Papers

Advanced Topics in High-Performance Computing (MCSC 6230G/7230G)

Faisal Qureshi

http://faculty.uoit.ca/qureshi

Computer Science, Faculty of Science

University of Ontario Institute of Technology

26 September 2017

Machine Learning Papers

Courtesy: http://www.kdnuggets.com/2017/04/top-20-papers-machine-learning.html and Tony Joseph for useful comments.

Dropout: a simple way to prevent neural networks from overfitting, by Hinton, G.E., Krizhevsky, A., Srivastava, N., Sutskever, I., & Salakhutdinov, R. (2014). Journal of Machine Learning Research, 15, 1929-1958. (cited 2084 times, HIC: 142 , CV: 536).

Summary

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. This significantly reduces overfitting and gives major improvements over other regularization methods

Deep Residual Learning for Image Recognition, by He, K., Ren, S., Sun, J., & Zhang, X. (2016). CoRR, abs/1512.03385. (cited 1436 times, HIC: 137 , CV: 582).

Summary

We present a residual learning framework to ease the training of deep neural networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. We provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth.

Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, by Sergey Ioffe, Christian Szegedy (2015) ICML. (cited 946 times, HIC: 56 , CV: 0).

Summary

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. We refer to this phenomenon as internal covariate shift, and address the problem by normalizing layer inputs. Applied to a state-of-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.

Large-Scale Video Classification with Convolutional Neural Networks , by Fei-Fei, L., Karpathy, A., Leung, T., Shetty, S., Sukthankar, R., & Toderici, G. (2014). IEEE Conference on Computer Vision and Pattern Recognition (cited 865 times, HIC: 24 , CV: 239)

Summary

Convolutional Neural Networks (CNNs) have been established as a powerful class of models for image recognition problems. Encouraged by these results, we provide an extensive empirical evaluation of CNNs on large-scale video classification using a new dataset of 1 million YouTube videos belonging to 487 classes .

Learning deep features for scene recognition using places database , by Lapedriza, À., Oliva, A., Torralba, A., Xiao, J., & Zhou, B. (2014). NIPS. (cited 644 times, HIC: 65 , CV: 0)

Summary

We introduce a new scene-centric database called Places with over 7 million labeled pictures of scenes. We propose new methods to compare the density and diversity of image datasets and show that Places is as dense as other scene datasets and has more diversity.

Generative adversarial nets, by Bengio, Y., Courville, A.C., Goodfellow, I.J., Mirza, M., Ozair, S., Pouget-Abadie, J., Warde-Farley, D., & Xu, B. (2014) NIPS. (cited 463 times, HIC: 55 , CV: 0)

Summary

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.

High-Speed Tracking with Kernelized Correlation Filters, by Batista, J., Caseiro, R., Henriques, J.F., & Martins, P. (2015). CoRR, abs/1404.7584. (cited 439 times, HIC: 43 , CV: 0)

Summary

In most modern trackers, to cope with natural image changes, a classifier is typically trained with translated and scaled sample patches. We propose an analytic model for datasets of thousands of translated patches. By showing that the resulting data matrix is circulant, we can diagonalize it with the discrete Fourier transform, reducing both storage and computation by several orders of magnitude.

A Review on Multi-Label Learning Algorithms, by Zhang, M., & Zhou, Z. (2014). IEEE TKDE, (cited 436 times, HIC: 7 , CV: 91)

Summary

This paper aims to provide a timely review on multi-label learning studies the problem where each example is represented by a single instance while associated with a set of labels simultaneously.

How transferable are features in deep neural networks, by Bengio, Y., Clune, J., Lipson, H., & Yosinski, J. (2014) CoRR, abs/1411.1792. (cited 402 times, HIC: 14 , CV: 0)

Summary

We experimentally quantify the generality versus specificity of neurons in each layer of a deep convolutional neural network and report a few surprising results. Transferability is negatively affected by two distinct issues: (1) the specialization of higher layer neurons to their original task at the expense of performance on the target task, which was expected, and (2) optimization difficulties related to splitting networks between co-adapted neurons, which was not expected.

Do we need hundreds of classifiers to solve real world classification problems, by Amorim, D.G., Barro, S., Cernadas, E., & Delgado, M.F. (2014). Journal of Machine Learning Research (cited 387 times, HIC: 3 , CV: 0)

Summary

We evaluate 179 classifiers arising from 17 families (discriminant analysis, Bayesian, neural networks, support vector machines, decision trees, rule-based classifiers, boosting, bagging, stacking, random forests and other ensembles, generalized linear models, nearest-neighbors, partial least squares and principal component regression, logistic and multinomial regression, multiple adaptive regression splines and other methods). We use 121 data sets from UCI data base to study the classifier behavior, not dependent on the data set collection. The winners are the random forest (RF) versions implemented in R and accessed via caret) and the SVM with Gaussian kernel implemented in C using LibSVM.

Scalable Nearest Neighbor Algorithms for High Dimensional Data, by Lowe, D.G., & Muja, M. (2014). IEEE Trans. Pattern Anal. Mach. Intell., (cited 324 times, HIC: 11 , CV: 69).

Summary

We propose new algorithms for approximate nearest neighbor matching and evaluate and compare them with previous algorithms. In order to scale to very large data sets that would otherwise not fit in the memory of a single machine, we propose a distributed nearest neighbor matching framework that can be used with any of the algorithms described in the paper.

Trends in extreme learning machines: a review, by Huang, G., Huang, G., Song, S., & You, K. (2015). Neural Networks, (cited 323 times, HIC: 0 , CV: 0)

Summary

We aim to report the current state of the theoretical research and practical advances on Extreme learning machine (ELM). Apart from classification and regression, ELM has recently been extended for clustering, feature selection, representational learning and many other learning tasks. Due to its remarkable efficiency, simplicity, and impressive generalization performance, ELM have been applied in a variety of domains, such as biomedical engineering, computer vision, system identification, and control and robotics.

A survey on concept drift adaptation, by Bifet, A., Bouchachia, A., Gama, J., Pechenizkiy, M., & Zliobaite, I. ACM Comput. Surv., 2014 , (cited 314 times, HIC: 4 , CV: 23)

Summary

This work aims at providing a comprehensive introduction to the concept drift adaptation that refers to an online supervised learning scenario when the relation between the input data and the target variable changes over time.

Multi-scale Orderless Pooling of Deep Convolutional Activation Features, by Gong, Y., Guo, R., Lazebnik, S., & Wang, L. (2014). ECCV(cited 293 times, HIC: 23 , CV: 95)

Summary

To improve the invariance of CNN activations without degrading their discriminative power, this paper presents a simple but effective scheme called multi-scale orderless pooling (MOP-CNN).

Simultaneous Detection and Segmentation, by Arbeláez, P.A., Girshick, R.B., Hariharan, B., & Malik, J. (2014) ECCV , (cited 286 times, HIC: 23 , CV: 94)

Summary

We aim to detect all instances of a category in an image and, for each instance, mark the pixels that belong to it. We call this task Simultaneous Detection and Segmentation (SDS).

A survey on feature selection methods, by Chandrashekar, G., & Sahin, F. Int. J. on Computers & Electrical Engineering, (cited 279 times, HIC: 1 , CV: 58)

Summary

Plenty of feature selection methods are available in literature due to the availability of data with hundreds of variables leading to data with very high dimension.

One Millisecond Face Alignment with an Ensemble of Regression Trees, by Kazemi, Vahid, and Josephine Sullivan, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014, (cited 277 times, HIC: 15 , CV: 0)

Summary

This paper addresses the problem of Face Alignment for a single image. We show how an ensemble of regression trees can be used to estimate the face's landmark positions directly from a sparse subset of pixel intensities, achieving super-realtime performance with high quality predictions.

A survey of multiple classifier systems as hybrid systems , by Corchado, E., Graña, M., & Wozniak, M. (2014). Information Fusion, 16, 3-17. (cited 269 times, HIC: 1 , CV: 22)

Summary

A current focus of intense research in pattern classification is the combination of several classifier systems, which can be built following either the same or different models and/or datasets building.

Deformable Convolutional Networks

https://arxiv.org/abs/1703.06211

Abstract

Convolutional neural networks (CNNs) are inherently limited to model geometric transformations due to the fixed geometric structures in its building modules. In this work, we introduce two new modules to enhance the transformation modeling capacity of CNNs, namely, deformable convolution and deformable RoI pooling. Both are based on the idea of augmenting the spatial sampling locations in the modules with additional offsets and learning the offsets from target tasks, without additional supervision. The new modules can readily replace their plain counterparts in existing CNNs and can be easily trained end-to-end by standard back-propagation, giving rise to deformable convolutional networks. Extensive experiments validate the effectiveness of our approach on sophisticated vision tasks of object detection and semantic segmentation. The code would be released.

Fully convolutional networks for semantic segmentation

https://people.eecs.berkeley.edu/~jonlong/long_shelhamer_fcn.pdf

Abstract

Convolutional networks are powerful visual models that yield hierarchies of features. We show that convolu- tional networks by themselves, trained end-to-end, pixels- to-pixels, exceed the state-of-the-art in semantic segmen- tation. Our key insight is to build "fully convolutional" networks that take input of arbitrary size and produce correspondingly-sized output with efficient inference and learning. We define and detail the space of fully convolu- tional networks, explain their application to spatially dense prediction tasks, and draw connections to prior models. We adapt contemporary classification networks (AlexNet [22], the VGG net [34], and GoogLeNet [35]) into fully convolu- tional networks and transfer their learned representations by fine-tuning [5] to the segmentation task. We then define a skip architecture that combines semantic information from a deep, coarse layer with appearance information from a shallow, fine layer to produce accurate and detailed seg- mentations. Our fully convolutional network achieves state- of-the-art segmentation of PASCAL VOC (20% relative improvement to 62.2% mean IU on 2012), NYUDv2, and SIFT Flow, while inference takes less than one fifth of a second for a typical image.

InterpoNet, A brain inspired neural network for optical flow dense interpolation

https://arxiv.org/abs/1611.09803

Abstract

Sparse-to-dense interpolation for optical flow is a fundamental phase in the pipeline of most of the leading optical flow estimation algorithms. The current state-of-the-art method for interpolation, EpicFlow, is a local average method based on an edge aware geodesic distance. We propose a new data-driven sparse-to-dense interpolation algorithm based on a fully convolutional network. We draw inspiration from the filling-in process in the visual cortex and introduce lateral dependencies between neurons and multi-layer supervision into our learning process. We also show the importance of the image contour to the learning process. Our method is robust and outperforms EpicFlow on competitive optical flow benchmarks with several underlying matching algorithms. This leads to state-of-the-art performance on the Sintel and KITTI 2012 benchmarks.

Deep Roots: Improving CNN Efficiency with Hierarchical Filter Groups

https://arxiv.org/abs/1605.06489

Abstract

We propose a new method for creating computationally efficient and compact convolutional neural networks (CNNs) using a novel sparse connection structure that resembles a tree root. This allows a significant reduction in computational cost and number of parameters compared to state-of-the-art deep CNNs, without compromising accuracy, by exploiting the sparsity of inter-layer filter dependencies. We validate our approach by using it to train more efficient variants of state-of-the-art CNN architectures,

evaluated on the CIFAR10 and ILSVRC datasets. Our results show similar or higher accuracy than the baseline architectures with much less computation, as measured by CPU and GPU timings. For example, for ResNet 50, our model has 40% fewer parameters, 45% fewer floating point operations, and is 31% (12%) faster on a CPU (GPU). For the deeper ResNet 200 our model has 25% fewer floating point operations and 44% fewer parameters, while maintaining state-of-the-art accuracy. For GoogLeNet, our model has 7% fewer parameters and is 21% (16%) faster on a CPU (GPU).