumar
Pages	9
Extra	00022
Library Catalog	Zotero
Language	en
Abstract	Rectified Linear Units (ReLUs) have been shown to ameliorate the vanishing gradient problem, allow for efficient backpropagation, and empirically promote sparsity in the learned parameters. They have led to state-of-the-art results in a variety of applications. However, unlike threshold and sigmoid networks, ReLU networks are less explored from the perspective of their

expressiveness. This paper studies the expressiveness of ReLU networks. We characterize the decision boundary of two-layer ReLU networks by constructing functionally equivalent threshold networks. We show that while the decision boundary of a two-layer ReLU network can be captured by a threshold network, the latter may require an exponentially larger number of hidden units. We also formulate sufficient conditions for a corresponding logarithmic reduction in the number of hidden units to represent a sign network as a ReLU network. Finally, we experimentally compare threshold networks and their much smaller ReLU counterparts with respect to their ability to learn from synthetically generated data.

Date Added 6/21/2019, 11:43:18 AM
Modified 6/21/2019, 11:43:27 AM

Tags:

⛔ No DOI found

Attachments

- Pan and Srikumar - Expressiveness of Rectifier Networks.pdf

---

FCNN: Fourier Convolutional Neural Networks

Type Book Section
Editor Michelangelo Ceci
Editor Jaakko Hollmén
Editor Ljupčo Todorovski
Editor Celine Vens
Editor Sašo Džeroski
Author Harry Pratt
Author Bryan Williams
Author Frans Coenen
Author Yalin Zheng
URL http://link.springer.com/10.1007/978-3-319-71249-9_47
Volume 10534
Place Cham
Publisher Springer International Publishing
Pages 786-798
ISBN 978-3-319-71248-2 978-3-319-71249-9
Date 2017
Extra DOI: 10.1007/978-3-319-71249-9_47
Accessed 6/21/2019, 11:33:35 AM
Library Catalog DOI.org (Crossref)

Language en

Abstract The Fourier domain is used in computer vision and machine learning as image analysis tasks in the Fourier domain are analogous to spatial domain methods but are achieved using different operations. Convolutional Neural Networks (CNNs) use machine learning to achieve state-of-the-art results with respect to many computer vision tasks. One of the main limiting aspects of CNNs is the computational cost of updating a large number of convolution parameters. Further, in the spatial domain, larger images take exponentially longer than smaller image to train on CNNs due to the operations involved in convolution methods. Consequently, CNNs are often not a viable solution for large image computer vision tasks. In this paper a Fourier Convolution Neural Network (FCNN) is proposed whereby training is conducted entirely within the Fourier domain. The advantage offered is that there is a significant speed up in training time without loss of effectiveness. Using the proposed approach larger images can therefore be processed within viable computation time. The FCNN is fully described and evaluated. The evaluation was conducted using the benchmark Cifar10 and MNIST datasets, and a bespoke fundus retina image dataset. The results demonstrate that convolution in the Fourier domain gives a significant speed up without adversely affecting accuracy. For simplicity the proposed FCNN concept is presented in the context of a basic CNN architecture, however, the FCNN concept has the potential to improve the speed of any neural network system involving convolution.

Book Title Machine Learning and Knowledge Discovery in Databases

Short Title FCNN

Date Added 6/21/2019, 11:33:35 AM

Modified 6/21/2019, 11:33:35 AM

Attachments

- Pratt et al. - 2017 - FCNN Fourier Convolutional Neural Networks.pdf

---

Forgetting Memories and Their Attractiveness

Type Journal Article

Author Enzo Marinari

URL https://www.mitpressjournals.org/doi/10.1162/neco_a_01162

Volume 31

Issue 3

Pages 503-516

Publication Neural Computation

ISSN 0899-7667, 1530-888X

Date 03/2019

Extra 00001

Journal Abbr Neural Computation

DOI 10/gf35pq
Accessed 6/21/2019, 12:55:50 PM
Library Catalog DOI.org (Crossref)
Language en
Abstract We study numerically the memory that forgets, introduced in 1986 by Parisi by bounding the synaptic strength, with a mechanism that avoids confusion; allows remembering the pattern learned more recently; and has a physiologically very well-defined meaning. We analyze a number of features of this learning for a finite number of neurons and finite number of patterns. We discuss how the system behaves in the large but finite [Formula: see text] limit. We analyze the basin of attraction of the patterns that have been learned, and we show that it is exponentially small in the age of the pattern.
Date Added 6/21/2019, 12:55:50 PM
Modified 6/21/2019, 12:55:52 PM

Attachments

- Marinari - 2019 - Forgetting Memories and Their Attractiveness.pdf

---

Generative Adversarial Networks

Type Journal Article
Author Ian J. Goodfellow
Author Jean Pouget-Abadie
Author Mehdi Mirza
Author Bing Xu
Author David Warde-Farley
Author Sherjil Ozair
Author Aaron Courville
Author Yoshua Bengio
URL https://arxiv.org/abs/1406.2661v1
Date 2014/06/10
Extra 00485
Accessed 6/21/2019, 11:55:55 AM
Library Catalog arxiv.org
Language en
Abstract We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a

unique solution exists, with G recovering the training data distribution and D equal to 1/2 everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

Date Added 6/21/2019, 11:55:55 AM
Modified 6/21/2019, 12:05:26 PM

Attachments

- Goodfellow et al_2014_Generative Adversarial Networks.pdf
- Snapshot

Glove: Global Vectors for Word Representation

Type Conference Paper
Author Jeffrey Pennington
Author Richard Socher
Author Christopher Manning
URL http://aclweb.org/anthology/D14-1162
Place Doha, Qatar
Publisher Association for Computational Linguistics
Pages 1532-1543
Date 2014
Extra 08011
DOI 10/gfshwg
Accessed 6/21/2019, 11:34:26 AM
Library Catalog DOI.org (Crossref)
Conference Name Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)
Language en
Abstract Recent methods for learning vector space representations of words have succeeded in capturing fine-grained semantic and syntactic regularities using vector arithmetic, but the origin of these regularities has remained opaque. We analyze and make explicit the model properties needed for such regularities to emerge in word vectors. The result is a new global logbilinear regression model that combines the advantages of the two major model families in the literature: global matrix factorization and local context window methods. Our model efficiently leverages statistical information by training only on the nonzero elements in a word-word cooccurrence matrix, rather than on the entire sparse matrix or on individual context windows in a large corpus. The model produces a vector

	space with meaningful substructure, as evidenced by its performance of 75% on a recent word analogy task. It also outperforms related models on similarity tasks and named entity recognition.
Proceedings Title	Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)
Short Title	Glove
Date Added	6/21/2019, 11:34:26 AM
Modified	6/21/2019, 11:34:29 AM

Attachments

- Pennington et al. - 2014 - Glove Global Vectors for Word Representation.pdf

---

Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

Type	Journal Article
Author	Yonghui Wu
Author	Mike Schuster
Author	Zhifeng Chen
Author	Quoc V. Le
Author	Mohammad Norouzi
Author	Wolfgang Macherey
Author	Maxim Krikun
Author	Yuan Cao
Author	Qin Gao
Author	Klaus Macherey
Author	Jeff Klingner
Author	Apurva Shah
Author	Melvin Johnson
Author	Xiaobing Liu
Author	Łukasz Kaiser
Author	Stephan Gouws
Author	Yoshikiyo Kato
Author	Taku Kudo
Author	Hideto Kazawa
Author	Keith Stevens
Author	George Kurian
Author	Nishant Patil
Author	Wei Wang
Author	Cliff Young

Author	Jason Smith
Author	Jason Riesa
Author	Alex Rudnick
Author	Oriol Vinyals
Author	Greg Corrado
Author	Macduff Hughes
Author	Jeffrey Dean
URL	http://arxiv.org/abs/1609.08144
Publication	arXiv:1609.08144 [cs]
Date	2016-09-26
Extra	01464 arXiv: 1609.08144
Accessed	6/21/2019, 5:03:58 PM
Library Catalog	arXiv.org
Abstract	Neural Machine Translation (NMT) is an end-to-end learning approach for automated translation, with the potential to overcome many of the weaknesses of conventional phrase-based translation systems. Unfortunately, NMT systems are known to be computationally expensive both in training and in translation inference. Also, most NMT systems have difficulty with rare words. These issues have hindered NMT's use in practical deployments and services, where both accuracy and speed are essential. In this work, we present GNMT, Google's Neural Machine Translation system, which attempts to address many of these issues. Our model consists of a deep LSTM network with 8 encoder and 8 decoder layers using attention and residual connections. To improve parallelism and therefore decrease training time, our attention mechanism connects the bottom layer of the decoder to the top layer of the encoder. To accelerate the final translation speed, we employ low-precision arithmetic during inference computations. To improve handling of rare words, we divide words into a limited set of common sub-word units ("wordpieces") for both input and output. This method provides a good balance between the flexibility of "character"-delimited models and the efficiency of "word"-delimited models, naturally handles translation of rare words, and ultimately improves the overall accuracy of the system. Our beam search technique employs a length-normalization procedure and uses a coverage penalty, which encourages generation of an output sentence that is most likely to cover all the words in the source sentence. On the WMT'14 English-to-French and English-to-German benchmarks, GNMT achieves competitive results to state-of-the-art. Using a human side-by-side evaluation on a set of isolated simple sentences, it reduces translation errors by an average of 60% compared to Google's phrase-based production system.
Short Title	Google's Neural Machine Translation System
Date Added	6/21/2019, 5:03:58 PM
Modified	6/21/2019, 5:04:17 PM

Tags:

Computer Science - Computation and Language, Computer Science - Machine Learning, Computer Science - Artificial Intelligence

Attachments

- arXiv.org Snapshot
- Wu et al_2016_Google's Neural Machine Translation System.pdf

Contents

Graphical Models, Exponential Families, and Variational Inference

Type Journal Article
Author Martin J. Wainwright
Author Michael I. Jordan
URL http://www.nowpublishers.com/article/Details/MAL-001
Volume 1
Issue 1–2
Pages 1-305

Publication	Foundations and Trends® in Machine Learning
ISSN	1935-8237, 1935-8245
Date	2007
Extra	03181
Journal Abbr	FNT in Machine Learning
DOI	10/bpnwrm
Accessed	6/21/2019, 6:23:56 PM
Library Catalog	DOI.org (Crossref)
Language	en
Date Added	6/21/2019, 6:23:56 PM
Modified	6/21/2019, 6:23:58 PM

Attachments

- Wainwright and Jordan - 2007 - Graphical Models, Exponential Families, and Variat.pdf

---

Introduction to Information Retrieval

Type	Journal Article
Author	Christopher Manning
Author	Prabhakar Raghavan
Author	Hinrich Schuetze
Pages	581
Date	2009
Extra	16180
Library Catalog	Zotero
Language	en
Date Added	6/21/2019, 2:01:47 PM
Modified	6/21/2019, 2:01:50 PM

Tags:

⛔ No DOI found

Attachments

- Manning et al. - 2009 - Introduction to Information Retrieval.pdf

---

lecun-89.pdf

Type	Attachment

Date Added 6/21/2019, 11:45:20 AM
Modified 6/21/2019, 11:45:20 AM

Long Short-Term Memory

Type Journal Article
Author Sepp Hochreiter
Author Schmidhuber, Jurgen
Volume 9
Pages 1735-1780
Publication Neural Computation
Extra 19087
DOI 10/bxd65w
Date Added 6/21/2019, 12:51:09 PM
Modified 6/21/2019, 12:53:23 PM

Attachments

- Hochreiter and Schmidhuber, Jurgen - Long Short-Term Memory.pdf

---

Maxout Networks

Type Journal Article
Author Ian J. Goodfellow
Author David Warde-Farley
Author Mehdi Mirza
Author Aaron Courville
Author Yoshua Bengio
URL http://arxiv.org/abs/1302.4389
Publication arXiv:1302.4389 [cs, stat]
Date 2013-02-18
Extra 01610 arXiv: 1302.4389
Accessed 6/21/2019, 11:53:02 AM
Library Catalog arXiv.org
Abstract We consider the problem of designing models to leverage a recently introduced approximate model averaging technique called dropout. We define a simple new model called maxout (so named because its output is the max of a set of inputs, and because it is a natural companion to dropout) designed to both facilitate optimization by dropout and improve the accuracy of dropout's fast approximate model averaging technique. We empirically verify that the model successfully accomplishes both of these

	tasks. We use maxout and dropout to demonstrate state of the art classification performance on four benchmark datasets: MNIST, CIFAR-10, CIFAR-100, and SVHN.
Date Added	6/21/2019, 11:53:02 AM
Modified	6/21/2019, 11:54:10 AM

Tags:

Computer Science - Machine Learning, Statistics - Machine Learning

Notes:

Comment: This is the version of the paper that appears in ICML 2013

Attachments

- Goodfellow et al_2013_Maxout Networks.pdf

Neocognitron: A hierarchical neural network capable of visual pattern recognition

Type	Journal Article
Author	Kunihiko Fukushima
URL	https://linkinghub.elsevier.com/retrieve/pii/0893608088900147
Volume	1
Issue	2
Pages	119-130
Publication	Neural Networks
ISSN	08936080
Date	1/1988
Extra	00924
Journal Abbr	Neural Networks
DOI	10/bw8fwp
Accessed	6/21/2019, 11:33:40 AM
Library Catalog	DOI.org (Crossref)
Language	en
Abstract	A neural network modelfor visual pattern recognition, called the "neocognitron, "' was previously proposed by the author In this paper, we discuss the mechanism of the model in detail. In order to demonstrate the ability of the neocognitron, we also discuss a pattern-recognition system which works with the mechanism o f the neocognitron.
Short Title	Neocognitron

Date Added 6/21/2019, 11:33:40 AM
Modified 6/21/2019, 11:33:42 AM

Attachments

- Fukushima - 1988 - Neocognitron A hierarchical neural network capabl.pdf

---

Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position

Type Journal Article
Author Kunihiko Fukushima
URL http://link.springer.com/10.1007/BF00344251
Volume 36
Issue 4
Pages 193-202
Publication Biological Cybernetics
ISSN 0340-1200, 1432-0770
Date 4/1980
Extra 03322
Journal Abbr Biol. Cybernetics
DOI 10/fhrjjv
Accessed 6/21/2019, 11:33:47 AM
Library Catalog DOI.org (Crossref)
Language en
Abstract A neural network model for a mechanism of visual pattern recognition is proposed in this paper. The network is self-organized by "learning without a teacher", and acquires an ability to recognize stimulus patterns based on the geometrical similarity (Gestalt) of their shapes without affected by their positions. This network is given a nickname "neocognitron". After completion of self-organization, the network has a structure similar to the hierarchy model of the visual nervous system proposed by Hubel and Wiesel. The network consists of an input layer (photoreceptor array) followed by a cascade connection of a number of modular structures, each of which is composed of two layers of cells connected in a cascade. The first layer of each module consists of "S-cells', which show characteristics similar to simple cells or lower order hypercomplex cells, and the second layer consists of "C-cells" similar to complex cells or higher order hypercomplex cells. The afferent synapses to each S-cell have plasticity and are modifiable. The network has an ability of unsupervised learning: We do not need any "teacher" during the process of selforganization, and it is only needed to present a set of stimulus patterns repeatedly to the input layer of the network. The network has been simulated on a digital computer. After repetitive presentation of a set of stimulus patterns, each stimulus pattern

has become to elicit an output only from one of the C-cells of the last layer, and conversely, this C-cell has become selectively responsive only to that stimulus pattern. That is, none of the C-cells of the last layer responds to more than one stimulus pattern. The response of the C-cells of the last layer is not affected by the pattern's position at all. Neither is it affected by a small change in shape nor in size of the stimulus pattern.

Short Title Neocognitron
Date Added 6/21/2019, 11:33:47 AM
Modified 6/21/2019, 11:33:48 AM

Attachments

- Fukushima - 1980 - Neocognitron A self-organizing neural network mod.pdf

---

Neural Machine Translation by Jointly Learning to Align and Translate

Type Journal Article
Author Dzmitry Bahdanau
Author Kyunghyun Cho
Author Yoshua Bengio
URL http://arxiv.org/abs/1409.0473
Publication arXiv:1409.0473 [cs, stat]
Date 2014-09-01
Extra 07290 arXiv: 1409.0473
Accessed 6/21/2019, 5:03:31 PM
Library Catalog arXiv.org
Abstract Neural machine translation is a recently proposed approach to machine translation. Unlike the traditional statistical machine translation, the neural machine translation aims at building a single neural network that can be jointly tuned to maximize the translation performance. The models proposed recently for neural machine translation often belong to a family of encoder-decoders and consists of an encoder that encodes a source sentence into a fixed-length vector from which a decoder generates a translation. In this paper, we conjecture that the use of a fixed-length vector is a bottleneck in improving the performance of this basic encoder-decoder architecture, and propose to extend this by allowing a model to automatically (soft-)search for parts of a source sentence that are relevant to predicting a target word, without having to form these parts as a hard segment explicitly. With this new approach, we achieve a translation performance comparable to the existing state-of-the-art phrase-based system on the task of English-to-French translation. Furthermore, qualitative analysis reveals that the (soft-)alignments found by the model agree well with our intuition.
Date Added 6/21/2019, 5:03:31 PM
Modified 6/21/2019, 5:03:52 PM

Tags:

Computer Science - Computation and Language, Computer Science - Neural and Evolutionary Computing, Computer Science - Machine Learning, Statistics - Machine Learning

Notes:

Comment: Accepted at ICLR 2015 as oral presentation

Attachments

- arXiv.org Snapshot
- Bahdanau et al_2014_Neural Machine Translation by Jointly Learning to Align and Translate.pdf

Contents

Neural Transfer Learning for Natural Language Processing

Type	Journal Article
Author	Sebastian Ruder
Pages	329
Extra	00002
Library Catalog	Zotero
Language	en
Date Added	6/21/2019, 11:34:13 AM
Modified	6/21/2019, 11:34:18 AM

Tags:

⛔ No DOI found

Attachments

- Ruder - Neural Transfer Learning for Natural Language Proc.pdf

---

Pattern recognition and machine learning

Type	Book
Author	Christopher M. Bishop
Series	Information science and statistics
Place	New York
Publisher	Springer
ISBN	978-0-387-31073-2
Date	2006
Call Number	Q327 .B52 2006
Extra	36259
Library Catalog	Library of Congress ISBN
Language	en
# of Pages	738
Date Added	6/21/2019, 2:08:42 PM
Modified	6/21/2019, 2:08:43 PM

Tags:

Machine learning, Pattern perception

Attachments

- Bishop - 2006 - Pattern recognition and machine learning.pdf

Random Search for Hyper-Parameter Optimization

Type	Journal Article
Author	James Bergstra
Author	Yoshua Bengio
Pages	25
Extra	02538
Library Catalog	Zotero
Language	en
Abstract	Grid search and manual search are the most widely used strategies for hyper-parameter optimization. This paper shows empirically and theoretically that randomly chosen trials are more efficient for hyper-parameter optimization than trials on a grid. Empirical evidence comes from a comparison with a large previous study that used grid search and manual search to configure neural networks and deep belief networks. Compared with neural networks configured by a pure grid search, we find that random search over the same domain is able to find models that are as good or better within a small fraction of the computation time. Granting random search the same computational budget, random search finds better models by effectively searching a larger, less promising configuration space. Compared with deep belief networks configured by a thoughtful combination of manual search and grid search, purely random search over the same 32-dimensional configuration space found statistically equal performance on four of seven data sets, and superior performance on one of seven. A Gaussian process analysis of the function from hyper-parameters to validation set performance reveals that for most data sets only a few of the hyper-parameters really matter, but that different hyper-parameters are important on different data sets. This phenomenon makes grid search a poor choice for configuring algorithms for new data sets. Our analysis casts some light on why recent "High Throughput" methods achieve surprising success—they appear to search through a large number of hyper-parameters because most hyper-parameters do not matter much. We anticipate that growing interest in large hierarchical models will place an increasing burden on techniques for hyper-parameter optimization; this work shows that random search is a natural baseline against which to judge progress in the development of adaptive (sequential) hyper-parameter optimization algorithms.
Date Added	6/21/2019, 11:30:14 AM
Modified	6/21/2019, 11:31:18 AM

Tags:

⛔ No DOI found

Attachments

- Bergstra and Bengio - Random Search for Hyper-Parameter Optimization.pdf

Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network

Type	Journal Article
Author	Wenzhe Shi
Author	Jose Caballero
Author	Ferenc Huszár
Author	Johannes Totz
Author	Andrew P. Aitken
Author	Rob Bishop
Author	Daniel Rueckert
Author	Zehan Wang
URL	http://arxiv.org/abs/1609.05158
Publication	arXiv:1609.05158 [cs, stat]
Date	2016-09-16
Extra	00751 arXiv: 1609.05158
Accessed	6/21/2019, 4:49:34 PM
Library Catalog	arXiv.org
Abstract	Recently, several models based on deep neural networks have achieved great success in terms of both reconstruction accuracy and computational performance for single image super-resolution. In these methods, the low resolution (LR) input image is upscaled to the high resolution (HR) space using a single filter, commonly bicubic interpolation, before reconstruction. This means that the super-resolution (SR) operation is performed in HR space. We demonstrate that this is sub-optimal and adds computational complexity. In this paper, we present the first convolutional neural network (CNN) capable of real-time SR of 1080p videos on a single K2 GPU. To achieve this, we propose a novel CNN architecture where the feature maps are extracted in the LR space. In addition, we introduce an efficient sub-pixel convolution layer which learns an array of upscaling filters to upscale the final LR feature maps into the HR output. By doing so, we effectively replace the handcrafted bicubic filter in the SR pipeline with more complex upscaling filters specifically trained for each feature map, whilst also reducing the computational complexity of the overall SR operation. We evaluate the proposed approach using images and videos from publicly available datasets and show that it performs significantly better (+0.15dB on Images and +0.39dB on Videos) and is an order of magnitude faster than previous CNN-based methods.
Date Added	6/21/2019, 4:49:34 PM
Modified	6/21/2019, 4:49:40 PM

Tags:

Computer Science - Computer Vision and Pattern Recognition, Statistics - Machine Learning

Notes:

Comment: CVPR 2016 paper with updated affiliations and supplemental material, fixed typo in equation 4

Attachments

- arXiv.org Snapshot
- Shi et al_2016_Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel.pdf

---

Realistic Evaluation of Deep Semi-Supervised Learning Algorithms

Type	Journal Article
Author	Avital Oliver
Author	Augustus Odena
Author	Colin Raffel
Author	Ekin D. Cubuk
Author	Ian J. Goodfellow
URL	http://arxiv.org/abs/1804.09170
Publication	arXiv:1804.09170 [cs, stat]
Date	2018-04-24
Extra	00063 arXiv: 1804.09170
Accessed	6/21/2019, 11:55:34 AM
Library Catalog	arXiv.org
Abstract	Semi-supervised learning (SSL) provides a powerful framework for leveraging unlabeled data when labels are limited or expensive to obtain. SSL algorithms based on deep neural networks have recently proven successful on standard benchmark tasks. However, we argue that these benchmarks fail to address many issues that these algorithms would face in real-world applications. After creating a unified reimplementation of various widely-used SSL techniques, we test them in a suite of experiments designed to address these issues. We find that the performance of simple baselines which do not use unlabeled data is often underreported, that SSL methods differ in sensitivity to the amount of labeled and unlabeled data, and that performance can degrade substantially when the unlabeled dataset contains out-of-class examples. To help guide SSL research towards real-world applicability, we make our unified reimplemention and evaluation platform publicly available.
Date Added	6/21/2019, 11:55:34 AM
Modified	6/21/2019, 11:56:14 AM

Tags:

⛔ No DOI found, Computer Science - Machine Learning, Statistics - Machine Learning

Attachments

- Oliver et al_2018_Realistic Evaluation of Deep Semi-Supervised Learning Algorithms.pdf

---

Recurrent Convolutional Neural Networks for Scene Labeling

Type	Journal Article
Author	Pedro O Pinheiro
Author	Ronan Collobert
Pages	9
Extra	00563
Library Catalog	Zotero
Language	en
Abstract	The goal of the scene labeling task is to assign a class label to each pixel in an image. To ensure a good visual coherence and a high class accuracy, it is essential for a model to capture long range (pixel) label dependencies in images. In a feed-forward architecture, this can be achieved simply by considering a sufficiently large input context patch, around each pixel to be labeled. We propose an approach that consists of a recurrent convolutional neural network which allows us to consider a large input context while limiting the capacity of the model. Contrary to most standard approaches, our method does not rely on any segmentation technique nor any taskspecific features. The system is trained in an end-to-end manner over raw pixels, and models complex spatial dependencies with low inference cost. As the context size increases with the built-in recurrence, the system identifies and corrects its own errors. Our approach yields state-ofthe-art performance on both the Stanford Background Dataset and the SIFT Flow Dataset, while remaining very fast at test time.
Date Added	6/21/2019, 11:45:24 AM
Modified	6/21/2019, 11:45:30 AM

Tags:

⛔ No DOI found

Attachments

- Pinheiro and Collobert - Recurrent Convolutional Neural Networks for Scene .pdf

---

Show and Tell: A Neural Image Caption Generator

Type	Journal Article
Author	Oriol Vinyals
Author	Alexander Toshev
Author	Samy Bengio
Author	Dumitru Erhan
URL	http://arxiv.org/abs/1411.4555
Publication	arXiv:1411.4555 [cs]
Date	2014-11-17
Extra	02562 arXiv: 1411.4555
Accessed	6/21/2019, 12:44:10 PM
Library Catalog	arXiv.org
Abstract	Automatically describing the content of an image is a fundamental problem in artificial intelligence that connects computer vision and natural language processing. In this paper, we present a generative model based on a deep recurrent architecture that combines recent advances in computer vision and machine translation and that can be used to generate natural sentences describing an image. The model is trained to maximize the likelihood of the target description sentence given the training image. Experiments on several datasets show the accuracy of the model and the fluency of the language it learns solely from image descriptions. Our model is often quite accurate, which we verify both qualitatively and quantitatively. For instance, while the current state-of-the-art BLEU-1 score (the higher the better) on the Pascal dataset is 25, our approach yields 59, to be compared to human performance around 69. We also show BLEU-1 score improvements on Flickr30k, from 56 to 66, and on SBU, from 19 to 28. Lastly, on the newly released COCO dataset, we achieve a BLEU-4 of 27.7, which is the current state-of-the-art.
Short Title	Show and Tell
Date Added	6/21/2019, 12:44:10 PM
Modified	6/21/2019, 12:44:24 PM

Tags:

Computer Science - Computer Vision and Pattern Recognition

Attachments

- arXiv.org Snapshot
- Vinyals et al_2014_Show and Tell.pdf

Contents

Show and Tell: Lessons learned from the 2015 MSCOCO Image Captioning Challenge

Type	Journal Article
Author	Oriol Vinyals
Author	Alexander Toshev
Author	Samy Bengio
Author	Dumitru Erhan
URL	http://arxiv.org/abs/1609.06647
Volume	39
Issue	4
Pages	652-663
Publication	IEEE Transactions on Pattern Analysis and Machine Intelligence
ISSN	0162-8828, 2160-9292
Date	2017-4-1
Extra	00281 arXiv: 1609.06647
Journal Abbr	IEEE Trans. Pattern Anal. Mach. Intell.
DOI	10/f9xnjx

Accessed 6/21/2019, 11:30:39 AM
Library Catalog arXiv.org
Language en
Abstract Automatically describing the content of an image is a fundamental problem in artificial intelligence that connects computer vision and natural language processing. In this paper, we present a generative model based on a deep recurrent architecture that combines recent advances in computer vision and machine translation and that can be used to generate natural sentences describing an image. The model is trained to maximize the likelihood of the target description sentence given the training image. Experiments on several datasets show the accuracy of the model and the fluency of the language it learns solely from image descriptions. Our model is often quite accurate, which we verify both qualitatively and quantitatively. Finally, given the recent surge of interest in this task, a competition was organized in 2015 using the newly released COCO dataset. We describe and analyze the various improvements we applied to our own baseline and show the resulting performance in the competition, which we won ex-aequo with a team from Microsoft Research, and provide an open source implementation in TensorFlow.
Short Title Show and Tell
Date Added 6/21/2019, 11:30:39 AM
Modified 6/21/2019, 11:30:43 AM

Tags:

Computer Science - Computer Vision and Pattern Recognition

Notes:

Comment: arXiv admin note: substantial text overlap with arXiv:1411.4555

Attachments

- Vinyals et al. - 2017 - Show and Tell Lessons learned from the 2015 MSCOC.pdf

---

State-Space Representations of Deep Neural Networks

Type Journal Article
Author Michael Hauser
Author Sean Gunn
Author Samer Saab
Author Asok Ray
URL https://www.mitpressjournals.org/doi/10.1162/neco_a_01165
Volume 31

Issue 3
Pages 538-554
Publication Neural Computation
ISSN 0899-7667, 1530-888X
Date 03/2019
Extra 00001
Journal Abbr Neural Computation
DOI 10/gf35pr
Accessed 6/21/2019, 12:55:58 PM
Library Catalog DOI.org (Crossref)
Language en
Abstract This letter deals with neural networks as dynamical systems governed by finite difference equations. It shows that the introduction of [Formula: see text]-many skip connections into network architectures, such as residual networks and additive dense networks, defines [Formula: see text]th order dynamical equations on the layer-wise transformations. Closed-form solutions for the state-space representations of general [Formula: see text]th order additive dense networks, where the concatenation operation is replaced by addition, as well as [Formula: see text]th order smooth networks, are found. The developed provision endows deep neural networks with an algebraic structure. Furthermore, it is shown that imposing [Formula: see text]th order smoothness on network architectures with [Formula: see text]-many nodes per layer increases the state-space dimension by a multiple of [Formula: see text], and so the effective embedding dimension of the data manifold by the neural network is [Formula: see text]-many dimensions. It follows that network architectures of these types reduce the number of parameters needed to maintain the same embedding dimension by a factor of [Formula: see text] when compared to an equivalent first-order, residual network. Numerical simulations and experiments on CIFAR10, SVHN, and MNIST have been conducted to help understand the developed theory and efficacy of the proposed concepts.
Date Added 6/21/2019, 12:55:58 PM
Modified 6/21/2019, 12:55:59 PM

Attachments

- Hauser et al. - 2019 - State-Space Representations of Deep Neural Network.pdf

---

Statistical learning with sparsity: the lasso and generalizations

Type Book
Author Trevor Hastie
Author Robert Tibshirani
Author Martin Wainwright

Series	Monographs on statistics and applied probability
Place	Boca Raton
Publisher	CRC Press, Taylor & Francis Group
ISBN	978-1-4987-1216-3
Date	2015
Call Number	QA275 .H38 2015
Extra	00982
Series Number	143
Library Catalog	Library of Congress ISBN
Short Title	Statistical learning with sparsity
# of Pages	351
Date Added	6/21/2019, 6:33:56 PM
Modified	6/21/2019, 6:33:58 PM

Tags:

Least squares, Linear models (Statistics), Mathematical statistics, Proof theory

Notes:

"A Chapman & Hall book."

Attachments

- Hastie et al. - 2015 - Statistical learning with sparsity the lasso and .pdf

---

T-CGAN: Conditional Generative Adversarial Network for Data Augmentation in Noisy Time Series with Irregular Sampling

Type	Journal Article
Author	Giorgia Ramponi
Author	Pavlos Protopapas
Author	Marco Brambilla
Author	Ryan Janssen
URL	http://arxiv.org/abs/1811.08295
Publication	arXiv:1811.08295 [cs, stat]
Date	2018-11-20
Extra	00001 arXiv: 1811.08295
Accessed	6/21/2019, 11:44:10 AM
Library Catalog	arXiv.org
Language	en

Abstract	In this paper we propose a data augmentation method for time series with irregular sampling, Time-Conditional Generative Adversarial Network (T-CGAN). Our approach is based on Conditional Generative Adversarial Networks (CGAN), where the generative step is implemented by a deconvolutional NN and the discriminative step by a convolutional NN. Both the generator and the discriminator are conditioned on the sampling timestamps, to learn the hidden relationship between data and timestamps, and consequently to generate new time series.
Short Title	T-CGAN
Date Added	6/21/2019, 11:44:10 AM
Modified	6/21/2019, 11:44:12 AM

Tags:

⛔ No DOI found, Computer Science - Machine Learning, Statistics - Machine Learning

Attachments

- Ramponi et al. - 2018 - T-CGAN Conditional Generative Adversarial Network.pdf

---

The elements of statistical learning: data mining, inference, and prediction

Type	Book
Author	Trevor Hastie
Author	Robert Tibshirani
Author	J. H. Friedman
Series	Springer series in statistics
Edition	2nd ed
Place	New York, NY
Publisher	Springer
ISBN	978-0-387-84857-0 978-0-387-84858-7
Date	2009
Call Number	Q325.5 .H39 2009
Extra	00355
Library Catalog	Library of Congress ISBN
Short Title	The elements of statistical learning
# of Pages	745
Date Added	6/21/2019, 6:37:21 PM
Modified	6/21/2019, 6:37:24 PM

Tags:

Machine learning, Bioinformatics, Computational intelligence, Data mining, Forecasting, Inference, Methodology, Statistics

Attachments

- Hastie et al. - 2009 - The elements of statistical learning data mining,.pdf

---

The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances

Type	Journal Article
Author	Anthony Bagnall
Author	Jason Lines
Author	Aaron Bostrom
Author	James Large
Author	Eamonn Keogh
URL	http://link.springer.com/10.1007/s10618-016-0483-9
Volume	31
Issue	3
Pages	606-660
Publication	Data Mining and Knowledge Discovery
ISSN	1384-5810, 1573-756X
Date	5/2017
Journal Abbr	Data Min Knowl Disc
DOI	10/f9586n
Accessed	6/21/2019, 11:44:25 AM
Library Catalog	DOI.org (Crossref)
Language	en
Abstract	In the last 5 years there have been a large number of new time series classification algorithms proposed in the literature. These algorithms have been evaluated on subsets of the 47 data sets in the University of California, Riverside time series classification archive. The archive has recently been expanded to 85 data sets, over half of which have been donated by researchers at the University of East Anglia. Aspects of previous evaluations have made comparisons between algorithms difficult. For example, several different programming languages have been used, experiments involved a single train/test split and some used normalised data whilst others did not. The relaunch of the archive provides a timely opportunity to thoroughly evaluate algorithms on a larger number of datasets. We have implemented 18 recently proposed algorithms in a common Java framework and compared them against two standard benchmark classifiers (and each other) by performing 100 resampling experiments on each of the 85 datasets. We use these results to test several hypotheses relating to whether the algorithms

	are significantly more accurate than the benchmarks and each other. Our results indicate that only nine of these algorithms are significantly more accurate than both benchmarks and that one classifier, the collective of transformation ensembles, is significantly more accurate than all of the others. All of our experiments and results are reproducible: we release all of our code, results and experimental details and we hope these experiments form the basis for more robust testing of new algorithms in the future.
Short Title	The great time series classification bake off
Date Added	6/21/2019, 11:44:25 AM
Modified	6/21/2019, 11:44:28 AM

Attachments

- Bagnall et al. - 2017 - The great time series classification bake off a r.pdf

---

U-Net: Convolutional Networks for Biomedical Image Segmentation

Type	Journal Article
Author	Olaf Ronneberger
Author	Philipp Fischer
Author	Thomas Brox
URL	http://arxiv.org/abs/1505.04597
Publication	arXiv:1505.04597 [cs]
Date	2015-05-18
Extra	06228 arXiv: 1505.04597
Accessed	6/21/2019, 4:42:59 PM
Library Catalog	arXiv.org
Abstract	There is large consent that successful training of deep networks requires many thousand annotated training samples. 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. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. 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 512x512 image takes less than a second on a recent GPU. The full implementation (based on Caffe) and the trained networks are available at http://lmb.informatik.uni-freiburg.de/people /ronneber/u-net .
Short Title	U-Net
Date Added	6/21/2019, 4:42:59 PM

Modified 6/21/2019, 4:43:04 PM

Tags:

Computer Science - Computer Vision and Pattern Recognition

Notes:

Comment: conditionally accepted at MICCAI 2015

Attachments

- arXiv.org Snapshot
- Ronneberger et al_2015_U-Net.pdf

Contents

U-Net: Convolutional Networks for Biomedical Image Segmentation

---

Very Deep Convolutional Networks for Large-Scale Image Recognition

Type Journal Article
Author Karen Simonyan
Author Andrew Zisserman
URL http://arxiv.org/abs/1409.1556
Publication arXiv:1409.1556 [cs]
Date 2014-09-04
Extra 23902 arXiv: 1409.1556
Accessed 6/21/2019, 11:33:56 AM
Library Catalog arXiv.org
Language en
Abstract In this work we investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth using an architecture with very small (3×3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers. These findings were the basis of our ImageNet Challenge 2014 submission, where our team secured the first and the second places in the localisation and classification tracks respectively. We also show that our representations generalise well to other datasets, where they achieve state-of-the-art results. We have made our two best-performing ConvNet models publicly available to facilitate further research on the use of deep visual representations in computer vision.
Date Added 6/21/2019, 11:33:56 AM

Modified 6/21/2019, 11:34:01 AM

Tags:

⛔ No DOI found, Computer Science - Computer Vision and Pattern Recognition

Attachments

- Simonyan and Zisserman - 2014 - Very Deep Convolutional Networks for Large-Scale I.pdf

---

Visualizing and Understanding Convolutional Networks

Type	Book Section
Editor	David Fleet
Editor	Tomas Pajdla
Editor	Bernt Schiele
Editor	Tinne Tuytelaars
Author	Matthew D. Zeiler
Author	Rob Fergus
URL	http://link.springer.com/10.1007/978-3-319-10590-1_53
Volume	8689
Place	Cham
Publisher	Springer International Publishing
Pages	818-833
ISBN	978-3-319-10589-5 978-3-319-10590-1
Date	2014
Extra	00000 DOI: 10.1007/978-3-319-10590-1_53
Accessed	6/21/2019, 11:34:07 AM
Library Catalog	DOI.org (Crossref)
Language	en
Abstract	Large Convolutional Network models have recently demonstrated impressive classification performance on the ImageNet benchmark Krizhevsky et al. [18]. However there is no clear understanding of why they perform so well, or how they might be improved. In this paper we explore both issues. We introduce a novel visualization technique that gives insight into the function of intermediate feature layers and the operation of the classifier. Used in a diagnostic role, these visualizations allow us to find model architectures that outperform Krizhevsky et al. on the ImageNet classification benchmark. We also perform an ablation study to discover the performance contribution from different model layers. We show our ImageNet model generalizes well to other datasets: when the softmax classifier is retrained, it convincingly beats the current state-ofthe-art results on Caltech-101 and Caltech-256 datasets.

Book Title Computer Vision – ECCV 2014
Date Added 6/21/2019, 11:34:07 AM
Modified 6/21/2019, 11:34:09 AM

Attachments

- Zeiler and Fergus - 2014 - Visualizing and Understanding Convolutional Networ.pdf

---

Visualizing the Loss Landscape of Neural Nets

Type Journal Article
Author Hao Li
Author Zheng Xu
Author Gavin Taylor
Author Christoph Studer
Author Tom Goldstein
URL http://arxiv.org/abs/1712.09913
Publication arXiv:1712.09913 [cs, stat]
Date 2017-12-28
Extra 00129 arXiv: 1712.09913
Accessed 6/21/2019, 1:05:39 PM
Library Catalog arXiv.org
Abstract Neural network training relies on our ability to find "good" minimizers of highly non-convex loss functions. It is well-known that certain network architecture designs (e.g., skip connections) produce loss functions that train easier, and well-chosen training parameters (batch size, learning rate, optimizer) produce minimizers that generalize better. However, the reasons for these differences, and their effects on the underlying loss landscape, are not well understood. In this paper, we explore the structure of neural loss functions, and the effect of loss landscapes on generalization, using a range of visualization methods. First, we introduce a simple "filter normalization" method that helps us visualize loss function curvature and make meaningful side-by-side comparisons between loss functions. Then, using a variety of visualizations, we explore how network architecture affects the loss landscape, and how training parameters affect the shape of minimizers.
Date Added 6/21/2019, 1:05:39 PM
Modified 6/21/2019, 1:06:23 PM

Tags:

Computer Science - Computer Vision and Pattern Recognition, Computer Science - Machine Learning, Statistics - Machine Learning

Notes:

Comment: NIPS 2018 (extended version, 10.5 pages), code is available at https://github.com/tomgoldstein/loss-landscape

Attachments

- arXiv.org Snapshot
- Li et al_2017_Visualizing the Loss Landscape of Neural Nets.pdf

Contents

What is the Role of Recurrent Neural Networks (RNNs) in an Image Caption Generator?

Type	Journal Article
Author	Marc Tanti
Author	Albert Gatt
Author	Kenneth P. Camilleri
URL	http://arxiv.org/abs/1708.02043
Publication	arXiv:1708.02043 [cs]
Date	2017-08-07

Extra	00005 arXiv: 1708.02043
Accessed	6/21/2019, 12:44:55 PM
Library Catalog	arXiv.org
Abstract	In neural image captioning systems, a recurrent neural network (RNN) is typically viewed as the primary `generation' component. This view suggests that the image features should be `injected' into the RNN. This is in fact the dominant view in the literature. Alternatively, the RNN can instead be viewed as only encoding the previously generated words. This view suggests that the RNN should only be used to encode linguistic features and that only the final representation should be `merged' with the image features at a later stage. This paper compares these two architectures. We find that, in general, late merging outperforms injection, suggesting that RNNs are better viewed as encoders, rather than generators.
Date Added	6/21/2019, 12:44:55 PM
Modified	6/21/2019, 12:45:09 PM

Tags:

Computer Science - Computation and Language, Computer Science - Neural and Evolutionary Computing, Computer Science - Computer Vision and Pattern Recognition

Notes:

Comment: Appears in: Proceedings of the 10th International Conference on Natural Language Generation (INLG'17)

Attachments

- arXiv.org Snapshot
- Tanti et al_2017_What is the Role of Recurrent Neural Networks (RNNs) in an Image Caption.pdf

---

Where to put the Image in an Image Caption Generator

Type	Journal Article
Author	Marc Tanti
Author	Albert Gatt
Author	Kenneth P. Camilleri
URL	http://arxiv.org/abs/1703.09137
Volume	24
Issue	3
Pages	467-489
Publication	Natural Language Engineering

ISSN 1351-3249, 1469-8110
Date 05/2018
Extra 00015 arXiv: 1703.09137
Journal Abbr Nat. Lang. Eng.
DOI 10/gf35pm
Accessed 6/21/2019, 12:42:39 PM
Library Catalog arXiv.org
Abstract When a recurrent neural network language model is used for caption generation, the image information can be fed to the neural network either by directly incorporating it in the RNN -- conditioning the language model by `injecting' image features -- or in a layer following the RNN -- conditioning the language model by `merging' image features. While both options are attested in the literature, there is as yet no systematic comparison between the two. In this paper we empirically show that it is not especially detrimental to performance whether one architecture is used or another. The merge architecture does have practical advantages, as conditioning by merging allows the RNN's hidden state vector to shrink in size by up to four times. Our results suggest that the visual and linguistic modalities for caption generation need not be jointly encoded by the RNN as that yields large, memory-intensive models with few tangible advantages in performance; rather, the multimodal integration should be delayed to a subsequent stage.
Date Added 6/21/2019, 12:42:39 PM
Modified 6/21/2019, 12:42:42 PM

Tags:

Computer Science - Computation and Language, Computer Science - Neural and Evolutionary Computing, Computer Science - Computer Vision and Pattern Recognition

Notes:

Comment: Accepted in JNLE Special Issue: Language for Images (24.3) (expanded with content that was removed from journal paper in order to reduce number of pages), 28 pages, 5 figures, 6 tables

Attachments

- arXiv.org Snapshot
- Tanti et al_2018_Where to put the Image in an Image Caption Generator.pdf