The Journal of Open Source Software

Delve: Neural Network Layer Saturation Computation

Justin Shenk[*1,2], **Mats L. Richter**[†2], **and Wolf Byttner**[3]

1 VisioLab, Berlin, Germany **2** Institute of Cognitive Science, University of Osnabrueck, Osnabrueck, Germany **3** Rapid Health, London, England, United Kingdom

DOI: DOIunavailable

Software

- Review
- Repository
- Archive

Editor: Pending Editor

Reviewers:

- @Pending Reviewers

Submitted: N/A
Published: N/A

Summary

Designing neural networks is a complex task.

Several tools exist which allow analayzing neural networks after and during training. Tools such as … allow …

[Limitation of these methods]

Delve is a Python package for statistical analysis of neural network layer eigenspaces. Delve hooks into PyTorch (Paszke et al., 2019) models and allows saving statistics via TensorBoard (Abadi et al., 2015) events or CSV writers. A comprehensive source of documentation is provided on the home page (http://delve-docs.readthedocs.io).

Statement of Need

Research on changes in neural network representations has exploded in the past years. [add citations] Furthermore, researchers who are interested in developing novel algorithms must implement from scratch much of the computational and algorithmic infrastructure for analysis and visualization. By packaging a library that is particularly useful for extracting statistics from neuarl network training, future researchers can benefit from access to a high-level interface and clearly documented methods for their work.

Overview of the Library

The software is structured into several modules which distribute tasks. Full details are available at https://delve-docs.readthedocs.io/. The … module provides …

Subclassing the TensorBoardX `SummaryWriter` (Abadi et al., 2015)…

Eigendecomposition of the feature covariance matrix

Saturation is a measure of the rank of the layer feature eigenspace (Shenk, 2018; Shenk et al., 2020) Covariance matrix of features is computed as described in (Shenk et al., 2020)…

$$Q(Z_l, Z_l) = \frac{\sum_{b=0}^{B} A_{l,b}^T A_{l,b}}{n} - (\bar{A}_l \bigotimes \bar{A}_l)$$

for B batches of layer output matrix A_l and n number of samples.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia,

[*] co-first author
[†] co-first author

Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*. https://www.tensorflow.org/

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32* (pp. 8024–8035). Curran Associates, Inc. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Shenk, J. (2018). *Spectral Decomposition for Live Guidance of Neural Network Architecture Design* [Master's Thesis]. University of Osnabrück.

Shenk, J., Richter, M. L., Byttner, W., Arpteg, A., & Huss, M. (2020). Feature space saturation during training. *CoRR*, *abs/2006.08679*. https://arxiv.org/abs/2006.08679