Showing 1 – 152 of 152 results for all: tensorflow

arXiv:1804.05879
[pdf, other] cs.CV

M-PACT: Michigan Platform for Activity Classification in Tensorflow

Authors: Eric Hofesmann, Madan Ravi Ganesh, Jason J Corso

Abstract: …unified platform capable of delivering results while removing the burden of developing an entire system cannot be overstated. To try and overcome these issues, we develop a tensorflow-based unified platform to abstract away unnecessary overheads in terms of an end-to-end pipeline setup in order to allow the user to quickly and easily prototype action classif… ▽ More

Submitted 16 April, 2018; originally announced April 2018.

arXiv:1804.05497
[pdf, other] cs.LG

Deep Learning on Key Performance Indicators for Predictive Maintenance in SAP HANA

Authors: Jaekoo Lee, Byunghan Lee, Jongyoon Song, Jaesik Yoon, Yongsik Lee, Donghun Lee, Sungroh Yoon

Abstract: …historical data, whereas a spatial learning approach is used to classify known anomalies based on labeled data. We implement a system in SAP HANA integrated with Google TensorFlow. The experimental results with real-world data confirm the effectiveness of the system and models. ▽ More

Submitted 15 April, 2018; originally announced April 2018.

arXiv:1804.05283
[pdf] stat.ML

OmicsMapNet: Transforming omics data to take advantage of Deep Convolutional Neural Network for discovery

Authors: Shiyong Ma, Zhen Zhang

Abstract: …glioma samples as treemaps to capture the functional hierarchical structure of genes in 2D images. Deep Convolutional Neural Networks (CNN) were derived using tools from TensorFlow to learn the grade of TCGA LGG and GBM samples with relatively high accuracy. The most contributory features in the trained CNN were confirmed in pathway analysis for their plausi… ▽ More

Submitted 14 April, 2018; originally announced April 2018.

arXiv:1804.04806
[pdf, other] cs.LG

μ-cuDNN: Accelerating Deep Learning Frameworks with Micro-Batching

Authors: Yosuke Oyama, Tal Ben-Nun, Torsten Hoefler, Satoshi Matsuoka

Abstract: … unchanged, so that it decouples statistical efficiency from the hardware efficiency safely. We demonstrate the effectiveness of {\mu}-cuDNN over two frameworks, Caffe and TensorFlow, achieving speedups of 1.63x for AlexNet and 1.21x for ResNet-18 on P100-SXM2 GPU. These results indicate that using micro-batches can seamlessly increase the performance of deep… ▽ More

Submitted 13 April, 2018; originally announced April 2018.

Comments: 11 pages, 14 figures. Part of the content have been published in IPSJ SIG Technical Report, Vol. 2017-HPC-162, No. 22, pp. 1-9, 2017. (DOI: http://id.nii.ac.jp/1001/00184814)

ACM Class: I.2.6

arXiv:1804.03159
[pdf, other] quant-ph

Strawberry Fields: A Software Platform for Photonic Quantum Computing

Authors: Nathan Killoran, Josh Izaac, Nicolás Quesada, Ville Bergholm, Matthew Amy, Christian Weedbrook

Abstract: … components: (i) an API for quantum programming based on an easy-to-use language named Blackbird; (ii) a suite of three virtual quantum computer backends, built in NumPy and Tensorflow, each targeting specialized uses; and (iii) an engine which can compile Blackbird programs on various backends, including the three built-in simulators, and -- in the near futu… ▽ More

Submitted 9 April, 2018; originally announced April 2018.

Comments: Try the Strawberry Fields Interactive website, located at http://strawberryfields.ai . Source code available at https://github.com/XanaduAI/strawberryfields

arXiv:1804.01138
[pdf] cs.DC

Designing a Micro-Benchmark Suite to Evaluate gRPC for TensorFlow: Early Experiences

Authors: Rajarshi Biswas, Xiaoyi Lu, Dhabaleswar K Panda

Abstract: … Google's gRPC is one of the most popular open source RPC frameworks available in the community. gRPC is the main communication engine for Google's Deep Learning framework TensorFlow. TensorFlow primarily uses gRPC for communicating tensors and administrative tasks among different processes. Tensor updates during the t… ▽ More

Submitted 3 April, 2018; originally announced April 2018.

Comments: 9 Pages, 14 Figures, This paper was presented at BPOE - 9 @ ASPLOS 2018

arXiv:1803.10228
[pdf, other] cs.LG

Demystifying Differentiable Programming: Shift/Reset the Penultimate Backpropagator

Authors: Fei Wang, Xilun Wu, Gregory Essertel, James Decker, Tiark Rompf

Abstract: … (staging), leading to a highly efficient implementation that combines the performance benefits of deep learning frameworks based on explicit reified computation graphs (e.g., TensorFlow) with the expressiveness of pure library approaches (e.g., PyTorch). ▽ More

Submitted 27 March, 2018; originally announced March 2018.

arXiv:1803.09926
[pdf, other] cs.CV

Diagonalwise Refactorization: An Efficient Training Method for Depthwise Convolutions

Authors: Zheng Qin, Zhaoning Zhang, Dongsheng Li, Yiming Zhang, Yuxing Peng

Abstract: … on TensorFlow, compared to their original implementations of depthwise convolutions. ▽ More

Submitted 27 March, 2018; originally announced March 2018.

Comments: 8 pages, 5 figures

arXiv:1803.09492
[pdf, other] cs.CV

Latency and Throughput Characterization of Convolutional Neural Networks for Mobile Computer Vision

Authors: Jussi Hanhirova, Teemu Kämäräinen, Sipi Seppälä, Matti Siekkinen, Vesa Hirvisalo, Antti Ylä-Jääski

Abstract: …(network-side server) computation. The measurements are conducted using real workloads and real processing platforms. On the platform side, we concentrate especially on TensorFlow and TensorRT. Our measurements include embedded processors found on mobile devices and high-performance processors that can be used on the network side of mobile systems. We show t… ▽ More

Submitted 26 March, 2018; originally announced March 2018.

Comments: 13 pages, 18 figures

arXiv:1803.09383
[pdf, ps, other] stat.ML

On the Performance of Preconditioned Stochastic Gradient Descent

Authors: XiLin Li

Abstract: …to equilibrated stochastic gradient descent (ESGD) and feature normalization, and provided a software package (https://github.com/lixilinx/psgd_tf) implemented in Tensorflow to compare PSGD with four different preconditioners and variations of stochastic gradient descent (SGD) on a wide range of benchmark problems with commonly used neural network models, e.… ▽ More

Submitted 1 April, 2018; v1 submitted 25 March, 2018; originally announced March 2018.

arXiv:1803.08165
[pdf, other] cs.NE

Comparing Fixed and Adaptive Computation Time for Recurrent Neural Networks

Authors: Daniel Fojo, Víctor Campos, Xavier Giro-i-Nieto

Abstract: …based on repeating each sample a fixed number of times. We found surprising results, where Repeat-RNN performs as good as ACT in the selected tasks. Source code in TensorFlow and PyTorch is publicly available at https://imatge-upc.github.io/danifojo-2018-repeatrnn/ ▽ More

Submitted 21 March, 2018; originally announced March 2018.

Comments: Accepted as workshop paper at ICLR 2018

arXiv:1803.07480
[pdf, other] cs.DB

AC/DC: In-Database Learning Thunderstruck

Authors: Mahmoud Abo Khamis, Hung Q Ngo, XuanLong Nguyen, Dan Olteanu, Maximilian Schleich

Abstract: …up to 86M tuples, AC/DC needs up to 30 minutes on one core of a commodity machine. This is up to three orders of magnitude faster than its competitors R, MadLib, libFM, and TensorFlow whenever they finish and thus do not exceed memory limitation, 24-hour timeout, or internal design limitations. ▽ More

Submitted 20 March, 2018; originally announced March 2018.

Comments: 10 pages, 3 figures

ACM Class: H.2.4; I.2.6

arXiv:1803.07015
[pdf] cs.CV

Live Target Detection with Deep Learning Neural Network and Unmanned Aerial Vehicle on Android Mobile Device

Authors: Ali Canberk Anar, Erkan Bostanci, Mehmet Serdar Guzel

Abstract: …stages faced during the development of an Android program which obtains and decodes live images from DJI Phantom 3 Professional Drone and implements certain features of the TensorFlow Android Camera Demo application. Test runs were made and outputs of the application were noted. A lake was classified as seashore, breakwater and pier with the proximities of 2… ▽ More

Submitted 22 March, 2018; v1 submitted 19 March, 2018; originally announced March 2018.

Comments: 6 pages

arXiv:1803.06905
[pdf, other] cs.LG

TBD: Benchmarking and Analyzing Deep Neural Network Training

Authors: Hongyu Zhu, Mohamed Akrout, Bojian Zheng, Andrew Pelegris, Amar Phanishayee, Bianca Schroeder, Gennady Pekhimenko

Abstract: … networks, reinforcement learning, and (ii) by performing an extensive performance analysis of training these different applications on three major deep learning frameworks (TensorFlow, MXNet, CNTK) across different hardware configurations (single-GPU, multi-GPU, and multi-machine). TBD currently covers six major application domains and eight different state-… ▽ More

Submitted 13 April, 2018; v1 submitted 16 March, 2018; originally announced March 2018.

arXiv:1803.04884
[pdf, other] cs.DB

IDEL: In-Database Entity Linking with Neural Embeddings

Authors: Torsten Kilias, Alexander Löser, Felix A Gers, Richard Koopmanschap, Ying Zhang, Martin Kersten

Abstract: … to support in-database-analytics with user defined functions (UDFs) implemented in Python. These functions call machine learning libraries for neural text mining, such as TensorFlow. The system achieves zero cost for data shipping and transformation by utilizing MonetDB's ability to embed Python processes in the database kernel and exchange data in NumPy arr… ▽ More

Submitted 13 March, 2018; originally announced March 2018.

Comments: This manuscript is a preprint for a paper submitted to VLDB2018

arXiv:1803.04559
[pdf, other] stat.ME

Weighted Bayesian Bootstrap for Scalable Bayes

Authors: Michael Newton, Nicholas G Polson, Jianeng Xu

Abstract: …quantification by sampling from a high dimensional posterior distribution. WBB is computationally fast and scalable using only off-theshelf optimization software such as TensorFlow. We provide regularity conditions which apply to a wide range of machine learning and statistical models. We illustrate our methodology in regularized regression, trend filtering… ▽ More

Submitted 12 March, 2018; originally announced March 2018.

arXiv:1803.03759
[pdf, other] stat.ML

Speech Recognition: Keyword Spotting Through Image Recognition

Authors: Sanjay Krishna Gouda, Salil Kanetkar, David Harrison, Manfred K Warmuth

Abstract: …whether a one second audio clip contains a particular word (out of a set of 10), an unknown word, or silence. The models to be implemented are a CNN recommended by the Tensorflow Speech Recognition tutorial, a low-latency CNN, and an adversarially trained CNN. The result is a demonstration of how to convert a problem in audio recognition to the better-studie… ▽ More

Submitted 10 March, 2018; originally announced March 2018.

arXiv:1803.03434
[pdf] cs.CV

Solving Fourier ptychographic imaging problems via neural network modeling and TensorFlow

Authors: Shaowei Jiang, Kaikai Guo, Jun Liao, Guoan Zheng

Abstract: … batch size of the network corresponds to the number of captured low-resolution images in one forward / backward pass. We use a popular open-source machine learning library, TensorFlow, for setting up the network and conducting the optimization process. We analyze the performance of different learning rates, different solvers, and different batch sizes. It is… ▽ More

Submitted 9 March, 2018; originally announced March 2018.

arXiv:1803.03288
[pdf, other] cs.DC

Communication Scheduling as a First-Class Citizen in Distributed Machine Learning Systems

Authors: Sayed Hadi Hashemi, Sangeetha Abdu Jyothi, Roy H Campbell

Abstract: …we develop a system for communication scheduling which realizes near-optimal overlap of communication and computation in graph-based models. Our system is implemented over TensorFlow and requires no changes in the model or developer inputs. Our system improves the throughput by up to 82% in inference and 20% in training, while also reducing straggler effect… ▽ More

Submitted 8 March, 2018; originally announced March 2018.

arXiv:1803.02875
[pdf, other] physics.optics

Machine Learning Inverse Problem for Topological Photonics

Authors: Laura Pilozzi, Francis A Farrelly, Giulia Marcucci, Claudio Conti

Abstract: … method to only select physically relevant solutions. We demonstrate our technique in a realistic topological laser design and by resorting to the widely available open-source TensorFlow library. Our results are general and scalable to thousands of topological components. This new inverse design technique based on machine learning potentially extends the appl… ▽ More

Submitted 7 March, 2018; originally announced March 2018.

arXiv:1802.09113
[pdf, other] cs.LG

GPU Accelerated Sub-Sampled Newton\textsf{'}s Method

Authors: Sudhir B Kylasa, Farbod Roosta-Khorasani, Michael W Mahoney, Ananth Grama

Abstract: …algorithms can be significantly accelerated, and can easily outperform state of the art implementations of existing techniques in popular ML/ DA software packages such as TensorFlow. Additionally these randomized methods incur a small memory overhead compared to first-order methods. In particular, we show that for million-dimensional problems, our GPU accele… ▽ More

Submitted 25 February, 2018; originally announced February 2018.

arXiv:1802.08960
[pdf, other] cs.RO

Bonnet: An Open-Source Training and Deployment Framework for Semantic Segmentation in Robotics using CNNs

Authors: Andres Milioto, Cyrill Stachniss

Abstract: …vision research and its use in robotics research. We provide an open-source framework for training and deployment. The training interface is implemented in Python using TensorFlow and the deployment interface provides a C++ library that can be easily integrated in an existing robotics codebase, a ROS node, and two standalone applications for label prediction… ▽ More

Submitted 25 February, 2018; originally announced February 2018.

Comments: Submitted to IEEE Robotics and Automation Letters (RA-L) 2018

arXiv:1802.08800
[pdf, other] cs.DB

Stochastic Gradient Descent on Highly-Parallel Architectures

Authors: Yujing Ma, Florin Rusu, Martin Torres

Abstract: There is an increased interest in building data analytics frameworks with advanced algebraic capabilities both in industry and academia. Many of these frameworks, e.g., TensorFlow and BIDMach, implement their compute-intensive primitives in two flavors---as multi-thread routines for multi-core CPUs and as highly-parallel kernels executed on GPU. Stochastic g… ▽ More

Submitted 24 February, 2018; originally announced February 2018.

arXiv:1802.06944
[pdf, other] cs.LG

DeepThin: A Self-Compressing Library for Deep Neural Networks

Authors: Matthew Sotoudeh, Sara S Baghsorkhi

Abstract: …by combining rank factorization with a reshaping process that adds nonlinearity to the approximation function. We deploy DeepThin as a plug-gable library integrated with TensorFlow that enables users to seamlessly compress models at different granularities. We evaluate DeepThin on two state-of-the-art acoustic models, TFKaldi and DeepSpeech, comparing it to… ▽ More

Submitted 19 February, 2018; originally announced February 2018.

arXiv:1802.05799
[pdf, other] cs.LG

Horovod: fast and easy distributed deep learning in TensorFlow

Authors: Alexander Sergeev, Mike Del Balso

Abstract: …Depending on the training library's API, the modification required may be either significant or minimal. Existing methods for enabling multi-GPU training under the TensorFlow library entail non-negligible communication overhead and require users to heavily modify their model-building code, leading many researchers to avoid the whole mess and stick with sl… ▽ More

Submitted 20 February, 2018; v1 submitted 15 February, 2018; originally announced February 2018.

arXiv:1802.05074
[pdf, other] cs.LG

L4: Practical loss-based stepsize adaptation for deep learning

Authors: Michal Rolinek, Georg Martius

Abstract: … increase in computational cost. The performance is validated on multiple architectures including ResNets and the Differential Neural Computer. A prototype implementation as a TensorFlow optimizer is released. ▽ More

Submitted 21 February, 2018; v1 submitted 14 February, 2018; originally announced February 2018.

arXiv:1802.04799
[pdf, other] cs.LG

TVM: End-to-End Optimization Stack for Deep Learning

Authors: Tianqi Chen, Thierry Moreau, Ziheng Jiang, Haichen Shen, Eddie Yan, Leyuan Wang, Yuwei Hu, Luis Ceze, Carlos Guestrin, Arvind Krishnamurthy

Abstract: Scalable frameworks, such as TensorFlow, MXNet, Caffe, and PyTorch drive the current popularity and utility of deep learning. However, these frameworks are optimized for a narrow range of server-class GPUs and deploying workloads to other platforms such as mobile phones, embedded devices, and specialized accelerators (e.g., FPGAs, ASICs) requires laborious m… ▽ More

Submitted 12 February, 2018; originally announced February 2018.

Comments: Longer version of SysML publication, arxiv version of UW techreport https://www.cs.washington.edu/tr/2017/12/UW-CSE-17-12-01.pdf

arXiv:1802.04730
[pdf, other] cs.PL

Tensor Comprehensions: Framework-Agnostic High-Performance Machine Learning Abstractions

Authors: Nicolas Vasilache, Oleksandr Zinenko, Theodoros Theodoridis, Priya Goyal, Zachary DeVito, William S Moses, Sven Verdoolaege, Andrew Adams, Albert Cohen

Abstract: …in automatic translation, speech-to-text, scene understanding, ranking user preferences, ad placement, etc. Competing frameworks for building these networks such as TensorFlow, Chainer, CNTK, Torch/PyTorch, Caffe1/2, MXNet and Theano, explore different tradeoffs between usability and expressiveness, research or production orientation and supported hardware.… ▽ More

Submitted 13 February, 2018; originally announced February 2018.

arXiv:1802.02736
[pdf, other] eess.SP

Completely Distributed Power Allocation using Deep Neural Network for Device to Device communication Underlaying LTE

Authors: Jeehyeong Kim, Joohan Park, Jaewon Noh, Sunghyun Cho

Abstract: …The deep learning can optimize total cell throughput while keeping constraints such as interference to eNB. The proposed scheme, which is implemented model using Tensorflow, can provide same throughput with the conventional method even it operates completely on distributed manner. ▽ More

Submitted 11 February, 2018; v1 submitted 8 February, 2018; originally announced February 2018.

Comments: 12 pages, 10 figures

arXiv:1802.02611
[pdf, other] cs.CV

Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation

Authors: LiangChieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam

Abstract: …a performance of 89% on the test set without any post-processing. Our paper is accompanied with a publicly available reference implementation of the proposed models in Tensorflow at https://github.com/tensorflow/models/tree/master/research/deeplab. ▽ More

Submitted 8 March, 2018; v1 submitted 7 February, 2018; originally announced February 2018.

Comments: Our Tensorflow-based implementation is publicly available at https://github.com/tensorflow/models/tree/master/research/deeplab

arXiv:1802.01115
[pdf, other] cs.CV

End2You -- The Imperial Toolkit for Multimodal Profiling by End-to-End Learning

Authors: Panagiotis Tzirakis, Stefanos Zafeiriou, Bjorn W Schuller

Abstract: … End2You -- the Imperial College London toolkit for multimodal profiling by end-to-end deep learning. End2You is an open-source toolkit implemented in Python and is based on Tensorflow. It provides capabilities to train and evaluate models in an end-to-end manner, i.e., using raw input. It supports input from raw audio, visual, physiological or other types of… ▽ More

Submitted 4 February, 2018; originally announced February 2018.

arXiv:1801.08058
[pdf, other] cs.DC

Intel nGraph: An Intermediate Representation, Compiler, and Executor for Deep Learning

Authors: Scott Cyphers, Arjun K Bansal, Anahita Bhiwandiwalla, Jayaram Bobba, Matthew Brookhart, Avijit Chakraborty, Will Constable, Christian Convey, Leona Cook, Omar Kanawi, Robert Kimball, Jason Knight, Nikolay Korovaiko, Varun Kumar, Yixing Lao, Christopher R Lishka, Jaikrishnan Menon, Jennifer Myers, Sandeep Aswath Narayana, Adam Procter, Tristan J Webb

Abstract: …open-sourced C++ library to simplify the realization of optimized deep learning performance across frameworks and hardware platforms. Initially-supported frameworks include TensorFlow, MXNet, and Intel neon framework. Initial backends are Intel Architecture CPUs (CPU), the Intel(R) Nervana Neural Network Processor(R) (NNP), and NVIDIA GPUs. Currently support…

▽ More

Submitted 29 January, 2018; v1 submitted 24 January, 2018; originally announced January 2018.

arXiv:1801.07972
[pdf] physics.chem-ph
doi
10.1273/cbij.18.58

Application of TensorFlow to recognition of visualized results of fragment molecular orbital (FMO) calculations

Authors: Sona Saitou, Jun Iijima, Mayu Fujimoto, Yuji Mochizuki, Koji Okuwaki, Hideo Doi, Yuto Komeiji

Abstract: We have applied Google's TensorFlow deep learning toolkit to recognize the visualized results of the fragment molecular orbital (FMO) calculations. Typical protein structures of alpha-helix and beta-sheet provide some characteristic patterns in the two-dimensional map of inter-fragment interaction energy termed as IFIE-map (Kurisaki et al., Biophys. Chem. 13… ▽ More

Submitted 24 January, 2018; originally announced January 2018.

Comments: 26 pages, 3 figures, 4 tables

Journal ref: Chem-Bio Informatics Journal, 18 (2018) 58-69

arXiv:1801.07232
[pdf, ps, other] physics.geo-ph

Seismic Full-Waveform Inversion Using Deep Learning Tools and Techniques

Authors: Alan Richardson

Abstract: … that the conventional seismic full-waveform inversion algorithm can be constructed as a recurrent neural network and so implemented using deep learning software such as TensorFlow. Applying another deep learning concept, the Adam optimizer with minibatches of data, produces quicker convergence toward the true wave speed model on a 2D dataset than Stochastic… ▽ More

Submitted 31 January, 2018; v1 submitted 22 January, 2018; originally announced January 2018.

Comments: 18 pages, 5 figures

arXiv:1801.04380
[pdf, other] cs.DC
doi
10.1145/3178487.3178491

SuperNeurons: Dynamic GPU Memory Management for Training Deep Neural Networks

Authors: Linnan Wang, Jinmian Ye, Yiyang Zhao, Wei Wu, Ang Li, Shuaiwen Leon Song, Zenglin Xu, Tim Kraska

Abstract: … memory for the training, but also dynamically allocates the memory for convolution workspaces to achieve the high performance. Evaluations against Caffe, Torch, MXNet and TensorFlow have demonstrated that SuperNeurons trains at least 3.2432 deeper network than current ones with the leading performance. Particularly, SuperNeurons can train ResNet2500 that has… ▽ More

Submitted 12 January, 2018; originally announced January 2018.

Comments: PPoPP '2018: 23nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming

arXiv:1801.03138
[pdf, other] cs.AI

Deep In-GPU Experience Replay

Authors: Ben Parr

Abstract: Experience replay allows a reinforcement learning agent to train on samples from a large amount of the most recent experiences. A simple in-RAM experience replay stores these most recent experiences in a list in RAM, and then copies sampled batches to the GPU for training. I moved this list to the GPU, thus creating an in-GPU experience replay, and a training step that no longer has inputs copied… ▽ More

Submitted 9 January, 2018; originally announced January 2018.

Comments: Source code (uses TensorFlow): https://github.com/bparr/gpu-experience-replay

arXiv:1801.03137
[pdf, other] cs.LG

Convergence Analysis of Gradient Descent Algorithms with Proportional Updates

Authors: Igor Gitman, Deepak Dilipkumar, Ben Parr

Abstract: The rise of deep learning in recent years has brought with it increasingly clever optimization methods to deal with complex, non-linear loss functions. These methods are often designed with convex optimization in mind, but have been shown to work well in practice even for the highly non-convex optimization associated with neural networks. However, one significant drawback of these methods when the… ▽ More

Submitted 9 January, 2018; originally announced January 2018.

Comments: Source code (uses TensorFlow): https://github.com/bparr/lars

arXiv:1801.01928
[pdf, ps, other] cs.MS

Tensor Train decomposition on TensorFlow (T3F)

Authors: Alexander Novikov, Pavel Izmailov, Valentin Khrulkov, Michael Figurnov, Ivan Oseledets

Abstract: Tensor Train decomposition is used across many branches of machine learning, but until now it lacked an implementation with GPU support, batch processing, automatic differentiation, and versatile functionality for Riemannian optimization framework, which takes in account the underlying manifold structure in order to construct efficient optimization methods. In this work, we propose a library that… ▽ More

Submitted 5 January, 2018; originally announced January 2018.

arXiv:1712.09401
[pdf, other] cs.CV

Robust Minutiae Extractor: Integrating Deep Networks and Fingerprint Domain Knowledge

Authors: DinhLuan Nguyen, Kai Cao, Anil K Jain

Abstract: …detection will be useful to train network-based fingerprint matching algorithms as well as for evaluating fingerprint individuality at scale. MinutiaeNet is implemented in Tensorflow: https://github.com/luannd/MinutiaeNet ▽ More

Submitted 26 December, 2017; originally announced December 2017.

Comments: Accepted to International Conference on Biometrics (ICB 2018)

arXiv:1712.09388
[pdf, other] cs.DC

Scaling GRPC Tensorflow on 512 nodes of Cori Supercomputer

Authors: Amrita Mathuriya, Thorsten Kurth, Vivek Rane, Mustafa Mustafa, Lei Shao, Debbie Bard, Prabhat, Victor W Lee

Abstract: We explore scaling of the standard distributed Tensorflow with GRPC primitives on up to 512 Intel Xeon Phi (KNL) nodes of Cori supercomputer with synchronous stochastic gradient descent (SGD), and identify causes of scaling inefficiency at higher node counts. To our knowledge, this is the first exploration of distributed GRPC … ▽ More

Submitted 26 December, 2017; originally announced December 2017.

Comments: Published as a poster in NIPS 2017 Workshop: Deep Learning At Supercomputer Scale

arXiv:1712.07805
[pdf, other] cs.CR

Wolf in Sheep's Clothing - The Downscaling Attack Against Deep Learning Applications

Authors: Qixue Xiao, Kang Li, Deyue Zhang, Yier Jin

Abstract: …usually assume a fixed scale for their training and input data. To allow deep learning applications to handle a wide range of input data, popular frameworks, such as Caffe, TensorFlow, and Torch, all provide data scaling functions to resize input to the dimensions used by deep learning models. Image scaling algorithms are intended to preserve the visual feat… ▽ More

Submitted 21 December, 2017; originally announced December 2017.

arXiv:1712.06272
[pdf, other] cs.AR

Automated flow for compressing convolution neural networks for efficient edge-computation with FPGA

Authors: Farhan Shafiq, Takato Yamada, Antonio T Vilchez, Sakyasingha Dasgupta

Abstract: … Moreover, the transition to lower bit-widths opens new avenues for performance optimizations and model improvement. In this paper, we present an automatic flow from trained TensorFlow models to FPGA system on chip implementation of binarized CNN. This flow involves quantization of model parameters and activations, generation of network and model in embedded… ▽ More

Submitted 18 December, 2017; originally announced December 2017.

Comments: 7 pages, 9 figures. Accepted and presented at MLPCD workshop, NIPS 2017 (LongBeach, California)

arXiv:1712.06139
[pdf, other] cs.DC

TensorFlow-Serving: Flexible, High-Performance ML Serving

Authors: Christopher Olston, Noah Fiedel, Kiril Gorovoy, Jeremiah Harmsen, Li Lao, Fangwei Li, Vinu Rajashekhar, Sukriti Ramesh, Jordan Soyke

Abstract: We describe TensorFlow-Serving, a system to serve machine learning models inside Google which is also available in the cloud and via open-source. It is extremely flexible in terms of the types of ML platforms it supports, and ways to integrate with systems that convey new models and updated versions from training to serving. At the same time, the core code p… ▽ More

Submitted 27 December, 2017; v1 submitted 17 December, 2017; originally announced December 2017.

Comments: Presented at NIPS 2017 Workshop on ML Systems (http://learningsys.org/nips17/acceptedpapers.html)

arXiv:1712.05902
[pdf, other] cs.LG

NSML: A Machine Learning Platform That Enables You to Focus on Your Models

Authors: Nako Sung, Minkyu Kim, Hyunwoo Jo, Youngil Yang, Jingwoong Kim, Leonard Lausen, Youngkwan Kim, Gayoung Lee, Donghyun Kwak, JungWoo Ha, Sunghun Kim

Abstract: Machine learning libraries such as TensorFlow and PyTorch simplify model implementation. However, researchers are still required to perform a non-trivial amount of manual tasks such as GPU allocation, training status tracking, and comparison of models with different hyperparameter settings. We propose a system to handle these tasks and help researchers focus… ▽ More

Submitted 15 December, 2017; originally announced December 2017.

Comments: 8 pages, 4figures

arXiv:1712.04048
[pdf, other] cs.LG

Cavs: A Vertex-centric Programming Interface for Dynamic Neural Networks

Authors: Hao Zhang, Shizhen Xu, Graham Neubig, Wei Dai, Qirong Ho, Guangwen Yang, Eric P Xing

Abstract: …, and naturally exposes batched execution opportunities over different graphs. Experiments comparing Cavs to two state-of-the-art frameworks for dynamic NNs (TensorFlow Fold and DyNet) demonstrate the efficacy of this approach: Cavs achieves a near one order of magnitude speedup on training of various dynamic NN architectures, and ablations demonstrate the c… ▽ More

Submitted 11 December, 2017; originally announced December 2017.

Comments: Short versions of this paper were presented at AISys workshop@SOSP 2017 and MLSys workshop@NIPS 2017

arXiv:1712.03641
[pdf, other] physics.comp-ph

DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics

Authors: Han Wang, Linfeng Zhang, Jiequn Han, Weinan E

Abstract: …Potential applications of DeePMD-kit span from finite molecules to extended systems and from metallic systems to chemically bonded systems. DeePMD-kit is interfaced with TensorFlow, one of the most popular deep learning frameworks, making the training process highly automatic and efficient. On the other end, DeePMD-kit is interfaced with high-performance cl… ▽ More

Submitted 30 December, 2017; v1 submitted 10 December, 2017; originally announced December 2017.

arXiv:1712.03376
[pdf, other] cs.CL

Word Sense Disambiguation with LSTM: Do We Really Need 100 Billion Words?

Authors: Minh Le, Marten Postma, Jacopo Urbani

Abstract: …code was released. This paper presents the results of a reproduction study of this technique using only openly available datasets (GigaWord, SemCore, OMSTI) and software (TensorFlow). From them, it emerged that state-of-the-art results can be obtained with much less data than hinted by Yuan et al. All code and trained models are made freely available. ▽ More

Submitted 16 December, 2017; v1 submitted 9 December, 2017; originally announced December 2017.

arXiv:1712.03073
[pdf, other] cs.CY

Enabling Cooperative Inference of Deep Learning on Wearables and Smartphones

Authors: Mengwei Xu, Feng Qian, Saumay Pushp

Abstract: …requirements, and user preference. Deployed as a user-space library, CoINF offers developer-friendly APIs that are as simple as those in traditional DL libraries such as TensorFlow, with all complicated offloading details hidden. We have implemented a prototype of CoINF on Android OS, and used real deep learning models to evaluate its performance on commerci… ▽ More

Submitted 1 December, 2017; originally announced December 2017.

arXiv:1712.00559
[pdf, other] cs.CV

Progressive Neural Architecture Search

Authors: Chenxi Liu, Barret Zoph, Maxim Neumann, Jonathon Shlens, Wei Hua, LiJia Li, Li Fei-Fei, Alan Yuille, Jonathan Huang, Kevin Murphy

Abstract: We propose a new method for learning the structure of convolutional neural networks (CNNs) that is more efficient than recent state-of-the-art methods based on reinforcement learning and evolutionary algorithms. Our approach uses a sequential model-based optimization (SMBO) strategy, in which we search for structures in order of increasing complexity, while simultaneously learning a surrogate mode… ▽ More

Submitted 23 March, 2018; v1 submitted 2 December, 2017; originally announced December 2017.

Comments: The code and checkpoint for the PNAS model trained on ImageNet can now be downloaded from https://github.com/tensorflow/models/tree/master/research/slim#Pretrained

arXiv:1711.11008
[pdf, other] cs.CR

Security Risks in Deep Learning Implementations

Authors: Qixue Xiao, Kang Li, Deyue Zhang, Weilin Xu

Abstract: …algorithms overshadows their security risk in software implementations. This paper discloses a set of vulnerabilities in popular deep learning frameworks including Caffe, TensorFlow, and Torch. Contrast to the small code size of deep learning models, these deep learning frameworks are complex and contain heavy dependencies on numerous open source packages. T… ▽ More

Submitted 29 November, 2017; originally announced November 2017.

arXiv:1711.10604
[pdf, ps, other] cs.LG

TensorFlow Distributions

Authors: Joshua V Dillon, Ian Langmore, Dustin Tran, Eugene Brevdo, Srinivas Vasudevan, Dave Moore, Brian Patton, Alex Alemi, Matt Hoffman, Rif A Saurous

Abstract: The TensorFlow Distributions library implements a vision of probability theory adapted to the modern deep-learning paradigm of end-to-end differentiable computation. Building on two basic abstractions, it offers flexible building blocks for probabilistic computation. Distributions provide fast, numerically stable methods for generating samples and computing… ▽ More

Submitted 28 November, 2017; originally announced November 2017.

arXiv:1711.09545
[pdf, other] stat.CO

OSTSC: Over Sampling for Time Series Classification in R

Authors: Matthew Dixon, Diego Klabjan, Lan Wei

Abstract: …the functionality in the package. To demonstrate the performance impact of OSTSC, we then provide two medium size imbalanced time series datasets. Each example applies a TensorFlow implementation of a Long Short-Term Memory (LSTM) classifier - a type of a Recurrent Neural Network (RNN) classifier - to imbalanced time series. The classifier performance is com… ▽ More

Submitted 27 November, 2017; originally announced November 2017.

arXiv:1711.09268
[pdf, other] stat.ML

Generalizing Hamiltonian Monte Carlo with Neural Networks

Authors: Daniel Levy, Matthew D Hoffman, Jascha Sohl-Dickstein

Abstract: … no measurable progress in a second. Finally, we show quantitative and qualitative gains on a real-world task: latent-variable generative modeling. We release an open source TensorFlow implementation of the algorithm. ▽ More

Submitted 12 January, 2018; v1 submitted 25 November, 2017; originally announced November 2017.

arXiv:1711.09181
[pdf, ps, other] cs.CL

Towards Accurate Deceptive Opinion Spam Detection based on Word Order-preserving CNN

Authors: Siyuan Zhao, Zhiwei Xu, Limin Liu, Mengjie Guo

Abstract: … in its convolution layer and pooling layer, which makes convolution neural network more suitable for various text classification and deceptive opinions detection. The TensorFlow-based experiments demonstrate that the detection mechanism proposed in this paper achieve more accurate deceptive opinion detection results. ▽ More

Submitted 19 March, 2018; v1 submitted 24 November, 2017; originally announced November 2017.

arXiv:1711.06853
[pdf, other] cs.CV

DLTK: State of the Art Reference Implementations for Deep Learning on Medical Images

Authors: Nick Pawlowski, Sofia Ira Ktena, Matthew C H Lee, Bernhard Kainz, Daniel Rueckert, Ben Glocker, Martin Rajchl

Abstract: We present DLTK, a toolkit providing baseline implementations for efficient experimentation with deep learning methods on biomedical images. It builds on top of TensorFlow and its high modularity and easy-to-use examples allow for a low-threshold access to state-of-the-art implementations for typical medical imaging problems. A comparison of DLTK's reference… ▽ More

Submitted 18 November, 2017; originally announced November 2017.

Comments: Submitted to Medical Imaging Meets NIPS 2017, Code at https://github.com/DLTK/DLTK

arXiv:1711.05979
[pdf, other] cs.DC

Performance Modeling and Evaluation of Distributed Deep Learning Frameworks on GPUs

Authors: Shaohuai Shi, Xiaowen Chu

Abstract: …on the same GPU hardware. In this paper, we evaluate the running performance of four state-of-the-art distributed deep learning frameworks (i.e., Caffe-MPI, CNTK, MXNet and TensorFlow) over single-GPU, multi-GPU and multi-node environments. We first build performance models of standard processes in training DNNs with SGD, and then we benchmark the running pe… ▽ More

Submitted 8 December, 2017; v1 submitted 16 November, 2017; originally announced November 2017.

Comments: 11 pages

arXiv:1711.03845
[pdf, other] stat.ML

GPflowOpt: A Bayesian Optimization Library using TensorFlow

Authors: Nicolas Knudde, Joachim van der Herten, Tom Dhaene, Ivo Couckuyt

Abstract: …framework for Bayesian optimization known as GPflowOpt is introduced. The package is based on the popular GPflow library for Gaussian processes, leveraging the benefits of TensorFlow including automatic differentiation, parallelization and GPU computations for Bayesian optimization. Design goals focus on a framework that is easy to extend with custom acquisi… ▽ More

Submitted 10 November, 2017; originally announced November 2017.

arXiv:1711.03278
[pdf, other] cs.CV

Feed Forward and Backward Run in Deep Convolution Neural Network

Authors: Pushparaja Murugan

Abstract: …by krizhevsky, the architecture of deep Convolution Neural Network is attracted many researchers. This has led to the major development in Deep learning frameworks such as Tensorflow, caffe, keras, theno. Though the implementation of deep learning is quite possible by employing deep learning frameworks, mathematical theory and concepts are harder to understa… ▽ More

Submitted 9 November, 2017; originally announced November 2017.

Comments: 20 pages, 20th International Conference on Computer Vision and Image Processing

arXiv:1711.02712
[pdf, other] cs.MS

Tangent: Automatic Differentiation Using Source Code Transformation in Python

Authors: Bart van Merriënboer, Alexander B Wiltschko, Dan Moldovan

Abstract: …new Python functions which calculate a derivative. This approach to automatic differentiation is different from existing packages popular in machine learning, such as TensorFlow and Autograd. Advantages are that Tangent generates gradient code in Python which is readable by the user, easy to understand and debug, and has no runtime overhead. Tangent also int… ▽ More

Submitted 7 November, 2017; originally announced November 2017.

arXiv:1711.01912
[pdf, other] cs.DC
doi
10.1145/3154842.3154843

The TensorFlow Partitioning and Scheduling Problem: It's the Critical Path!

Authors: Ruben Mayer, Christian Mayer, Larissa Laich

Abstract: State-of-the-art data flow systems such as TensorFlow impose iterative calculations on large graphs that need to be partitioned on heterogeneous devices such as CPUs, GPUs, and TPUs. However, partitioning can not be viewed in isolation. Each device has to select the next graph vertex to be executed, i.e., perform local scheduling decisions. Both problems, pa… ▽ More

Submitted 6 November, 2017; originally announced November 2017.

Comments: 6 pages. To be published in Proceedings of DIDL '17: Workshop on Distributed Infrastructures for Deep Learning, hosted by ACM Middleware 2017 Conference. https://doi.org/10.1145/3154842.3154843

arXiv:1710.11555
[pdf, other] stat.ML

TF Boosted Trees: A scalable TensorFlow based framework for gradient boosting

Authors: Natalia Ponomareva, Soroush Radpour, Gilbert Hendry, Salem Haykal, Thomas Colthurst, Petr Mitrichev, Alexander Grushetsky

Abstract: TF Boosted Trees (TFBT) is a new open-sourced frame-work for the distributed training of gradient boosted trees. It is based on TensorFlow, and its distinguishing features include a novel architecture, automatic loss differentiation, layer-by-layer boosting that results in smaller ensembles and faster prediction, principled multi-class handling, and a number… ▽ More

Submitted 31 October, 2017; originally announced October 2017.

Comments: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2017). The final publication will be available at link.springer.com and is available on ECML website http://ecmlpkdd2017.ijs.si/papers/paperID705.pdf

arXiv:1710.11547
[pdf, other] stat.ML

Compact Multi-Class Boosted Trees

Authors: Natalia Ponomareva, Thomas Colthurst, Gilbert Hendry, Salem Haykal, Soroush Radpour

Abstract: …smaller steps in function space, which is empirically shown to lead to a faster convergence and to a more compact ensemble. We have added both improvements to the open-source TensorFlow Boosted trees (TFBT) package, and we demonstrate their efficacy on a variety of multiclass datasets. We expect these extensions will be of particular interest to boosted tree… ▽ More

Submitted 31 October, 2017; originally announced October 2017.

Comments: Accepted for publication in IEEE Big Data 2017 http://cci.drexel.edu/bigdata/bigdata2017/AcceptedPapers.html

arXiv:1710.10229
[pdf] q-bio.PE

Feature learning of virus genome evolution with the nucleotide skip-gram neural network

Authors: Hyunjin Shim

Abstract: …frequency of echovirus 11 in the presence or absence of the disinfectant from the experimental evolution data. Results from the training using a new open-source software TensorFlow show that the learned distributed vectors can be clustered using Principal Component Analysis and Hierarchical Clustering to reveal a list of non-synonymous mutations that arise o… ▽ More

Submitted 27 October, 2017; originally announced October 2017.

Comments: 16 pages, 4 figures

arXiv:1710.09967
[pdf, ps, other] cs.LG

Improving Deep Learning by Inverse Square Root Linear Units (ISRLUs)

Authors: Brad Carlile, Guy Delamarter, Paul Kinney, Akiko Marti, Brian Whitney

Abstract: …The significant performance advantage of ISRLU on traditional CPUs also carry over to more efficient HW implementations on HW/SW codesign for CNNs/RNNs. In experiments with TensorFlow, ISRLU leads to faster learning and better generalization than ReLU on CNNs. This work also suggests a computationally efficient variant called the "inverse square root unit"… ▽ More

Submitted 9 November, 2017; v1 submitted 26 October, 2017; originally announced October 2017.

Comments: 8 pages, 2 figures, 5 tables

arXiv:1710.08961
[pdf] cs.DC

Fast and Scalable Distributed Deep Convolutional Autoencoder for fMRI Big Data Analytics

Authors: Milad Makkie, Heng Huang, Yu Zhao, Athanasios V Vasilakos, Tianming Liu

Abstract: …the processing power of multiple GPUs in a distributed fashion. To implement such a model, we have created an enhanced processing pipeline on the top of Apache Spark and Tensorflow library, leveraging from a very large cluster of GPU machines. Experimental data from applying the model on the Human Connectome Project (HCP) show that the proposed model is effi… ▽ More

Submitted 13 December, 2017; v1 submitted 24 October, 2017; originally announced October 2017.

arXiv:1710.08717
[pdf, ps, other] cs.MS

Auto-Differentiating Linear Algebra

Authors: Matthias Seeger, Asmus Hetzel, Zhenwen Dai, Neil D Lawrence

Abstract: Development systems for deep learning, such as Theano, Torch, TensorFlow, or MXNet, are easy-to-use tools for creating complex neural network models. Since gradient computations are automatically baked in, and execution is mapped to high performance hardware, these models can be trained end-to-end on large amounts of data. However, it is currently not easy t… ▽ More

Submitted 30 October, 2017; v1 submitted 24 October, 2017; originally announced October 2017.

arXiv:1710.06554
[pdf, other] cs.CL

Honk: A PyTorch Reimplementation of Convolutional Neural Networks for Keyword Spotting

Authors: Raphael Tang, Jimmy Lin

Abstract: We describe Honk, an open-source PyTorch reimplementation of convolutional neural networks for keyword spotting that are included as examples in TensorFlow. These models are useful for recognizing "command triggers" in speech-based interfaces (e.g., "Hey Siri"), which serve as explicit cues for audio recordings of utterances that are sent to the cloud for fu… ▽ More

Submitted 28 November, 2017; v1 submitted 17 October, 2017; originally announced October 2017.

Comments: 3 pages, 2 figures

arXiv:1710.06390

[pdf, other] cs.LG

Fishing for Clickbaits in Social Images and Texts with Linguistically-Infused Neural Network Models

Authors: Maria Glenski, Ellyn Ayton, Dustin Arendt, Svitlana Volkova

Abstract: … combined with text yield significant performance improvement yet. Nevertheless, this work is the first to present preliminary analysis of objects extracted using Google Tensorflow object detection API from images in clickbait vs. non-clickbait Twitter posts. Finally, we outline several steps to improve model performance as a part of the future work. ▽ More

Submitted 17 October, 2017; originally announced October 2017.

Comments: Pineapplefish Clickbait Detector, Clickbait Challenge 2017

arXiv:1710.05758
[pdf, other] cs.CV

TensorQuant - A Simulation Toolbox for Deep Neural Network Quantization

Authors: Dominik Marek Loroch, Norbert Wehn, FranzJosef Pfreundt, Janis Keuper

Abstract: …theory available, which would allow users to derive the optimal quantization during the design of a DNN topology. In this paper, we present a quantization tool box for the TensorFlow framework. TensorQuant allows a transparent quantization simulation of existing DNN topologies during training and inference. TensorQuant supports generic quantization methods a… ▽ More

Submitted 13 October, 2017; originally announced October 2017.

arXiv:1710.05267
[pdf] cs.CV

Deep Learning for Rapid Sparse MR Fingerprinting Reconstruction

Authors: Ouri Cohen, Bo Zhu, Matthew S Rosen

Abstract: …a novel fast method for reconstruction of multi-dimensional MR Fingerprinting (MRF) data using Deep Learning methods. METHODS: A neural network (NN) is defined using the TensorFlow framework and trained on simulated MRF data computed using the Bloch equations. The accuracy of the NN reconstruction of noisy data is compared to conventional MRF template matc… ▽ More

Submitted 14 October, 2017; originally announced October 2017.

Comments: 21 pages, 7 figures

arXiv:1710.00578
[pdf, other] stat.CO

sgmcmc: An R Package for Stochastic Gradient Markov Chain Monte Carlo

Authors: Jack Baker, Paul Fearnhead, Emily B Fox, Christopher Nemeth

Abstract: …calculates these gradients itself using automatic differentiation, making the implementation of these methods much easier. To do this, the package uses the software library TensorFlow, which has a variety of statistical distributions and mathematical operations as standard, meaning a wide class of models can be built using this framework. SGMCMC has become w… ▽ More

Submitted 13 April, 2018; v1 submitted 2 October, 2017; originally announced October 2017.

arXiv:1709.09161
[pdf, other] stat.ML

EDEN: Evolutionary Deep Networks for Efficient Machine Learning

Authors: Emmanuel Dufourq, Bruce A Bassett

Abstract: …we propose Evolutionary DEep Networks (EDEN), a computationally efficient neuro-evolutionary algorithm which interfaces to any deep neural network platform, such as TensorFlow. We show that EDEN evolves simple yet successful architectures built from embedding, 1D and 2D convolutional, max pooling and fully connected layers along with their hyperparameters.… ▽ More

Submitted 26 September, 2017; originally announced September 2017.

Comments: 7 pages, 3 figures, 3 tables and see video https://vimeo.com/234510097

arXiv:1709.06680
[pdf, other] stat.ML

Deep Lattice Networks and Partial Monotonic Functions

Authors: Seungil You, David Ding, Kevin Canini, Jan Pfeifer, Maya Gupta

Abstract: …with appropriate constraints for monotonicity, and jointly training the resulting network. We implement the layers and projections with new computational graph nodes in TensorFlow and use the ADAM optimizer and batched stochastic gradients. Experiments on benchmark and real-world datasets show that six-layer monotonic deep lattice networks achieve state-of-… ▽ More

Submitted 19 September, 2017; originally announced September 2017.

Comments: 9 pages, NIPS 2017

arXiv:1709.06416
[pdf, other] cs.DC

Weld: Rethinking the Interface Between Data-Intensive Applications

Authors: Shoumik Palkar, James Thomas, Deepak Narayanan, Anil Shanbhag, Rahul Palamuttam, Holger Pirk, Malte Schwarzkopf, Saman Amarasinghe, Samuel Madden, Matei Zaharia

Abstract: …It then optimizes data movement across these functions and emits efficient code for diverse hardware. Weld can be integrated into existing frameworks such as Spark, TensorFlow, Pandas and NumPy without changing their user-facing APIs. We demonstrate that Weld can speed up applications using these frameworks by up to 29x. ▽ More

Submitted 24 October, 2017; v1 submitted 14 September, 2017; originally announced September 2017.

arXiv:1709.05871
[pdf] cs.DC

IBM Deep Learning Service

Authors: Bishwaranjan Bhattacharjee, Scott Boag, Chandani Doshi, Parijat Dube, Ben Herta, Vatche Ishakian, K R Jayaram, Rania Khalaf, Avesh Krishna, Yu Bo Li, Vinod Muthusamy, Ruchir Puri, Yufei Ren, Florian Rosenberg, Seetharami R Seelam, Yandong Wang, Jian Ming Zhang, Li Zhang

Abstract: …architecture behind IBM's deep learning as a service (DLaaS). DLaaS provides developers the flexibility to use popular deep learning libraries such as Caffe, Torch and TensorFlow, in the cloud in a scalable and resilient manner with minimal effort. The platform uses a distribution and orchestration layer that facilitates learning from a large amount of data… ▽ More

Submitted 18 September, 2017; originally announced September 2017.

arXiv:1709.05870
[pdf, other] stat.ML

ZhuSuan: A Library for Bayesian Deep Learning

Authors: Jiaxin Shi, Jianfei Chen, Jun Zhu, Shengyang Sun, Yucen Luo, Yihong Gu, Yuhao Zhou

Abstract: …python probabilistic programming library for Bayesian deep learning, which conjoins the complimentary advantages of Bayesian methods and deep learning. ZhuSuan is built upon Tensorflow. Unlike existing deep learning libraries, which are mainly designed for deterministic neural networks and supervised tasks, ZhuSuan is featured for its deep root into Bayesian… ▽ More

Submitted 18 September, 2017; originally announced September 2017.

Comments: The GitHub page is at https://github.com/thu-ml/zhusuan

arXiv:1709.03485
[pdf, other] cs.CV
doi
10.1016/j.cmpb.2018.01.025

NiftyNet: a deep-learning platform for medical imaging

Authors: Eli Gibson, Wenqi Li, Carole Sudre, Lucas Fidon, Dzhoshkun I Shakir, Guotai Wang, Zach Eaton-Rosen, Robert Gray, Tom Doel, Yipeng Hu, Tom Whyntie, Parashkev Nachev, Marc Modat, Dean C Barratt, Sébastien Ourselin, M Jorge Cardoso, Tom Vercauteren

Abstract: …functions and evaluation metrics are tailored to, and take advantage of, the idiosyncracies of medical image analysis and computer-assisted intervention. NiftyNet is built on TensorFlow and supports TensorBoard visualization of 2D and 3D images and computational graphs by default. We present 3 illustrative medical image analysis applications built using Ni… ▽ More

Submitted 16 October, 2017; v1 submitted 11 September, 2017; originally announced September 2017.

Comments: Wenqi Li and Eli Gibson contributed equally to this work. M. Jorge Cardoso and Tom Vercauteren contributed equally to this work. 26 pages, 6 figures; Update includes additional applications, updated author list and formatting for journal submission

arXiv:1709.02878

[pdf, other] cs.LG

TensorFlow Agents: Efficient Batched Reinforcement Learning in TensorFlow

Authors: Danijar Hafner, James Davidson, Vincent Vanhoucke

Abstract: We introduce TensorFlow Agents, an efficient infrastructure paradigm for building parallel reinforcement learning algorithms in TensorFlow. We simulate multiple environments in parallel, and group them to perform the neural network computation on a batch rather than individual observations. This allows the … ▽ More

Submitted 8 September, 2017; originally announced September 2017.

Comments: White paper, 7 pages

arXiv:1708.08670
[pdf] cs.LG
doi
10.1007/978-3-319-70581-1_17

Performance Analysis of Open Source Machine Learning Frameworks for Various Parameters in Single-Threaded and Multi-Threaded Modes

Authors: Yuriy Kochura, Sergii Stirenko, Oleg Alienin, Michail Novotarskiy, Yuri Gordienko

Abstract: The basic features of some of the most versatile and popular open source frameworks for machine learning (TensorFlow, Deep Learning4j, and H2O) are considered and compared. Their comparative analysis was performed and conclusions were made as to the advantages and disadvantages of these platforms. The performance tests for the de facto standard MNIST data se… ▽ More

Submitted 29 August, 2017; originally announced August 2017.

Comments: 15 pages, 11 figures, 4 tables; this paper summarizes the activities which were started recently and described shortly in the previous conference presentations arXiv:1706.02248 and arXiv:1707.04940; it is accepted for Springer book series "Advances in Intelligent Systems and Computing"

Journal ref: Advances in Intelligent Systems and Computing II. CSIT 2017. Advances in Intelligent Systems and Computing, vol 689, pp 243-256. Springer, Cham

arXiv:1708.07829
[pdf, other] cs.DS

Algorithms for Big Data: Graphs and PageRank

Authors: Sergio García Prado

Abstract: …case study. Finally, the development of a library for the resolution of graph problems, implemented on the top of the intensive mathematical computation platform known as TensorFlow has been started. ▽ More

Submitted 25 August, 2017; originally announced August 2017.

Comments: in Spanish, 143 pages, final degree project (bachelor's thesis)

arXiv:1708.04915
[pdf, other] cs.SE
doi
10.1109/ICSE-NIER.2017.13

DARVIZ: Deep Abstract Representation, Visualization, and Verification of Deep Learning Models

Authors: Anush Sankaran, Rahul Aralikatte, Senthil Mani, Shreya Khare, Naveen Panwar, Neelamadhav Gantayat

Abstract: …current large scale data-driven software development is challenging. Further, for deep learning development there are many libraries in multiple programming languages such as TensorFlow (Python), CAFFE (C++), Theano (Python), Torch (Lua), and Deeplearning4j (Java), driving a huge need for interoperability across libraries. ▽ More

Submitted 16 August, 2017; originally announced August 2017.

Comments: Accepted in ICSE NIER 2017. Preprint

arXiv:1708.04103
[pdf, ps, other] q-bio.BM

SECLAF: A Webserver and Deep Neural Network Design Tool for Biological Sequence Classification

Authors: Balazs Szalkai, Vince Grolmusz

Abstract: …in numerous areas, including biological sequence analysis. Here we introduce SECLAF, an artificial neural-net based biological sequence classifier framework, which uses the Tensorflow library of Google, Inc. By applying SECLAF for residue-sequences, we have reported (Methods (2017), https://doi.org/10.1016/j.ymeth.2017.06.034) the most accurate multi-label p… ▽ More

Submitted 14 August, 2017; originally announced August 2017.

arXiv:1708.03788
[pdf, other] cs.LG

Direct-Manipulation Visualization of Deep Networks

Authors: Daniel Smilkov, Shan Carter, D Sculley, Fernanda B Viégas, Martin Wattenberg

Abstract: … While the theory is important, it is also helpful for novices to develop an intuitive feel for the effect of different hyperparameters and structural variations. We describe TensorFlow Playground, an interactive, open sourced visualization that allows users to experiment via direct manipulation rather than coding, enabling them to quickly build an intuition… ▽ More

Submitted 12 August, 2017; originally announced August 2017.

arXiv:1708.03665
[pdf] stat.ML

Time Series Anomaly Detection; Detection of anomalous drops with limited features and sparse examples in noisy highly periodic data

Authors: Dominique T Shipmon, Jason M Gurevitch, Paolo M Piselli, Stephen T Edwards

Abstract: … Since we do not have a large body of labeled examples to directly apply supervised learning for anomaly classification, we approached the problem in two parts. First we used TensorFlow to train our various models including DNNs, RNNs, and LSTMs to perform regression and predict the expected value in the time series. Secondly we created anomaly detection rul… ▽ More

Submitted 11 August, 2017; originally announced August 2017.

arXiv:1708.03157
[pdf, other] cs.DC

TensorFlow Enabled Genetic Programming

Authors: Kai Staats, Edward Pantridge, Marco Cavaglia, Iurii Milovanov, Arun Aniyan

Abstract: Genetic Programming, a kind of evolutionary computation and machine learning algorithm, is shown to benefit significantly from the application of vectorized data and the

TensorFlow numerical computation library on both CPU and GPU architectures. The open source, Python Karoo GP is employed for a series of 190 tests across 6 platforms, with real-world dataset… ▽ More

Submitted 10 August, 2017; originally announced August 2017.

Comments: 8 pages, 5 figures; presented at GECCO 2017, Berlin, Germany

Journal ref: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO) Companion, ACM 2017, pp. 1872-1879

arXiv:1708.02637
[pdf, other] cs.DC
doi
10.1145/3097983.3098171

TensorFlow Estimators: Managing Simplicity vs. Flexibility in High-Level Machine Learning Frameworks

Authors: HengTze Cheng, Zakaria Haque, Lichan Hong, Mustafa Ispir, Clemens Mewald, Illia Polosukhin, Georgios Roumpos, D Sculley, Jamie Smith, David Soergel, Yuan Tang, Philipp Tucker, Martin Wicke, Cassandra Xia, Jianwei Xie

Abstract: We present a framework for specifying, training, evaluating, and deploying machine learning models. Our focus is on simplifying cutting edge machine learning for practitioners in order to bring such technologies into production. Recognizing the fast evolution of the field of deep learning, we make no attempt to capture the design space of all possible model architectures in a domain- specific lang… ▽ More

Submitted 8 August, 2017; originally announced August 2017.

Comments: 8 pages, Appeared at KDD 2017, August 13--17, 2017, Halifax, NS, Canada

arXiv:1708.02188
[pdf, ps, other] cs.DC

PowerAI DDL

Authors: Minsik Cho, Ulrich Finkler, Sameer Kumar, David Kung, Vaibhav Saxena, Dheeraj Sreedhar

Abstract: … and bandwidth and adapts to a variety of system configurations. The communication algorithm is implemented as a library for easy use. This library has been integrated into Tensorflow, Caffe, and Torch. We train Resnet-101 on Imagenet 22K with 64 IBM

Power8 S822LC servers (256 GPUs) in about 7 hours to an accuracy of 33.8 % validation accuracy. Microsoft's AD… ▽ More

Submitted 7 August, 2017; originally announced August 2017.

arXiv:1708.01580
[pdf] cs.CV

Sensing Urban Land-Use Patterns By Integrating Google Tensorflow And Scene-Classification Models

Authors: Yao Yao, Haolin Liang, Xia Li, Jinbao Zhang, Jialv He

Abstract: … method in detecting urban land-use patterns, we applied a transfer-learning-based remote-sensing image approach to extract and classify features. Using the Google Tensorflow framework, a powerful convolution neural network (CNN) library was created. First, the transferred model was previously trained on ImageNet, one of the largest object-image data sets, to… ▽ More

Submitted 4 August, 2017; originally announced August 2017.

Comments: 8 pages, 8 figures, 2 tables

arXiv:1707.05390
[pdf, other] cs.AI

TensorLog: Deep Learning Meets Probabilistic DBs

Authors: William W Cohen, Fan Yang, Kathryn Rivard Mazaitis

Abstract: …a probabilistic first-order logic called TensorLog, in which classes of logical queries are compiled into differentiable functions in a neural-network infrastructure such as Tensorflow or Theano. This leads to a close integration of probabilistic logical reasoning with deep-learning infrastructure: in particular, it enables high-performance deep learning fra… ▽ More

Submitted 17 July, 2017; originally announced July 2017.

ACM Class: I.2.4; I.2.6

arXiv:1707.04131
[pdf, ps, other] cs.LG

Foolbox: A Python toolbox to benchmark the robustness of machine learning models

Authors: Jonas Rauber, Wieland Brendel, Matthias Bethge

Abstract: … hyperparameter tuning to find the minimum adversarial perturbation. Additionally, Foolbox interfaces with most popular deep learning frameworks such as PyTorch, Keras, TensorFlow, Theano and MXNet and allows different adversarial criteria such as targeted misclassification and top-k misclassification as well as different distance measures. The code is licens… ▽ More

Submitted 20 March, 2018; v1 submitted 13 July, 2017; originally announced July 2017.

Comments: Code and examples available at https://github.com/bethgelab/foolbox and documentation available at http://foolbox.readthedocs.io

arXiv:1707.03750
[pdf, other] cs.SE

DeepProf: Performance Analysis for Deep Learning Applications via Mining GPU Execution Patterns

Authors: Jiazhen Gu, Huan Liu, Yangfan Zhou, Xin Wang

Abstract: … learning applications. Empirical study verifies the effectiveness of \texttt{DeepProf} in performance analysis and diagnosis. We also find out some interesting properties of Tensorflow, which can be used to guide the deep learning system setup. ▽ More

Submitted 12 July, 2017; originally announced July 2017.

arXiv:1707.00889
[pdf, other] cs.DC

ECHO: An Adaptive Orchestration Platform for Hybrid Dataflows across Cloud and Edge

Authors: Pushkara Ravindra, Aakash Khochare, Siva Prakash Reddy, Sarthak Sharma, Prateeksha Varshney, Yogesh Simmhan

Abstract: … resources. ECHO's hybrid dataflow composition can operate on diverse data models -- streams, micro-batches and files, and interface with native runtime engines like TensorFlow and Storm to execute them. It manages the application's lifecycle, including container-based deployment and a registry for state management. ECHO can schedule the dataflow on different… ▽ More

Submitted 4 July, 2017; originally announced July 2017.

Comments: 17 pages, 5 figures, 2 tables, submitted to ICSOC-2017

arXiv:1707.00703
[pdf, other] cs.NE

Automated Problem Identification: Regression vs Classification via Evolutionary Deep Networks

Authors: Emmanuel Dufourq, Bruce A Bassett

Abstract: …will be applied to the execution of the neural network. We propose the Automated Problem Identification (API) algorithm, which uses an evolutionary algorithm interface to TensorFlow to manipulate a deep neural network to decide if a dataset represents a classification or a regression problem. We test API on 16 different classification, regression and sentime… ▽ More

Submitted 3 July, 2017; originally announced July 2017.

Comments: 9 pages, 6 figures, 4 tables

arXiv:1706.08605
[pdf, other] cs.SE

Developing Bug-Free Machine Learning Systems With Formal Mathematics

Authors: Daniel Selsam, Percy Liang, David L Dill

Abstract: …unbiased estimates of the true mathematical gradients. We train a variational autoencoder using Certigrad and find the performance comparable to training the same model in TensorFlow. ▽ More

Submitted 26 June, 2017; originally announced June 2017.

Comments: To appear at the Thirty-fourth International Conference on Machine Learning (ICML) 2017

arXiv:1706.08217
[pdf, other] stat.ML

An Effective Way to Improve YouTube-8M Classification Accuracy in Google Cloud Platform

Authors: Zhenzhen Zhong, Shujiao Huang, Cheng Zhan, Licheng Zhang, Zhiwei Xiao, ChangChun Wang, Pei Yang

Abstract: …understanding and classification using deep learning algorithms and ensemble methods. We built several baseline predictions according to the benchmark paper and public github tensorflow code. Furthermore, we improved global prediction accuracy (GAP) from base level 77% to 80.7% through approaches of ensemble. ▽ More

Submitted 25 June, 2017; originally announced June 2017.

Comments: 5 pages, 2 figures

arXiv:1706.04972
[pdf, ps, other] cs.LG

Device Placement Optimization with Reinforcement Learning

Authors: Azalia Mirhoseini, Hieu Pham, Quoc V Le, Benoit Steiner, Rasmus Larsen, Yuefeng Zhou, Naveen Kumar, Mohammad Norouzi, Samy Bengio, Jeff Dean

Abstract: …models on devices is often made by human experts based on simple heuristics and intuitions. In this paper, we propose a method which learns to optimize device placement for TensorFlow computational graphs. Key to our method is the use of a sequence-to-sequence model to predict which subsets of operations in a TensorFlow… ▽ More

Submitted 25 June, 2017; v1 submitted 13 June, 2017; originally announced June 2017.

Comments: To appear at ICML 2017

arXiv:1706.04702
[pdf, other] math.NA

Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations

Authors: Weinan E, Jiequn Han, Arnulf Jentzen

Abstract: … condition and the solution of the BSDE. The policy function is then approximated by a neural network, as is done in deep reinforcement learning. Numerical results using TensorFlow illustrate the efficiency and accuracy of the proposed algorithms for several 100-dimensional nonlinear PDEs from physics and finance such as the Allen-Cahn equation, the Hamilton-… ▽ More

Submitted 14 June, 2017; originally announced June 2017.

Comments: 39 pages, 15 figures

MSC Class: 65M75; 60H35; 65C30

arXiv:1706.03292
[pdf, other] cs.LG

Poseidon: An Efficient Communication Architecture for Distributed Deep Learning on GPU Clusters

Authors: Hao Zhang, Zeyu Zheng, Shizhen Xu, Wei Dai, Qirong Ho, Xiaodan Liang, Zhiting Hu, Jinliang Wei, Pengtao Xie, Eric P Xing

Abstract: …each layer, according to layer properties and the number of machines. We show that Poseidon is applicable to different DL frameworks by plugging Poseidon into Caffe and TensorFlow. We show that Poseidon enables Caffe and TensorFlow to achieve 15.5x speed-up on 16 single-GPU machines, even with limited bandwidth (10GbE)… ▽ More

Submitted 10 June, 2017; originally announced June 2017.

Comments: To appear in 2017 USENIX Annual Technical Conference

arXiv:1706.02248
[pdf] cs.LG
doi
10.1109/STC-CSIT.2017.8098808

Comparative Analysis of Open Source Frameworks for Machine Learning with Use Case in Single-Threaded and Multi-Threaded Modes

Authors: Yuriy Kochura, Sergii Stirenko, Anis Rojbi, Oleg Alienin, Michail Novotarskiy, Yuri Gordienko

Abstract: The basic features of some of the most versatile and popular open source frameworks for machine learning (TensorFlow, Deep Learning4j, and H2O) are considered and compared. Their comparative analysis was performed and conclusions were made as to the advantages and disadvantages of these platforms. The performance tests for the de facto standard MNIST data se… ▽ More

Submitted 7 June, 2017; originally announced June 2017.

Comments: 4 pages, 6 figures, 4 tables; XIIth International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT 2017), Lviv, Ukraine

Journal ref: Proceedings of 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 5-8 Sept. 2017, (Lviv, Ukraine), vol.1, pp.

373-376, IEEE

arXiv:1706.01763
[pdf, other] cs.CR

Adversarial-Playground: A Visualization Suite for Adversarial Sample Generation

Authors: Andrew Norton, Yanjun Qi

Abstract: … web-based visualization tool, Adversarial-Playground, to demonstrate the efficacy of common adversarial methods against a deep neural network (DNN) model, built on top of the TensorFlow library. Adversarial-Playground provides users an efficient and effective experience in exploring techniques generating adversarial examples, which are inputs crafted by an a… ▽ More

Submitted 16 June, 2017; v1 submitted 6 June, 2017; originally announced June 2017.

Comments: 8 pages; 3 figures

arXiv:1706.00231
[pdf, other] cs.MS

Automatic Differentiation using Constraint Handling Rules in Prolog

Authors: Samer Abdallah

Abstract: … for parameter learning in probabilistic grammars, the CHR based implementations outperformed two well-known frameworks for optimising differentiable functions, Theano and TensorFlow, by a large margin. ▽ More

Submitted 1 June, 2017; originally announced June 2017.

arXiv:1705.10246
[pdf, other] stat.ML

The Principle of Logit Separation

Authors: Gil Keren, Sivan Sabato, Björn Schuller

Abstract: …it. We therefore conclude that the Principle of Logit Separation sheds light on an important property of the most common loss functions used by neural network classifiers. Tensorflow code for optimizing the new batch losses is publicly available in https://github.com/cruvadom/Logit_Separation. ▽ More

Submitted 26 October, 2017; v1 submitted 29 May, 2017; originally announced May 2017.

arXiv:1705.09056
[pdf, other] math.OC

Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent

Authors: Xiangru Lian, Ce Zhang, Huan Zhang, ChoJui Hsieh, Wei Zhang, Ji Liu

Abstract: Most distributed machine learning systems nowadays, including TensorFlow and CNTK, are built in a centralized fashion. One bottleneck of centralized algorithms lies on high communication cost on the central node. Motivated by this, we ask, can decentralized algorithms be faster than its centralized counterpart? Although decentralized PSGD (D-PSGD) algorith… ▽ More

Submitted 11 September, 2017; v1 submitted 25 May, 2017; originally announced May 2017.

arXiv:1705.07860
[pdf, other] cs.LG

On-the-fly Operation Batching in Dynamic Computation Graphs

Authors: Graham Neubig, Yoav Goldberg, Chris Dyer

Abstract: … more flexibility for implementing models that cope with data of varying dimensions and structure, relative to toolkits that operate on statically declared computations (e.g., TensorFlow, CNTK, and Theano). However, existing toolkits - both static and dynamic - require that the developer organize the computations into the batches necessary for exploiting high… ▽ More

Submitted 22 May, 2017; originally announced May 2017.

arXiv:1705.06936
[pdf, other] cs.DC
doi
10.1007/978-3-319-75931-9_1

Atari games and Intel processors

Authors: Robert Adamski, Tomasz Grel, Maciej Klimek, Henryk Michalewski

Abstract: … is spent performing convolutions. In this work we present our results on learning strategies in Atari games using a Convolutional Neural Network, the Math Kernel Library and TensorFlow 0.11rc0 machine learning framework. We also analyze effects of asynchronous computations on the convergence of reinforcement learning algorithms. ▽ More

Submitted 19 May, 2017; originally announced May 2017.

arXiv:1705.05922
[pdf, other] cs.CV

LCDet: Low-Complexity Fully-Convolutional Neural Networks for Object Detection in Embedded Systems

Authors: Subarna Tripathi, Gokce Dane, Byeongkeun Kang, Vasudev Bhaskaran, Truong Nguyen

Abstract: …In this work, we propose LCDet, a fully-convolutional neural network for generic object detection that aims to work in embedded systems. We design and develop an end-to-end TensorFlow(TF)-based model. Additionally, we employ 8-bit quantization on the learned weights. We use face detection as a use case. Our TF-Slim based network can predict different faces… ▽ More

Submitted 16 May, 2017; originally announced May 2017.

Comments: Embedded Vision Workshop in CVPR

arXiv:1705.05627
[pdf] cs.CV
doi
10.5334/jors.178

Picasso: A Modular Framework for Visualizing the Learning Process of Neural Network Image Classifiers

Authors: Ryan Henderson, Rasmus Rothe

Abstract: … visualizations which help reveal issues that evaluation metrics like loss and accuracy might hide: for example, learning a proxy classification task. Picasso works with the Tensorflow deep learning framework, and Keras (when the model can be loaded into the Tensorflow backend). Picasso can be used with minimal configur… ▽ More

Submitted 11 September, 2017; v1 submitted 16 May, 2017; originally announced May 2017.

Comments: 9 pages, submission to the Journal of Open Research Software, github.com/merantix/picasso

Journal ref: Journal of Open Research Software. 5(1), p.22 (2017)

arXiv:1705.02583
[pdf, other] cs.LG

A Design Methodology for Efficient Implementation of Deconvolutional Neural Networks on an FPGA

Authors: Xinyu Zhang, Srinjoy Das, Ojash Neopane, Ken Kreutz-Delgado

Abstract: …our FPGA deconvolutional accelerator design methodology we train DCNNs offline on two representative datasets using the generative adversarial network method (GAN) run on Tensorflow, and then map these DCNNs to an FPGA DCNN-plus-accelerator implementation to perform generative inference on a Xilinx Zynq-7000 FPGA. Our DCNN implementation achieves a peak perf… ▽ More

Submitted 7 May, 2017; originally announced May 2017.

arXiv:1705.02414
[pdf, other] cs.LG

A comprehensive study of batch construction strategies for recurrent neural networks in MXNet

Authors: Patrick Doetsch, Pavel Golik, Hermann Ney

Abstract: In this work we compare different batch construction methods for mini-batch training of recurrent neural networks. While popular implementations like TensorFlow and MXNet suggest a bucketing approach to improve the parallelization capabilities of the recurrent training process, we propose a simple ordering strategy that arranges the training sequences in a s… ▽ More

Submitted 5 May, 2017; originally announced May 2017.

arXiv:1705.01507
[pdf, other] cs.LG

XES Tensorflow - Process Prediction using the Tensorflow Deep-Learning Framework

Authors: Joerg Evermann, JanaRebecca Rehse, Peter Fettke

Abstract: …neural networks, so called deep-learning approaches, have been proposed to address this challenge. This demo paper describes a software application that applies the Tensorflow deep-learning framework to process prediction. The software application reads industry-standard XES files for training and presents the user with an easy-to-use graphical user interfac… ▽ More

Submitted 3 May, 2017; originally announced May 2017.

arXiv:1704.07511
[pdf, ps, other] cs.LG

Scalable Planning with Tensorflow for Hybrid Nonlinear Domains

Authors: Ga Wu, Buser Say, Scott Sanner

Abstract: … optimize high-dimensional non-convex functions with gradient descent optimization on GPUs, we ask in this paper whether symbolic gradient optimization tools such as Tensorflow can be effective for planning in hybrid (mixed discrete and continuous) nonlinear domains with high dimensional state and action spaces? To this end, we demonstrate that hybrid plannin… ▽ More

Submitted 4 November, 2017; v1 submitted 24 April, 2017; originally announced April 2017.

Comments: 9 pages

arXiv:1704.04760
[pdf] cs.AR

In-Datacenter Performance Analysis of a Tensor Processing Unit

Authors: Norman P Jouppi, Cliff Young, Nishant Patil, David Patterson, Gaurav Agrawal, Raminder Bajwa, Sarah Bates, Suresh Bhatia, Nan Boden, Al Borchers, Rick Boyle, Pierreluc Cantin, Clifford Chao, Chris Clark, Jeremy Coriell, Mike Daley, Matt Dau, Jeffrey Dean, Ben Gelb, Tara Vazir Ghaemmaghami, Rajendra Gottipati, William Gulland, Robert Hagmann, C Richard Ho, Doug Hogberg , et al. (50 additional authors not shown)

Abstract: …compare the TPU to a server-class Intel Haswell CPU and an Nvidia K80 GPU, which are contemporaries deployed in the same datacenters. Our workload, written in the high-level TensorFlow framework, uses production NN applications (MLPs, CNNs, and LSTMs) that represent 95% of our datacenters' NN inference demand. Despite low utilization for some applications, t… ▽ More

Submitted 16 April, 2017; originally announced April 2017.

Comments: 17 pages, 11 figures, 8 tables. To appear at the 44th International Symposium on Computer Architecture (ISCA), Toronto, Canada, June 24-28, 2017

arXiv:1704.04560
[pdf, other] cs.DC

User-transparent Distributed TensorFlow

Authors: Abhinav Vishnu, Joseph Manzano, Charles Siegel, Jeff Daily

Abstract: …Learning (DL) algorithms have become the {\em de facto} choice for data analysis. Several DL implementations -- primarily limited to a single compute node -- such as Caffe, TensorFlow, Theano and Torch have become readily available. Distributed DL implementations capable of execution on large scale systems are becoming important to address the computational… ▽ More

Submitted 14 April, 2017; originally announced April 2017.

Comments: 9 pages, 8 figures

arXiv:1704.03751
[pdf] cs.LG

Enabling Embedded Inference Engine with ARM Compute Library: A Case Study

Authors: Dawei Sun, Shaoshan Liu, JeanLuc Gaudiot

Abstract: …to build an inference engine from scratch compared to porting existing frameworks. In addition, by utilizing ACL, we managed to build an inference engine that outperforms TensorFlow by 25%. Our conclusion is that, on embedded devices, we most likely will use very simple deep learning models for inference, and with well-developed building blocks such as ACL,… ▽ More

Submitted 14 April, 2017; v1 submitted 12 April, 2017; originally announced April 2017.

Comments: 4 pages, 4 figures

arXiv:1703.05298
[pdf, ps, other] cs.LG

Neural Networks for Beginners. A fast implementation in Matlab, Torch, TensorFlow

Authors: Francesco Giannini, Vincenzo Laveglia, Alessandro Rossi, Dario Zanca, Andrea Zugarini

Abstract: This report provides an introduction to some Machine Learning tools within the most common development environments. It mainly focuses on practical problems, skipping any theoretical introduction. It is oriented to both students trying to approach Machine Learning and experts looking for new frameworks. ▽ More

Submitted 16 March, 2017; v1 submitted 10 March, 2017; originally announced March 2017.

arXiv:1703.01775
[pdf, ps, other] cs.CV

Building a Regular Decision Boundary with Deep Networks

Authors: Edouard Oyallon

Abstract: …regular. Besides, we defined and analyzed local support vectors that separate classes locally. All our experiments are reproducible and code is available online, based on TensorFlow. ▽ More

Submitted 6 March, 2017; originally announced March 2017.

Comments: CVPR 2017, 8 pages

arXiv:1702.07398
[pdf, other] stat.ML

Deep Nonparametric Estimation of Discrete Conditional Distributions via Smoothed Dyadic Partitioning

Authors: Wesley Tansey, Karl Pichotta, James G Scott

Abstract: …and real-world datasets, in some cases reducing the error by nearly half in comparison to other popular methods in the literature. All of our models are implemented in Tensorflow and publicly available at https://github.com/tansey/sdp . ▽ More

Submitted 28 February, 2017; v1 submitted 23 February, 2017; originally announced February 2017.

arXiv:1702.04683
[pdf, other] cs.LG

Distributed deep learning on edge-devices: feasibility via adaptive compression

Authors: Corentin Hardy, Erwan Le Merrer, Bruno Sericola

Abstract: …techniques, such as deep learning. The state-of-the-art results by deep learning come at the price of an intensive use of computing resources. The leading frameworks (e.g., TensorFlow) are executed on GPUs or on high-end servers in datacenters. On the other end, there is a proliferation of personal devices with possibly free CPU cycles; this can enable servi… ▽ More

Submitted 6 November, 2017; v1 submitted 15 February, 2017; originally announced February 2017.

Comments: Best paper award at IEEE International Symposium on Network Computing and Applications (NCA 2017)

arXiv:1702.04389
[pdf, other] cs.AI

Entropy Non-increasing Games for the Improvement of Dataflow Programming

Authors: Norbert Bátfai, Renátó Besenczi, Gergő Bogacsovics, Fanny Monori

Abstract: …this article, we introduce a new conception of a family of esport games called Samu Entropy to try to improve dataflow program graphs like the ones that are based on Google's TensorFlow. Currently, the Samu Entropy project specifies only requirements for new esport games to be developed with particular attention to the investigation of the relationship betwe… ▽ More

Submitted 14 February, 2017; originally announced February 2017.

Comments: 15 pages, 7 figures

MSC Class: 68T01 ACM Class: I.2.1

arXiv:1702.03865
[pdf, other] cs.LG

Next-Step Conditioned Deep Convolutional Neural Networks Improve Protein Secondary Structure Prediction

Authors: Akosua Busia, Navdeep Jaitly

Abstract: …71.4% Q8 accuracy on the same test set, improving upon the previous overall state of the art for the eight-class secondary structure problem. Our models are implemented using TensorFlow, an open-source machine learning software library available at TensorFlow.org; we aim to release the code for these experiments as part of the … ▽ More

Submitted 13 February, 2017; originally announced February 2017.

Comments: 11 pages, 3 figures, 4 tables, submitted to ISMB/ECCB 2017. arXiv admin note: text overlap with arXiv:1611.01503

arXiv:1702.02138
[pdf, ps, other] cs.CV

An Implementation of Faster RCNN with Study for Region Sampling

Authors: Xinlei Chen, Abhinav Gupta

Abstract: We adapted the join-training scheme of Faster RCNN framework from Caffe to TensorFlow as a baseline implementation for object detection. Our code is made publicly available. This report documents the simplifications made to the original pipeline, with justifications from ablation analysis on both PASCAL VOC 2007 and COCO 2014. We further investigated the rol… ▽ More

Submitted 8 February, 2017; v1 submitted 7 February, 2017; originally announced February 2017.

Comments: Technical Report, 3 pages

arXiv:1701.03980
[pdf, other] stat.ML

DyNet: The Dynamic Neural Network Toolkit

Authors: Graham Neubig, Chris Dyer, Yoav Goldberg, Austin Matthews, Waleed Ammar, Antonios Anastasopoulos, Miguel Ballesteros, David Chiang, Daniel Clothiaux, Trevor Cohn, Kevin Duh, Manaal Faruqui, Cynthia Gan, Dan Garrette, Yangfeng Ji, Lingpeng Kong, Adhiguna Kuncoro, Gaurav Kumar, Chaitanya Malaviya, Paul Michel, Yusuke Oda, Matthew Richardson, Naomi Saphra, Swabha Swayamdipta, Pengcheng Yin

Abstract: …for implementing neural network models based on dynamic declaration of network structure. In the static declaration strategy that is used in toolkits like Theano, CNTK, and TensorFlow, the user first defines a computation graph (a symbolic representation of the computation), and then examples are fed into an engine that executes this computation and computes… ▽ More

Submitted 14 January, 2017; originally announced January 2017.

Comments: 33 pages

arXiv:1701.03757
[pdf, ps, other] stat.ML

Deep Probabilistic Programming

Authors: Dustin Tran, Matthew D Hoffman, Rif A Saurous, Eugene Brevdo, Kevin Murphy, David M Blei

Abstract: …representation as part of inference, facilitating the design of rich variational models and generative adversarial networks. For efficiency, Edward is integrated into TensorFlow, providing significant speedups over existing probabilistic systems. For example, we show on a benchmark logistic regression task that Edward is at least 35x faster than Stan and 6x… ▽ More

Submitted 7 March, 2017; v1 submitted 13 January, 2017; originally announced January 2017.

Comments: Appears in International Conference on Learning Representations, 2017. A companion webpage for this paper is available at http://edwardlib.org/iclr2017

arXiv:1701.02284
[pdf, other] cs.PL

DeepDSL: A Compilation-based Domain-Specific Language for Deep Learning

Authors: Tian Zhao, Xiaobing Huang, Yu Cao

Abstract: …such as multimedia understanding. However, the complex nature of multimedia data makes it difficult to develop DL-based software. The state-of-the art tools, such as Caffe, TensorFlow, Torch7, and CNTK, while are successful in their applicable domains, are programming libraries with fixed user interface, internal representation, and execution environment. Th… ▽ More

Submitted 9 January, 2017; originally announced January 2017.

arXiv:1701.00609
[pdf, other] cs.LG

Akid: A Library for Neural Network Research and Production from a Dataism Approach

Authors: Shuai Li

Abstract: … named {\texttt akid}. It provides higher level of abstraction for entities (abstracted as blocks) in nature upon the abstraction done on signals (abstracted as tensors) by Tensorflow, characterizing the dataism observation that all entities in nature processes input and emit out in some ways. It includes a full stack of software that provides abstraction to… ▽

More

Submitted 3 January, 2017; originally announced January 2017.

arXiv:1612.08882
[pdf, other] cs.MM

Improving Blind Steganalysis in Spatial Domain using a Criterion to Choose the Appropriate Steganalyzer between CNN and SRM+EC

Authors: JeanFrancois Couchot, Raphaël Couturier, Michel Salomon

Abstract: …CNN or the SRM+EC method for a given input image. Our approach is studied with three different steganographic spatial domain algorithms: S-UNIWARD, MiPOD, and HILL, using the Tensorflow computing platform, and exhibits detection capabilities better than each method alone. Furthermore, as SRM+EC and the CNN are both only trained with a single embedding algori… ▽ More

Submitted 9 January, 2017; v1 submitted 28 December, 2016; originally announced December 2016.

arXiv:1612.06321
[pdf, other] cs.CV

Large-Scale Image Retrieval with Attentive Deep Local Features

Authors: Hyeonwoo Noh, Andre Araujo, Jack Sim, Tobias Weyand, Bohyung Han

Abstract: …the state-of-the-art global and local descriptors in the large-scale setting by significant margins. Code and dataset can be found at the project webpage: https://github.com/tensorflow/models/tree/master/research/delf . ▽ More

Submitted 2 February, 2018; v1 submitted 19 December, 2016; originally announced December 2016.

Comments: ICCV 2017. Code and dataset available: https://github.com/tensorflow/models/tree/master/research/delf

arXiv:1612.05086
[pdf, ps, other] cs.LG

Coupling Adaptive Batch Sizes with Learning Rates

Authors: Lukas Balles, Javier Romero, Philipp Hennig

Abstract: …two. On popular image classification benchmarks, our batch size adaptation yields faster optimization convergence, while simultaneously simplifying learning rate tuning. A TensorFlow implementation is available. ▽ More

Submitted 28 June, 2017; v1 submitted 15 December, 2016; originally announced December 2016.

Comments: Thirty-Third Conference on Uncertainty in Artificial Intelligence (UAI), 2017, (accepted)

arXiv:1612.04251
[pdf, ps, other] cs.DC

TF.Learn: TensorFlow's High-level Module for Distributed Machine Learning

Authors: Yuan Tang

Abstract: TF.Learn is a high-level Python module for distributed machine learning inside TensorFlow. It provides an easy-to-use Scikit-learn style interface to simplify the process of creating, configuring, training, evaluating, and experimenting a machine learning model. TF.Learn integrates a wide range of state-of-art machine learning algorithms built on top of Tens… ▽ More

Submitted 13 December, 2016; originally announced December 2016.

arXiv:1612.03079
[pdf, other] cs.DC

Clipper: A Low-Latency Online Prediction Serving System

Authors: Daniel Crankshaw, Xin Wang, Giulio Zhou, Michael J Franklin, Joseph E Gonzalez, Ion Stoica

Abstract: …benchmark datasets and demonstrate its ability to meet the latency, accuracy, and throughput demands of online serving applications. Finally, we compare Clipper to the TensorFlow Serving system and demonstrate that we are able to achieve comparable throughput and latency while enabling model composition and online learning to improve accuracy and render more… ▽ More

Submitted 28 February, 2017; v1 submitted 9 December, 2016; originally announced December 2016.

arXiv:1612.02485

[pdf, other] cs.DB

Comparative Evaluation of Big-Data Systems on Scientific Image Analytics Workloads

Authors: Parmita Mehta, Sven Dorkenwald, Dongfang Zhao, Tomer Kaftan, Alvin Cheung, Magdalena Balazinska, Ariel Rokem, Andrew Connolly, Jacob Vanderplas, Yusra AlSayyad

Abstract: …of large-scale image analysis systems using two real-world scientific image data processing use cases. We evaluate five representative systems (SciDB, Myria, Spark, Dask, and TensorFlow) and find that each of them has shortcomings that complicate implementation or hurt performance. Such shortcomings lead to new research opportunities in making large-scale im… ▽ More

Submitted 7 December, 2016; originally announced December 2016.

arXiv:1611.08903
[pdf, other] cs.LG

Should I use TensorFlow

Authors: Martin Schrimpf

Abstract: Google's Machine Learning framework TensorFlow was open-sourced in November 2015 [1] and has since built a growing community around it. TensorFlow is supposed to be flexible for research purposes while also allowing its models to be deployed productively. This work is aimed towards people with experience in Machine Lea… ▽ More

Submitted 27 November, 2016; originally announced November 2016.

Comments: Seminar Paper

arXiv:1611.06256
[pdf, other] cs.LG

Reinforcement Learning through Asynchronous Advantage Actor-Critic on a GPU

Authors: Mohammad Babaeizadeh, Iuri Frosio, Stephen Tyree, Jason Clemons, Jan Kautz

Abstract: … We introduce a system of queues and a dynamic scheduling strategy, potentially helpful for other asynchronous algorithms as well. Our hybrid CPU/GPU version of A3C, based on TensorFlow, achieves a significant speed up compared to a CPU implementation; we make it publicly available to other researchers at https://github.com/NVlabs/GA3C . ▽ More

Submitted 2 March, 2017; v1 submitted 18 November, 2016; originally announced November 2016.

arXiv:1611.01503
[pdf, other] cs.LG

Protein Secondary Structure Prediction Using Deep Multi-scale Convolutional Neural Networks and Next-Step Conditioning

Authors: Akosua Busia, Jasmine Collins, Navdeep Jaitly

Abstract: … in prediction accuracy through more sophisticated attempts to control overfitting of conditional models. We aim to release the code for these experiments as part of the TensorFlow repository. ▽ More

Submitted 4 November, 2016; originally announced November 2016.

Comments: 10 pages, 2 figures, submitted to RECOMB 2017

arXiv:1610.09787
[pdf, other] stat.CO

Edward: A library for probabilistic modeling, inference, and criticism

Authors: Dustin Tran, Alp Kucukelbir, Adji B Dieng, Maja Rudolph, Dawen Liang, David M Blei

Abstract: …to the data. Edward supports a broad class of probabilistic models, efficient algorithms for inference, and many techniques for model criticism. The library builds on top of TensorFlow to support distributed training and hardware such as GPUs. Edward enables the development of complex probabilistic models and their algorithms at a massive scale. ▽ More

Submitted 31 January, 2017; v1 submitted 31 October, 2016; originally announced October 2016.

arXiv:1610.08733
[pdf, other] stat.ML

GPflow: A Gaussian process library using TensorFlow

Authors: Alexander G de G Matthews, Mark van der Wilk, Tom Nickson, Keisuke Fujii, Alexis Boukouvalas, Pablo León-Villagrá, Zoubin Ghahramani, James Hensman

Abstract: GPflow is a Gaussian process library that uses TensorFlow for its core computations and Python for its front end. The distinguishing features of GPflow are that it uses variational inference as the primary approximation method, provides concise code through the use of automatic differentiation, has been engineered with a particular emphasis on software testi⋯ ▽ More

Submitted 27 October, 2016; originally announced October 2016.

arXiv:1610.01178
[pdf, other] cs.LG

A Tour of TensorFlow

Authors: Peter Goldsborough

Abstract: ⋯ has significantly advanced the state-of-the-art in computer vision, speech recognition, natural language processing and other domains. In November 2015, Google released TensorFlow, an open source deep learning software library for defining, training and deploying machine learning models. In this paper, we review Te⋯ ▽ More

Submitted 1 October, 2016; originally announced October 2016.

arXiv:1609.08675
[pdf, other] cs.CV

YouTube-8M: A Large-Scale Video Classification Benchmark

Authors: Sami Abu-El-Haija, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, Sudheendra Vijayanarasimhan

Abstract: ⋯ evaluation metrics, and report them as baselines. Despite the size of the dataset, some of our models train to convergence in less than a day on a single machine using TensorFlow. We plan to release code for training a TensorFlow model and for computing metrics. ▽ More

Submitted 27 September, 2016; originally announced September 2016.

Comments: 10 pages

arXiv:1609.06647
[pdf, other] cs.CV
doi
10.1109/TPAMI.2016.2587640

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

Authors: Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan

Abstract: …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. ▽ More

Submitted 21 September, 2016; originally announced September 2016.

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

Journal ref: IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: PP, Issue: 99 , July 2016)

arXiv:1609.03488
[pdf, other] math.OC

A New Architecture for Optimization Modeling Frameworks

Authors: Matt Wytock, Steven Diamond, Felix Heide, Stephen Boyd

Abstract: We propose a new architecture for optimization modeling frameworks in which solvers are expressed as computation graphs in a framework like TensorFlow rather than as standalone programs built on a low-level linear algebra interface. Our new architecture makes it easy for modeling frameworks to support high performance computational platforms like GPUs and di… ▽ More

Submitted 11 October, 2016; v1 submitted 12 September, 2016; originally announced September 2016.

arXiv:1608.07249
[pdf, other] cs.DC

Benchmarking State-of-the-Art Deep Learning Software Tools

Authors: Shaohuai Shi, Qiang Wang, Pengfei Xu, Xiaowen Chu

Abstract: …software and hardware. In this paper, we aim to make a comparative study of the state-of-the-art GPU-accelerated deep learning software tools, including Caffe, CNTK, MXNet, TensorFlow, and Torch. We first benchmark the running performance of these tools with three popular types of neural networks on two CPU platforms and three GPU platforms. We then benchmar… ▽ More

Submitted 17 February, 2017; v1 submitted 25 August, 2016; originally announced August 2016.

Comments: Revision history: 1. Revise ResNet-50 configuration in MXNet. 2. Add faster implementation of ResNet-56 in TensorFlow with multiple GPUs

arXiv:1608.06581
[pdf, other] cs.LG
doi
10.1109/IISWC.2016.7581275

Fathom: Reference Workloads for Modern Deep Learning Methods

Authors: Robert Adolf, Saketh Rama, Brandon Reagen, GuYeon Wei, David Brooks

Abstract: …online, and this paper focuses on understanding the fundamental performance characteristics of each model. We use a set of application-level modeling tools built around the TensorFlow deep learning framework in order to analyze the behavior of the Fathom workloads. We present a breakdown of where time is spent, the similarities between the performance profil… ▽ More

Submitted 23 August, 2016; originally announced August 2016.

Comments: Proceedings of the IEEE International Symposium on Workload Characterization, 2016

arXiv:1606.07792
[pdf, other] cs.LG

Wide & Deep Learning for Recommender Systems

Authors: HengTze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, Hemal Shah

Abstract: … results show that Wide & Deep significantly increased app acquisitions compared with wide-only and deep-only models. We have also open-sourced our implementation in TensorFlow. ▽ More

Submitted 24 June, 2016; originally announced June 2016.

arXiv:1606.04422
[pdf, ps, other] cs.AI

Logic Tensor Networks: Deep Learning and Logical Reasoning from Data and Knowledge

Authors: Luciano Serafini, Artur dAvila Garcez

Abstract: …on a knowledge-base and efficient data-driven relational machine learning. We show how Real Logic can be implemented in deep Tensor Neural Networks with the use of Google's tensorflow primitives. The paper concludes with experiments applying Logic Tensor Networks on a simple but representative example of knowledge completion. ▽ More

Submitted 7 July, 2016; v1 submitted 14 June, 2016; originally announced June 2016.

Comments: 12 pages, 2 figs, 1 table, 27 references

arXiv:1605.08695
[pdf, other] cs.DC

TensorFlow: A system for large-scale machine learning

Authors: Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, Xiaoqiang Zheng

Abstract: TensorFlow is a machine learning system that operates at large scale and in heterogeneous environments. TensorFlow uses dataflow graphs to represent computation, shared state, and the operations that mutate that state. It maps the nodes of a dataflow graph across many machines in a cluster, and within a machine across… ▽ More

Submitted 31 May, 2016; v1 submitted 27 May, 2016; originally announced May 2016.

Comments: 18 pages, 9 figures; v2 has a spelling correction in the metadata

arXiv:1605.04614
[pdf, other] cs.LG

DeepLearningKit - an GPU Optimized Deep Learning Framework for Apple's iOS, OS X and tvOS developed in Metal and Swift

Authors: Amund Tveit, Torbjørn Morland, Thomas Brox Røst

Abstract: …tvOS-based apps for the big screen, or OS X desktop applications. The goal is to support using deep learning models trained with popular frameworks such as Caffe, Torch, TensorFlow, Theano, Pylearn, Deeplearning4J and Mocha. Given the massive GPU resources and time required to train Deep Learning models we suggest an App Store like model to distribute and d… ▽ More

Submitted 15 May, 2016; originally announced May 2016.

Comments: 9 pages, 12 figures, open source documentation and code at deeplearningkit.org and github.com/deeplearningkit

arXiv:1605.02688
[pdf, other] cs.SC

Theano: A Python framework for fast computation of mathematical expressions

Authors: The Theano Development Team, Rami Al-Rfou, Guillaume Alain, Amjad Almahairi, Christof Angermueller, Dzmitry Bahdanau, Nicolas Ballas, Frédéric Bastien, Justin Bayer, Anatoly Belikov, Alexander Belopolsky, Yoshua Bengio, Arnaud Bergeron, James Bergstra, Valentin Bisson, Josh Bleecher Snyder, Nicolas Bouchard, Nicolas Boulanger-Lewandowski, Xavier Bouthillier, Alexandre de Brébisson, Olivier Breuleux, PierreLuc Carrier, Kyunghyun Cho, Jan Chorowski, Paul Christiano , et al. (88 additional authors not shown)

Abstract: … with other similar projects. Section III focuses on recently-introduced functionalities and improvements. Section IV compares the performance of Theano against Torch7 and TensorFlow on several machine learning models. Section V discusses current limitations of Theano and potential ways of improving it. ▽ More

Submitted 9 May, 2016; originally announced May 2016.

Comments: 19 pages, 5 figures

arXiv:1603.04467
[pdf, other] cs.DC

TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems

Authors: Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah , et al. (15 additional authors not shown)

Abstract: TensorFlow is an interface for expressing machine learning algorithms, and an implementation for executing such algorithms. A computation expressed using TensorFlow can be executed with little or no change on a wide variety of heterogeneous systems, ranging from mobile devices such as phones and tablets up to large-sca… ▽ More

Submitted 16 March, 2016; v1 submitted 14 March, 2016; originally announced March 2016.

Comments: Version 2 updates only the metadata, to correct the formatting of Martín Abadi's name

arXiv:1603.02339
[pdf, other] cs.DC

Distributed TensorFlow with MPI

Authors: Abhinav Vishnu, Charles Siegel, Jeffrey Daily

Abstract: …of open source MLDM software is limited to sequential execution with a few supporting multi-core/many-core execution. In this paper, we extend recently proposed Google TensorFlow for execution on large scale clusters using Message Passing Interface (MPI). Our approach requires minimal changes to the TensorFlow runtim… ▽ More

Submitted 18 August, 2017; v1 submitted 7 March, 2016; originally announced March 2016.

Comments: 6 pages; fixed significant typo

arXiv:1602.08191
[pdf, other] cs.LG

DeepSpark: A Spark-Based Distributed Deep Learning Framework for Commodity Clusters

Authors: Hanjoo Kim, Jaehong Park, Jaehee Jang, Sungroh Yoon

Abstract: … learning framework that exploits Apache Spark on commodity clusters. To support parallel operations, DeepSpark automatically distributes workloads and parameters to Caffe/Tensorflow-running nodes using Spark, and iteratively aggregates training results by a novel lock-free asynchronous variant of the popular elastic averaging stochastic gradient descent base… ▽ More

Submitted 30 September, 2016; v1 submitted 25 February, 2016; originally announced February 2016.

arXiv:1602.05875
[pdf, other] stat.ML

Convolutional RNN: an Enhanced Model for Extracting Features from Sequential Data

Authors: Gil Keren, Björn Schuller

Abstract: …features. Using our convolutional recurrent layers we obtain an improvement in performance in two audio classification tasks, compared to traditional convolutional layers.

Tensorflow code for the convolutional recurrent layers is publicly available in https://github.com/cruvadom/Convolutional-RNN. ▽ More

Submitted 20 July, 2017; v1 submitted 18 February, 2016; originally announced February 2016.

arXiv:1511.06435
[pdf, ps, other] cs.LG

Comparative Study of Deep Learning Software Frameworks

Authors: Soheil Bahrampour, Naveen Ramakrishnan, Lukas Schott, Mohak Shah

Abstract: …software frameworks have been developed to facilitate their implementation. This paper presents a comparative study of five deep learning frameworks, namely Caffe, Neon, TensorFlow, Theano, and Torch, on three aspects: extensibility, hardware utilization, and speed. The study is performed on several types of deep learning architectures and we evaluate the pe… ▽ More

Submitted 29 March, 2016; v1 submitted 19 November, 2015; originally announced November 2015.

Comments: Submitted to KDD 2016 with TensorFlow results added. At the time of submission to KDD, TensorFlow was available only with cuDNN v.2 and thus its performance is reported with that version