Krizhevsky et al. Analysis

Palani Johnson
A02231136

October 3, 2021

In the paper *ImageNet Classification with Deep Convolutional Neural Networks*, Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton use a deep convolutional neural net, trained on the ImageNet LSVRC-2010 dataset, to classify images into 1000 different classes. This dataset contains over 15 million labeled high-resolution images placed into roughly 22,000 categories. The images in this data set were not of fixed size, so during training the images were down-sampled to a fixed resolution of $256x256$ and the raw RGB values were passed into the net. The net itself consisted of 8 hidden layers (5 convolutional, 3 fully-connected). The net was trained on two GTX 580 GPUs over a period of six days.

Remarks

I feel that I can (mostly) trust the results of this paper, but I would have liked to know more about the researchers' process for choosing the architecture a little bit more. While they were good ad showing some of their process, most of their decisions for the architecture felt arbitrary. It appears that they did do a lot of testing to be able to get to the the final network that this paper covers, but the results were generally glossed over. The only graphs or figures that they provided that gave any insight into this process were Figure 1 and Table 2, and Figure 1 looks hand drawn! While I think the results are certainly impressive, a $> 10\%$ increase in accuracy is very significant, I don't think the paper has much generality. Most of what this paper says is "This network works for my dataset!" but it doesn't help to explain why it works.

This being said, the parts that did have some kind of process or comparison were the most interesting aspects of the paper to me. I was surprised by the comparison of ReLUs to tanh neurons. I would have thought that the ReLUs would be purely faster from a clock cycle point of view by having to perform fewer individual instructions than tanh, resulting in a performance increase. The researchers instead observed a faster training rate and a decrease in the number of epochs needed to train to a threshold. Of course, one test is not enough to convince me that this will always work. Most the techniques used in this paper seem to be useful tools to keep around for building neural nets, but I don't have much faith that they will consistently work.