

Analysis: ImageNet Classification with Deep Convolutional Neural Networks

This paper demonstrates a result on the ImageNet LSVRC-2012 competition, where the authors trained the model on the ImageNet dataset, which contains 1000 categories of about 1.2 million images.

Data Set:

The ImageNet images are of variable resolutions, and the model presented in this paper requires fixed size images, each image was scaled to 256x256 pixels. The network is trained on the raw RGB values of the pixels.

As per my understanding, the network works with 224x224 images, which are generated by randomly sampling patches of that size from each 256x256 image. They did the data augmentation by performing PCA on a set of RGB pixel values.

Model:

Their model had 5 convolutional layers and 3 fully connected layers.

Experiment:

Authors used the ReLU function as opposed to the sigmoid or tanh, yields better gains in training speed, because of the non-saturating nature of the function. When compared against the tanh, they were able to train their model 6 times faster with the ReLU activation.

Dropout was used in the first two fully connected layers. Momentum to speed up training by pushing batch gradient descent updates to move towards the optimum faster. Authors used a momentum parameter of 0.9, which roughly keeps the past 10 values in the moving average memory. Weight decay was used to scale down weights by a factor slightly less than 1 to prevent them from becoming too big. The training is done on two GPUs for parallelism, and the setup was quite interesting. The GPUs used each have 3GB memory. While layer 4 (CONV) operates on input which comes from the layer 3 activations from *both* GPUs, other conv layers in the network do not have this cross GPU communication going on, and only work with activations from the GPU local half of the network.

Likes:

1. This work was the first of its kind to have trained deep convolutional networks on GPUs to achieve impressive results on the ImageNet dataset for object detection.
2. I like the way they lay out their research, they selected the components that would enhance their results. They gave an efficient, parallelized GPU implementation of their model, their model helps in improving performance and reducing training time, along with extensive ablative studies. This gives me a feeling that they knew what they had to do.

Dislikes:

1. Although, most of the decisions made in the paper were based on experimental results and there was not a lot of theory to back it.

Their results were very impressive, I also read that it was after their results that big companies like Google and Microsoft switched to use DNN and as a researcher their results would have been promising to me too.

These papers are always insightful, and they always inspire me. Despite of all the technological advancements there are still new fields that researchers can investigate. These papers show that there would be no saturation in the industry.