ImageNet Classification with Deep Convolutional Neural Networks

Summary:

Network Architecture used:

- Contains eight learned layers **Five convolutional** and **three fully-connected** layers the network's size is **limited** mainly by the amount of **memory available** on current GPUs and by the amount of training time. There's high scope of improvement when better GPUs come into picture.
- The output of the last fully-connected layer is fed to a 1000-way **softmax** which produces a distribution over the **1000 class labels.**

Training the model:

- **Data Augmentation** by generating **image translations** and horizontal **reflections**, performing **PCA** on RGB channels, also a **dropout of 0.5** was used to prevent overfitting.
- **SGD optimizer**, batch size **128**, momentum of **0.9**, and weight decay of **0.0005**
- **Initialization** - For weights **Zero-mean Gaussian distribution** with standard deviation **0.01**. For biases in **second, fourth, and fifth** convolutional layers, as well as in the fully-connected hidden layers, with the **constant 1** and in the remaining layers with the **constant 0.**
- **Learning rate:** Initialized at **0.01** and reduced **three** times prior to termination.
- **ImageNet** - It is a dataset of over **15 million labeled** high-resolution images belonging to roughly **22,000 categories.**
- **Image pre-processing** - Rescaled the rectangular image such that the shorter side was of length 256, and cropped out the central **256×256** patch from the resulting image.

Key observations:

- A four-layer convolutional neural network with ReLUs reaches a **25%** training error rate on CIFAR-10 **six times faster** than an equivalent network with tanh neurons
- **Faster learning** has great influence on the performance of models trained on large datasets.
- For training models with **overlapping pooling** slightly reduces the **overfit**.
- ReLUs do not require **input normalization** to prevent them from saturating.
- Depth of CNN plays a cruicial role in the model's performance.
- No unsupervised pre-training was done.

Special advantages in CNN:

- Their capacity can be controlled by varying their depth and breadth, and they also make strong assumptions about the nature of images.
- Compared to standard feedforward neural networks with similarly-sized layers, CNNs have much **fewer connections and parameters** and so they are easier to train, while their theoretically-best performance is likely to be only slightly worse.
- For intensively large CNNs, **GPUs** are necessary to facilitate the training.

Result: The CNN described in this paper achieved a **top-5** error rate of **18.2%**. Averaging the predictions of five similar CNNs gives an error rate of **16.4%.**