ELAD HOFFER

PERSONAL INFORMATION

Born in Israel, 11 October 1986

email elad.hoffer@gmail.com

website http://www.DeepLearning.co.il

github eladhoffer

phone (M)054-4228891

PROFESSIONAL INTERESTS

Machine Learning, Deep Learning, Computer Vision, Signal Processing

WORK EXPERIENCE

2015–2016 Deep Learning researcher, ICRI-CI

Intel Deep Learning researcher at Intel's Collaborative Research Center for Computational Intelligence.

2013–2015 Visual Algorithms and Neural Networks, CVG

Intel Researching and developing Deep Learning capabilities for computer-vision tasks. Part of the Algorithms team in Intel's Computer-Vision Group.

2011-2013 Emulation student, Network Division

Intel Created FPGA prototypes and automation scripts for emulation of network devices.

2005-2009 Commanding Officer , Artillery Corps

IDF Rank: *Captain (Reserve duty)*
Served as a battery commander (artillery), leading 90 soldiers.
Currently serves on active reserve duty.

EDUCATION

2014-Present Technion, Israel Institute of Technology

PhD (Direct track) Electrical Engineering

Research: *Deep Learning of Representations*
Description: My research explores the machine learning technique known as "Deep Learning" which uses artificial neural networks to learn useful data representations.
Advisor: Prof. Daniel Soudry

2010-2014 Technion, Israel Institute of Technology

BSc Electrical Engineering GPA: 90 · *Cum Laude*
Specialized in Computer Engineering, Signal Processing.
Final Project: Real-Time Movie Subtitles Extraction - using image processing and computer-vision techniques.

2004 High-school "Itzhak Rabin", Gan-Yavne

Graduated with honors
Studied Computer Science and Physics
Final Project: Handwriting Recognition.

PUBLICATIONS - IN PROCEEDINGS

Oct. 2015 **Deep metric learning using Triplet network**

SIMBAD

In this paper we propose the deep triplet network model, which aims to learn useful representations by distance comparisons. We demonstrate using various datasets that our model learns a better representation than that of its immediate competitor, the Siamese network, and discuss future possible usage as a framework for unsupervised learning.
Authors: Elad Hoffer, Nir Ailon

Dec. 2017 **Train longer, generalize better: closing the generalization gap in large batch training of neural networks**

NIPS 2017
Oral presentation

We examine the initial high learning rate training phase. We find that the weight distance from its initialization grows logarithmically with the number of weight updates. We therefore propose a "random walk on random landscape" statistical model which is known to exhibit similar "ultra-slow" diffusion behavior. Following this hypothesis we conducted experiments to show empirically that the "generalization gap" stems from the relatively small number of updates rather than the batch size, and can be completely eliminated by adapting the training regime used.
Authors: Elad Hoffer, Itay Hubara, Daniel Soudry

April 2018 **Fix your classifier: the marginal value of training the last weight layer**

ICLR 2018

Neural networks are commonly used as models for classification for a wide variety of tasks. Typically, a learned affine transformation is placed at the end of such models, yielding a per-class value used for classification. This classifier can have a vast number of parameters, which grows linearly with the number of possible classes, thus requiring increasingly more resources. In this work we argue that this classifier can be fixed, up to a global scale constant, with little or no loss of accuracy for most tasks, allowing memory and computational benefits. Moreover, we show that by initializing the classifier with a Hadamard matrix we can speed up inference as well. We discuss the implications for current understanding of neural network models.
Authors: Elad Hoffer, Itay Hubara, Daniel Soudry

April 2018 **The Implicit Bias of Gradient Descent on Separable Data**

ICLR 2018

We show that gradient descent on an unregularized logistic regression problem with separable data converges to the max-margin solution. The result generalizes also to other monotone decreasing loss functions with an infimum at infinity, and we also discuss a multi-class generalizations to the cross entropy loss. Furthermore, we show this convergence is very slow, and only logarithmic in the convergence of the loss itself. This can help explain the benefit of continuing to optimize the logistic or cross-entropy loss even after the training error is zero and the training loss is extremely small, and, as we show, even if the validation loss increases. Our methodology can also aid in understanding implicit regularization in more complex models and with other optimization methods.
Authors: Daniel Soudry, Elad Hoffer, Mor Shpigel Nacson, Nathan Srebro

April 2018 **Norm matters: efficient and accurate normalization schemes in deep networks**

NIPS 2018
Spotlight

Over the past few years batch-normalization has been commonly used in deep networks, allowing faster training and high performance for a wide variety of applications. However, the reasons behind its merits remained unanswered, with several shortcomings that hindered its use for certain tasks. In this work we present a novel view on the purpose and function of normalization methods and weight-decay, as tools to decouple weights' norm from the underlying optimized objective. We also improve the use of weight-normalization and

show the connection between practices such as normalization, weight decay and learning-rate adjustments.
Authors: Elad Hoffer, Ron Banner, Itay Golan, Daniel Soudry

Dec. 2018 Scalable Methods for 8-bit Training of Neural Networks

NIPS 2018

Quantized Neural Networks (QNNs) are often used to improve network efficiency during the inference phase, i.e. after the network has been trained. Extensive research in the field suggests many different quantization schemes. Still, the number of bits required, as well as the best quantization scheme, are yet unknown. Our theoretical analysis suggests that most of the training process is robust to substantial precision reduction, and points to only a few specific operations that require higher precision. Armed with this knowledge, we quantize the model parameters, activations and layer gradients to 8-bit, leaving at a higher precision only the final step in the computation of the weight gradients.
Authors: Ron Banner, Itay Hubara, Elad Hoffer, Daniel Soudry

PUBLICATIONS - WORKSHOPS

May 2015 Deep metric learning using Triplet network

ICLR 2015 Workshop

In this paper we propose the deep triplet network model, which aims to learn useful representations by distance comparisons. We demonstrate using various datasets that our model learns a better representation than that of its immediate competitor, the Siamese network, and discuss future possible usage as a framework for unsupervised learning.
Authors: Elad Hoffer, Nir Ailon

Dec. 2016 Deep unsupervised learning through spatial contrasting

NIPS 2016 Workshop

In this work we present a novel approach for unsupervised training of Convolutional networks that is based on contrasting between spatial regions within images. This criterion can be employed within conventional neural networks and trained using standard techniques such as SGD and back-propagation, thus complementing supervised methods.
Authors: Elad Hoffer, Itay Hubara, Nir Ailon

May 2017 Semi-supervised deep learning by metric embedding

ICLR 2017 Workshop

In this work we explored a new training objective for deep networks that is targeting a semi-supervised regime with only a small subset of labeled data. This criterion is based on a deep metric embedding over distance relations within the set of labeled samples, together with constraints over the embeddings of the unlabeled set.
Authors: Elad Hoffer, Nir Ailon

April 2018 Exponentially vanishing sub-optimal local minima in multilayer neural networks

ICLR 2018 Workshop

Statistical mechanics results suggest that local minima with high error are exponentially rare in high dimensions. However, to prove low error guarantees for Multilayer Neural Networks (MNNs), previous works so far required either a heavily modified MNN model or training method, strong assumptions on the labels (e.g., ”near” linear separability), or an unrealistic hidden layer with (N) units. We examine a MNN with one hidden layer of piecewise linear units, a single output, and a quadratic loss. We prove that, with high probability in the limit of N datapoints, the volume of differentiable regions of the empiric loss containing sub-optimal differentiable local minima is exponentially vanishing in comparison with the same volume of global minima.
Authors: Daniel Soudry, Elad Hoffer

Dec. 2018 **Infer2Train: leveraging inference for better training of deep networks**

NIPS 2018 Workshop

Training large scale Deep Neural Networks (DNNs) requires ever growing computational resources. This growth is usually based on larger and faster training devices. However, a new category of inference-only accelerators is emerging, allowing fast and energy efficient forward pass using low precision operations. In this study, we explore how to leverage such inference-only accelerators for improving training performance. We examine several alternatives and show preliminary results with improved test accuracy on visual-classification tasks such as training ResNet model on the ImageNet and Cifar datasets.
Authors: Elad Hoffer, Berry Weinstein, Itay Hubara, Sergei Gofman, Daniel Soudry

PUBLICATIONS - PREPRINT

Feb. 2018 **On the Blindspots of Convolutional Networks**

Arxiv

In this work, we will demonstrate that convolutional networks have limitations that may, in some cases, hinder it from learning properties of the data, which are easily recognizable by traditional, less demanding, models. To this end, we present a series of competitive analysis studies on image recognition and text analysis tasks, for which convolutional networks are known to provide state-of-the-art results. In our studies, we inject a truth-revealing signal, indiscernible for the network, thus hitting time and again the network's blind spots. The various forms of the carefully designed signals shed a light on the strengths and weaknesses of convolutional network, which may provide insights for both theoreticians that study the power of deep architectures, and for practitioners that consider applying convolutional networks to the task at hand.
Authors: Elad Hoffer, Shai Fine, Daniel Soudry

April 2018 **Bayesian Gradient Descent: Online Variational Bayes Learning with Increased Robustness to Catastrophic Forgetting and Weight Pruning**

Arxiv

We suggest a novel approach for the estimation of the posterior distribution of the weights of a neural network, using an online version of the variational Bayes method. Having a confidence measure of the weights allows to combat several shortcomings of neural networks, such as their parameter redundancy, and their notorious vulnerability to the change of input distribution ("catastrophic forgetting"). Specifically, We show that this approach helps alleviate the catastrophic forgetting phenomenon - even without the knowledge of when the tasks are been switched. Furthermore, it improves the robustness of the network to weight pruning - even without re-training.
Authors: Chen Zeno, Itay Golan, Elad Hoffer, Daniel Soudry

COMPUTER SKILLS

Programming Python, C++, Matlab, Lua, Perl, CUDA, Erlang, Verilog, Julia

Environments Linux, Microsoft Windows, Microsoft Office

Other PyTorch/Torch, TensorFlow, Open-CV, LaTeX

OTHER INFORMATION

Awards 2010-2013 · Dean's honor list - Technion Electrical Engineering Dept.

2018 · The Porat Award for PhD students

2018 · KLA-TENCOR Award for excellent PhD paper

Invited Talks

March 2018 · IMVC 2018 - "Rethinking Common Practices in Deep Learning"

Sep. 2016 · Cx 2nd Annual Fall Workshop (NYC)

Languages

Hebrew · Native

English · Fluent

November 15, 2018