

## TERM PAPER

We are living in a special time in Machine Learning(ML), Neural Network(NN), and Deep Learning(DL) history. Never in the history of Machine Learning and Neural Networks have the available tools at our disposal been so exceptional. Deep Learning has been around since the 1940s, going by different names and incarnations based on various schools of thought and popular research trends at a given time. At the very core, Deep Learning belongs to the family of Artificial Neural Networks (ANNs), a set of algorithms that learn patterns inspired by the structure and function of the brain.

Though CNN's have become quite popular for tasks like image classification and object detection in recent times, the concept/idea was introduced much earlier in 1980's with Neocognitron followed by LeNet in 1998 for character and digit recognition. CNN's typically have conv layers followed by RELU and max pool layer followed by FC layers. However, these networks were not good enough for realistic, large category, image classification problems and Alexnet that came in 2012 was able to demonstrate good results on these. Concepts such as Local contrast normalization and data augmentation were introduced in Alexnet. Apart from these, it was much like LeNet with more number of layers, channels and parameters. Then came the VGGNet with idea to use smaller filters and short stride and achieved higher accuracies than AlexNet. Dropout was introduced in FC layers and intuition of dropout was that ensemble of networks gives higher accuracy than a single network and training a network with Dropout applied is similar to training an ensemble of networks. Also, used the concept of FC to conv layers so that network can be applied to larger image than the size it is trained on. Then came GoogLeNet with the concept of inception modules that considered 1x1 convolutions and showed that higher accuracy than VGGNet can be obtained with more layers and much less no of parameters. Then came ResNet with the concept of forward connections and showed that much deeper networks can be trained using this, achieving higher accuracies. Batch normalization was included in ResNet whose idea was to keep the distribution of data coming into the layer same. Then came ResNext that includes key concepts from ResNet and GoogLeNet followed by DenseNet that takes the concept of ResNet to its extreme. Much success in training these networks is due to algorithms available like SGD, SGD with momentum, Adam that lets us train networks on batches. We also have many other optimization algorithms like Nesterov momentum, RMSProp etc that guides us in the right direction, leading to convergence much faster.

Deep Learning also showed significant advances in detection tasks where R-CNN uses the classification networks mentioned above on the region proposals generated externally(say selective search) and uses regression techniques to adjust the bounding boxes. Then came fast R-CNN, reducing the repeated computations followed by Faster R-CNN that uses RPN to generate proposals internally. Then came YOLO that attempts to do away with the concept of proposals and comes up with 7x7x30 tensor that has all the information. Then came SSD that takes features from earlier layers so objects of various sizes can be detected. All these techniques/algorithms were able to show that neural networks were able to solve task of object detection with much higher accuracies than before at much faster rate(speed). Deep Learning also showed significant advances in segmentation tasks with FCN's, Conv/Deconv networks and Mask R-CNN all using the key ideas of CNN's.

Deep Learning is an exciting field, and due to these powerful GPUs, modular libraries, and unbridled, intelligent researchers, we're seeing new publications that push the state-of-the-art coming on a monthly basis. From a software perspective, we have libraries and packages such as Tensorflow, Keras etc enabling us to rapidly construct new architectures in a fraction of the time it took us just years before. Then, from a hardware view, memory and general purpose GPUs are becoming increasingly cheaper while still becoming more powerful and are able to meet the needs of DL which were not possible earlier.