ROSHAN POUDEL

CS 415: Deep Learning
HW 5/6

Summary of the paper **"ImageNet Classification with Deep Convolutional Neural Networks" by Krizhevsky, Sutskever, and Hinton**

“ImageNet Classification with Deep Convolutional Neural Networks” is a paper that introduced AlexNet, a deep convolutional neural network that was trained on the ImageNet dataset and achieved state-of-the-art performance on object recognition tasks.

The authors trained a large convolutional neural network with over 60 million parameters on a dataset of 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest. The network was able to classify images into one of 1000 different categories with an error rate of 37.5% for top-1 and 17% for top-5 which was a significant improvement over previous state-of-the-art methods.

The architecture of AlexNet consists of five convolutional layers (including max-pooling layers) followed by three fully connected layers. All convolutional layers use rectified linear units (ReLU) as activation functions. All fully connected layers except the last one use ReLU activation functions as well. The last fully connected layer has 1000 neurons, one for each class in the ImageNet dataset. The output of this layer is fed into a softmax function to produce class probabilities.

The learning algorithm used in AlexNet is stochastic gradient descent (SGD). The authors also applied a variation of SGD called "momentum," which helped accelerate convergence during training. The researchers used two GPUs to train their convolutional neural network (CNN) due to the limited memory (3GB) of a single GTX 580 GPU. The GPUs were able to communicate directly with each other's memory, without going through the host machine's memory, which made them well-suited for cross-GPU parallelization. The parallelization scheme they used involved distributing half of the kernels (or neurons) of the CNN on each GPU. However, there was a specific communication pattern between the GPUs in certain layers of the network. For example, the kernels in layer 3 took input from all kernel maps in layer 2, while the kernels in layer 4 took input only from those kernel maps in layer 3 that resided on the same GPU. This allowed them to control the amount of communication between the GPUs and tune it to an acceptable fraction of the computation.

The paper also introduced new types of layers that are commonly used in modern CNN architectures. For instance, the authors used the concept of local response normalization (LRN) layers, which are used to normalize the outputs of neurons within a local region of the input. The paper discusses various regularization techniques used to reduce overfitting, including weight decay (also known as L2 regularization) and dropout. Weight decay involves adding a penalty term to the loss function to discourage large weight values, while dropout helps prevent overfitting by randomly dropping out neurons during training, forcing the network to learn more robust features.

In summary, the main ideas presented in the paper include the introduction of the AlexNet architecture, which is a type of CNN specifically designed for image classification tasks, the use of SGD with momentum as the learning algorithm, the efficient use of GPUs for training, the introduction of new types of layers, and the application of regularization techniques to reduce overfitting.