

Article reviewed: ImageNet Classification with Deep Convolutional Neural Networks – Meetu Malhotra

Reference:

[https://proceedings.neurips.cc/paper\\_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf)

Summary: The paper talks about AlexNet model, deep CNN model, which was introduced in 2012 by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. It won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The model achieved top-1 error rate of 37.5% and top-5 error rate of 17.0%. It was implemented using 8 layers: 5 convolutional layers followed by 3 fully connected layers. The dataset used was 1.2m training image data, 50K validation data and 150K images of test data. There were 60m parameters used on 650k neurons, with softmax in final layer to classify 1000 classes. To improve the learning from training data, the authors introduced the concept of non-saturating (change in input value, leads to change in output instead of saturated value) functions ReLu and use of GPUs. Additionally, to avoid overfitting they implemented dropout neurons and data augmentation, meaning created transformed image data using Python code. It also used Local Response normalization and overlapping max pooling techniques to avoid overfitting.

Five C's:

Category: The paper falls under the category of Computer vision and Deep Learning

Context: The paper focuses on reducing error rate for ImageNet dataset classification by using different techniques such as data augmentation, dropout regularization, Local Response normalization and overlapping max pooling and GPUs, all implemented on 8-layer architecture.

Correctness: The paper presents accurate information on developing CNN framework as shown in the experiments section, covering comparison on top-1 and top-5 results

Contributions: This paper significantly changed the domain of deep learning, especially in computer vision and object detection, by training 8 layer neural networks and addressed the overfitting problem by using data augmentation, dropout regularization, LRN and overlap max pooling. The paper focuses on improving error rate, with significant improvement top-1 error rate of 37.5% and top-5 error rate of 17.0% on 150K test image dataset.

Clarity: The paper is written in a clear and understandable manner. It starts with the introduction of CNN and then results of top-1 error rate of 37.5% and top-5 error rate of 17.0% on imageNet test set, eliminating the overfitting problem. It shows clear architectural design through 8 convolutional layers, with the first convolutional layer filters the  $224 \times 224 \times 3$  input image with 96 kernels, 2<sup>nd</sup> layer of 256 kernels, 3<sup>rd</sup> layer, and 4<sup>th</sup> layer of 384 kernels and 5<sup>th</sup> of 256 kernels, followed by fully connected layers of 4096 neurons each. Further it explains non-linearity function of ReLu and the drop out regularization. Additionally, it shows experiments conducted on different models with comparisons on val and test dataset for top-1 and top-5 tiers.

Outline: The paper follows a logical structure, starting with an introduction to the issue of overfitting and the results. It then talks about its architecture, concept of data augmentation and dropout regularization, and Local Response normalization and the overlapping max pooling, explaining with architectural diagram in figure 2, section 3.5. The paper also discusses the analysis of implementation on different models and on top1, top 5 tiers for val and test datasets. Furthermore, it emphasizes good generalization of performance on object recognition tasks. It further covers the limitations in last section of discussion as pointed below.

Discussion:

Innovations: It brought the innovative idea of fast and better training using ReLu activation function and use of GPUs. The network was trained using two NVIDIA GTX 580 GPUs, which significantly sped up the training process. This made it feasible to train such a deep network on a large dataset like ImageNet. Another innovation was the data augmentation and dropout regularization for neurons to avoid overfitting. Furthermore, it also used Local Response normalization, involving normalizing output of neurons based on the neighboring neurons and the overlapping max pooling technique as opposed to normal pooling to avoid overfitting.

Assumptions: The paper assumes that the readers have fundamental understanding of neural network concept such as shallow and deep neural networks, ImageNet dataset

Faults: As discussed in section 7 of the paper, the limitation of paper includes not working on unsupervised pre-training dataset. Also, use very large and deep convolutional nets on video sequences is missing.

Terminology: Deep neural networks, Deep learning CNNs, overlapping max pooling, dropout, ReLu, softmax, activation functions, data augmentation, ImageNet classification, object recognition