TOWARDS GUIDED NEURAL ARCHITECTURE DESIGN BY SPECTRAL ANALYSIS

A PREPRINT

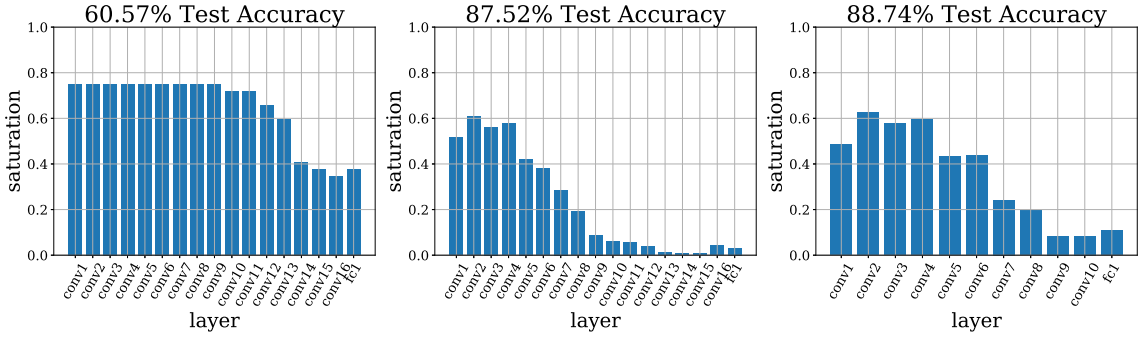
Mats L. Richter
Institute of Cognitive Science
Osnabrück University
Osnabrück, Germany
matrichter@uos.de

Justin Shenk
Peltarion
Stockholm, Sweden 111 60
justin@peltarion.com

August 30, 2019

ABSTRACT

We propose a metric, Layer Saturation, defined as the proportion of the number of eigenvalues needed to explain 99% of the variance of the latent representations, for analyzing the learned representations of neural network layers. [1] Saturation is based on spectral analysis (analog to SVCCA [2]) and can be computed efficiently, making live analysis of the representations practical during training. We provide an outlook on the application of this metric in neural architecture design and show how saturation may help reducing the number of experiments required to find a good architecture. We further show that saturation is related to the generalization and predictive performance of neural networks.



(a) not enough filters (b) too deep (c) just right

Figure 1: Layer-wise Saturation on three different VGG-style networks trained on CIFAR10.

References

[1] Justin Shenk, Mats L. Richter, Anders Arpteg, and Mikael Huss. Spectral analysis of latent representations. *CoRR*, abs/1907.08589, 2019.

[2] Maithra Raghu, Justin Gilmer, Jason Yosinski, and Jascha Sohl-Dickstein. Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6076–6085. Curran Associates, Inc., 2017.