







Visual Analysis of the Impact of Neural Network Hyper-Parameters

Supplementary material

Daniel Jönsson¹ , Gabriel Eilertsen¹ , Hezi Shi² , Jianmin Zheng² , Anders Ynnerman¹ , and Jonas Unger¹ 

¹Linköping University, Institute for Science and Technology, Sweden

²Nanyang Technological University, Institute for Media Innovation, Singapore

1. Introduction

The following supplementary material demonstrate a visualization of the hyper-parameters of the studied 13K CNN using parallel coordinates, and motivates why this technique was discarded during the design process.

2. Parallel coordinate visualization of hyper-parameters

Figure 1 shows a parallel coordinate plot of 6 different hyper-parameters of the 13K studied CNNs, as well as normalized test accuracy on the first axis (same normalization as in the main paper). The vertical locations of the categorical hyper-parameters have been randomly offset by a small amount, in order to better convey the individual samples. The color coding distinguishes between unsuccessful and successful trainings. This makes it possible to detect advantageous combinations of hyper-parameters. For example, it is evident how the *constant* and *random normal* initialization schemes correlate with low accuracy, especially together with *momentum SGD*. It is also possible to see more subtle patterns by investigating how the colors mix between the axes.

Figures 2 and 3 shows similar information as Figure 1, but displayed by selecting a subset of low performance and high performance samples, respectively. As can be seen, some of the connections are missing, for example between *random normal* initialization and *momentum SGD* in Figure 3.

While the more impactful hyper-parameters can be characterized to some extent, it is difficult to see any patterns in other hyper-parameters, such as filter size and learning rate. It is also in general difficult to see complicated dependencies between combinations of multiple hyper-parameters. Such dependencies are better revealed with the tools presented in the main paper.

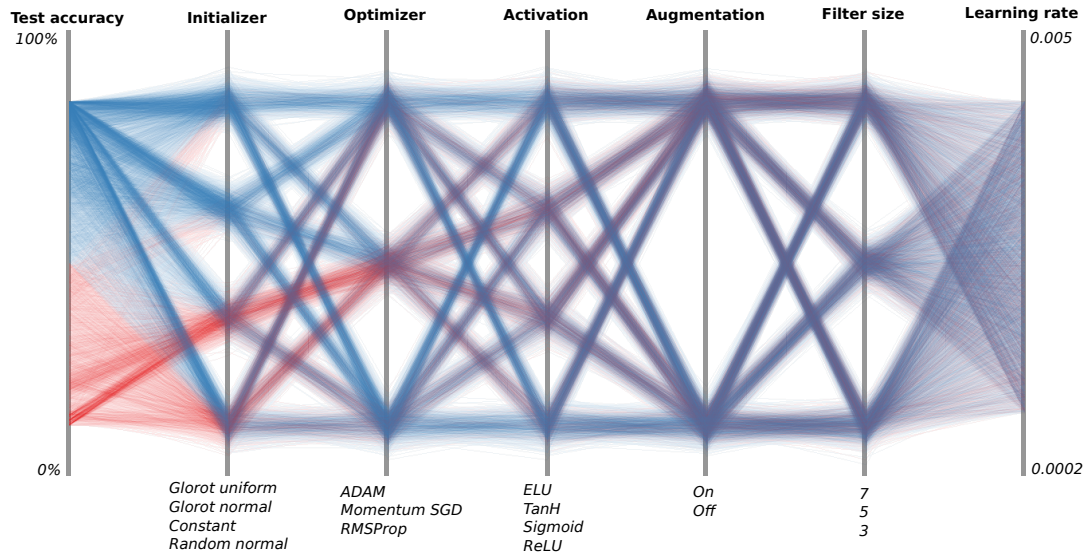


Figure 1

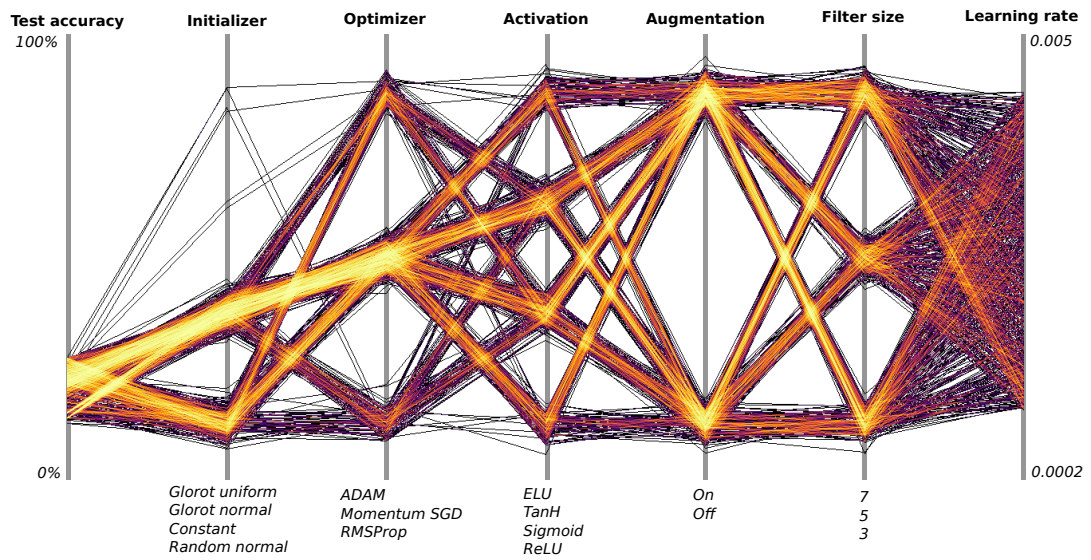


Figure 2

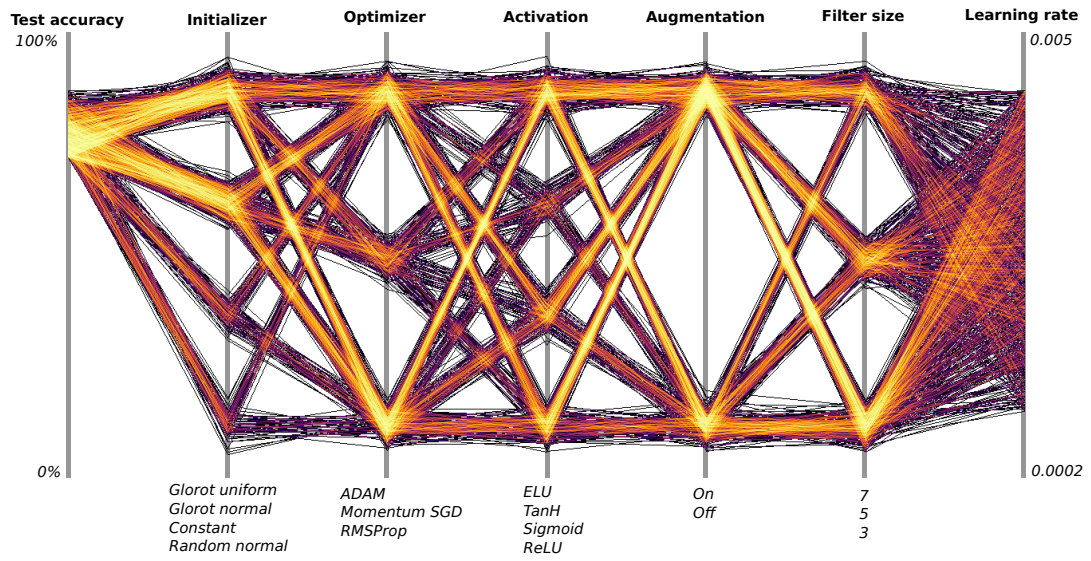


Figure 3