A Survey on Architecture of Deep Convolution Neural Network

Prachi Pandya[1] Pinal J. Patel[2]
[1]PG Student [2]Assistant Professor
[1,2]Department of Computer Engineering
[1,2]Government Engineering College Gandhinagar, Gujarat, India

Abstract— With the advancement of huge information age, Convolutional neural systems (CNNs) with increasingly shrouded layers have progressively complex system structure and all the more dominant element learning and highlight articulation capacities than customary AI strategies. The convolution neural system model prepared by the profound learning calculation has made amazing accomplishments in some vast scale ID errands in the field of PC vision since its introduction. The ground-breaking learning capacity of profound CNN is to a great extent accomplished with the utilization of numerous component extraction organizes that can consequently take in progressive portrayals from the information. Accessibility of a lot of information and enhancements in the equipment handling units have quickened the examination in CNNs and as of late fascinating profound CNN designs are reported. This paper right off the bat talk about the ascent and development of profound learning and Convolutional Neural Systems, essential model structure and activities and different profound CNN designs in a word afterwards. Lastly we will have the similar perspective on CNN structures and further extent of advancement.

Keywords: CNN, Neural Network, Convolution Layer

I. INTRODUCTION

Computer vision (CV) is an investigation of how to utilize Computer recreation of human visual science, its primary errand is through the accumulation of pictures (or video) examination and comprehension, to settle on decisions or choices. In the previous couple of decades, CV has gained incredible ground and advancement. The Picture acknowledgment is a sort of innovation that utilizes PC to process, examine and comprehend the picture to recognize the objective and object of various modes. At present, picture acknowledgment innovation has incredible business market and great application prospect in Web applications, for example, picture look, item suggestion, client conduct investigation and face acknowledgment. In the meantime, cutting edge, for example, insightful robot, unmanned driving and unmanned flying vehicle Industry and science, drug and topography and numerous different controls have wide application prospects.

In prior occasions highlight extraction strategies, for example, Scale-invariant feature transform (SIFT [1]) and histogram of oriented gradients (HOG [2]) were utilized, and afterward the separated Element input classifier for grouping and recognition. But with the ascent of profound CNN there have been noticiable rot in th fame of these customary element extraction calculations like SIFT,SURF,HOG etc. Deep CNN learns the element all the more effectively as well as utilized for characterization purposes over many AI techniques. Local just as worldwide features of picture ,audio, video information could be pleasantly learned and incorporated by profound CNN designs which furthered affects the grouping results and in general execution of the model. Deep CNNs are seen performing very well in comparison to conventional element extraction techniques and order techniques. The just thing to remember while working with profound CNN engineering is you need extensive measure of information to prepare the system so as to make it progressively proficient and accurate. For that regardless of whether you have huge dataset finetuning of the system additionally assumes critical job in the execution of profound CNN systems.

II. BASIC MODEL STRUCTURES, OPERATIONS OF DEEP CNN ARCHITECTURES

A typical convolution neural system model structure chart appeared in Fig. 1, comprises of Convolutional layer pursued by pooling/sub examining layer than again conv and pool and than couple of fully connected layers which ultimately gives yield of the offered contribution to the system.

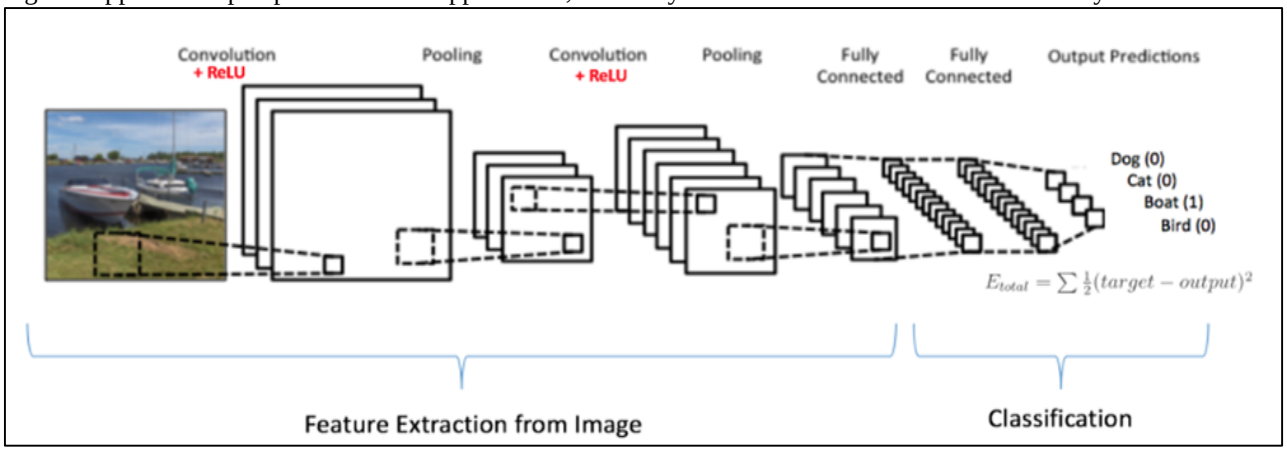

Fig. 1: A typical Convolutional Neural Network Architecture

A. Convolution Layer

The objective of a Conv layer is to extract features of the input volume. some portion of the picture is associated with the following Conv layer in such a case that every one of the pixels of the information is associated with the Conv layer, It will be excessively computationally costly. So we will apply dot products between a receptive field and a filter on all the dimensions. The result of this activity is a single integer of the yield volume (feature map). At that point we slide the filter throughout the following respective field of a similar information picture by a Stride and compute again the dot products between the new open field and a similar filter. We rehash this procedure until we go through the entire input image. The yield will be the contribution for the following layer.[3]

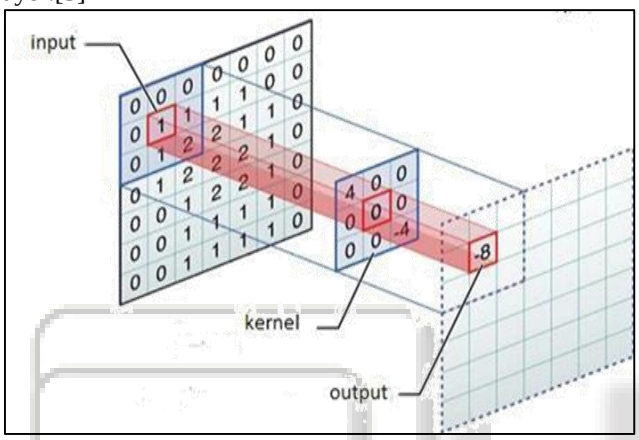

Fig. 2: How Convolution layer works

Here comes the challenge of deciding size of stride, no of filters and activation function which impacts on the output of the convolution layer.

B. Pooling Layer

Pool Layer performs a function to reduce the spatial dimensions of the input, and the computational complexity of our model. And it also controls overfitting. It operates independently on every depth slice of the input. There are different functions such as Max pooling, average pooling, or L2-norm pooling. However, Max pooling is the most used type of pooling which only takes the most important part (the value of the brightest pixel) of the input volume.[3]

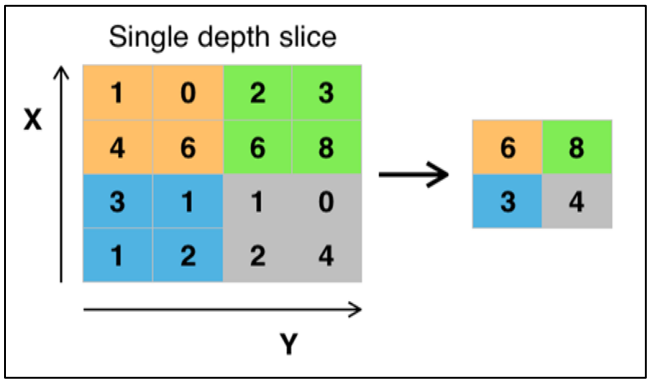

Fig. 3: Max Pooling with 2*2 filter and stride=2

C. Fully Connected Layer

Fully connected layers connect every neuron in one layer to every neuron in another layer. The last fully-connected layer uses a softmax/sigmoid activation function for classifying the generated features of the input image into various classes based on the training dataset.[3]

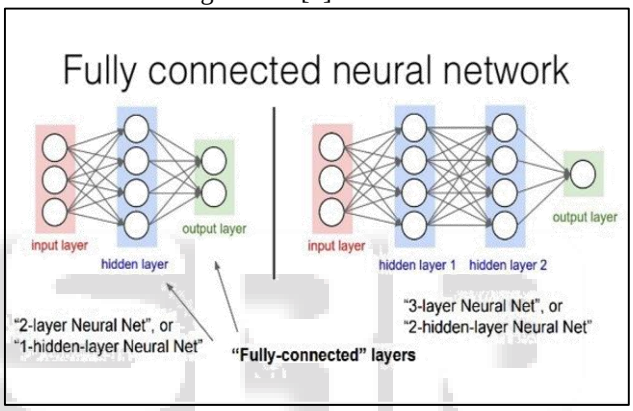

Fig. 4: Fully Connected Layer

III. Various Deep CNN Architectures in Brief

A. LeNet

The first successful applications of Convolutional Networks were developed by Yann LeCun in 1990's. Of these, the best known is the LeNet architecture that was used to read zip codes, digits, etc.[4]

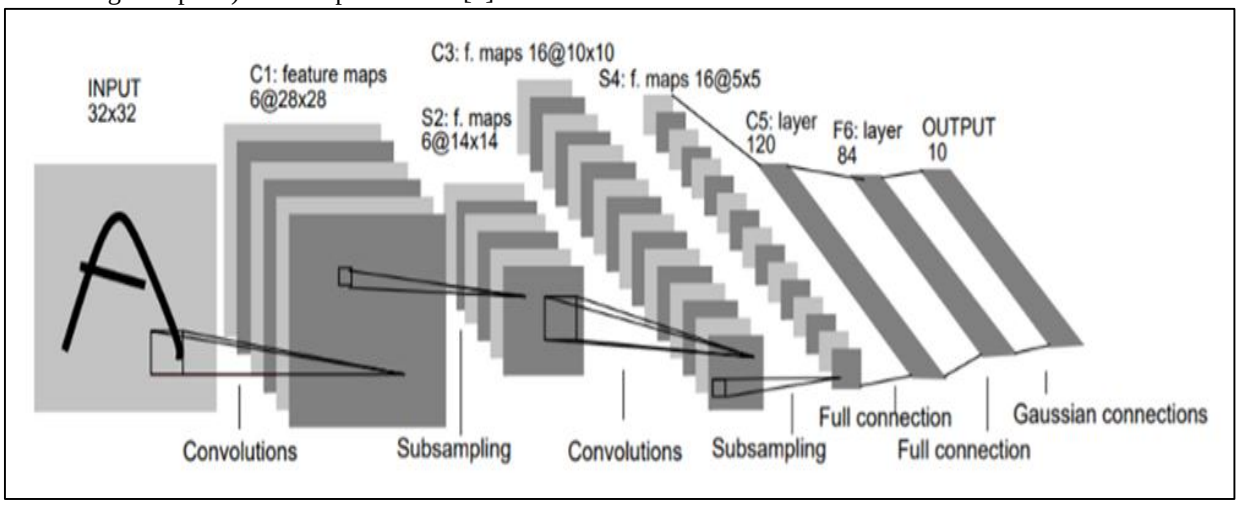

Fig. 5: LeNet Architecture

B. AlexNet

The first work that popularized Convolutional Networks in Computer Vision was the AlexNet, developed by Alex Krizhevsky, Ilya Sutskever and Geoff Hinton. The AlexNet was submitted to the ImageNet ILSVRC challenge in 2012 and altogether outflanked the second runner-up (top 5 error of 16% compared to runner-up with 26% error). The Network had a very similar architecture to LeNet, but was deeper, bigger, and featured Convolutional Layers stacked over one another (previously it was common to only have a single CONV layer always immediately followed by a POOL layer).[5]

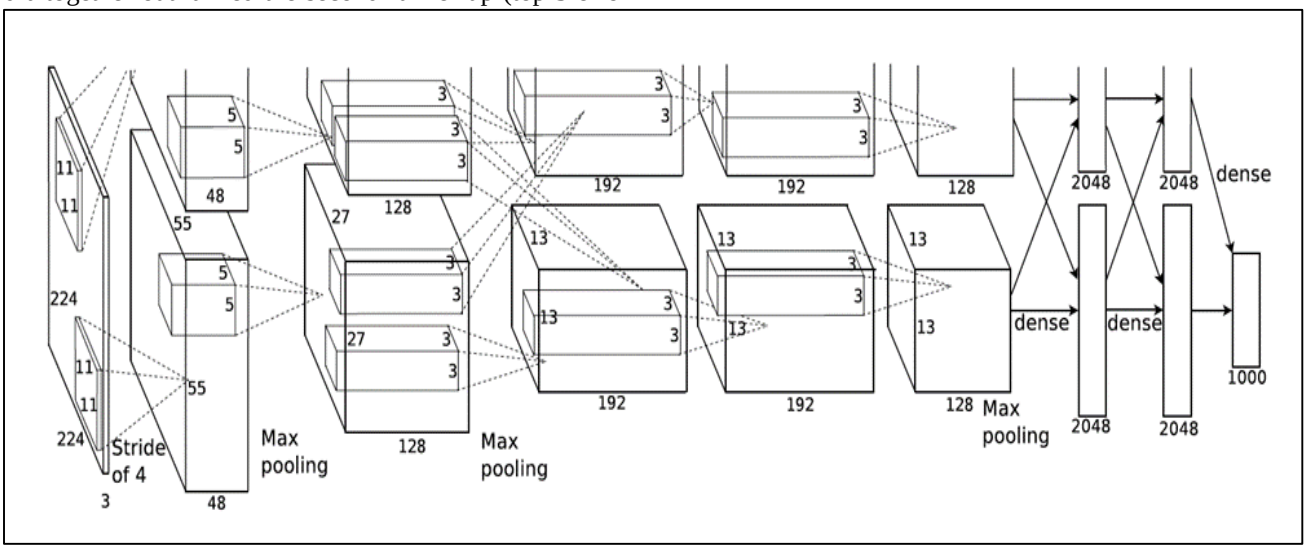

Fig. 6: AlexNet Architecture

C. ZFNet

The ILSVRC 2013 winner was a Convolutional Network from Matthew Zeiler and Rob Fergus. It became known as the ZFNet (short for Zeiler & Fergus Net). It was an improvement on AlexNet by tweaking the architecture hyperparameters, in particular by expanding the size of the middle convolutional layers and making the stride and filter size on the first layer smaller.[6]

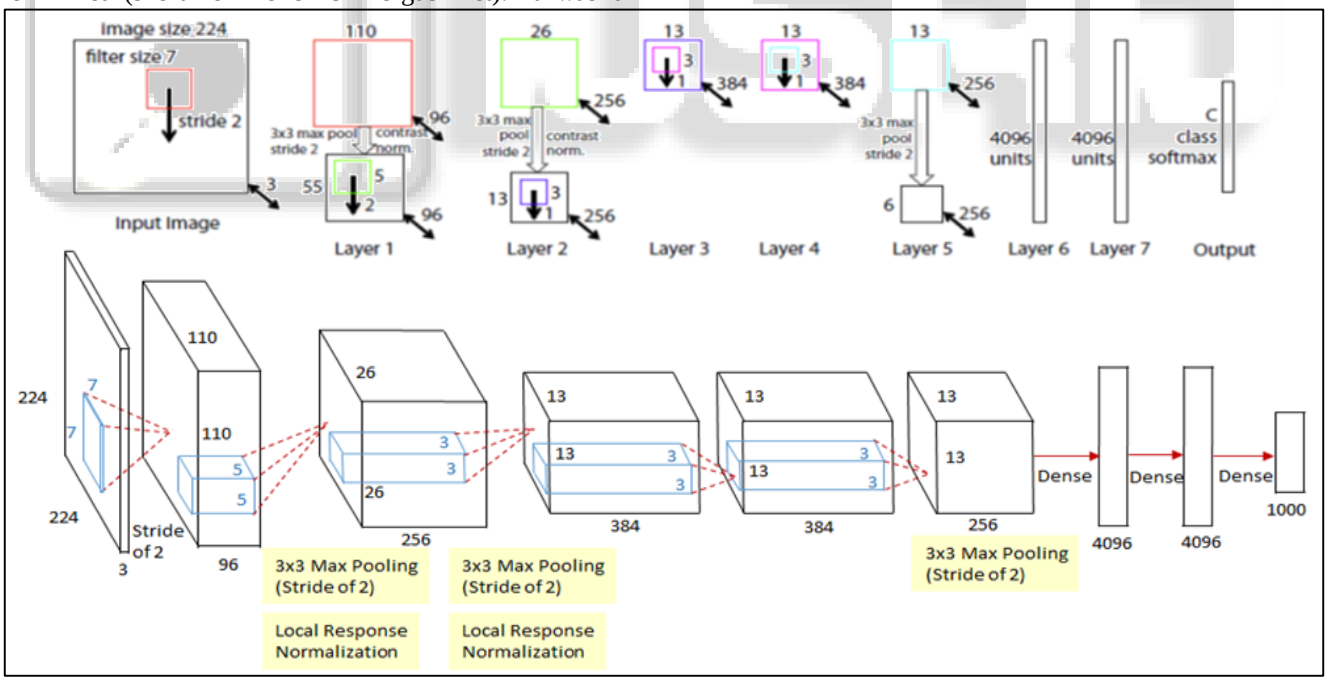

Fig. 7: ZFNet Architecture

D. GoogLeNet

The ILSVRC 2014 winner was a Convolutional Network from Szegedy et al. from Google. Its main contribution was the development of an Inception Module that dramatically diminished the quantity of parameters in the network (4M, compared to AlexNet with 60M). Furthermore, this paper uses Average Pooling instead of Fully Connected layers at the top of the ConvNet, dispensing with a lot of parameters that don't appear to make a difference much. There are also several followup versions to the GoogLeNet, most recently Inception-v4.[7]

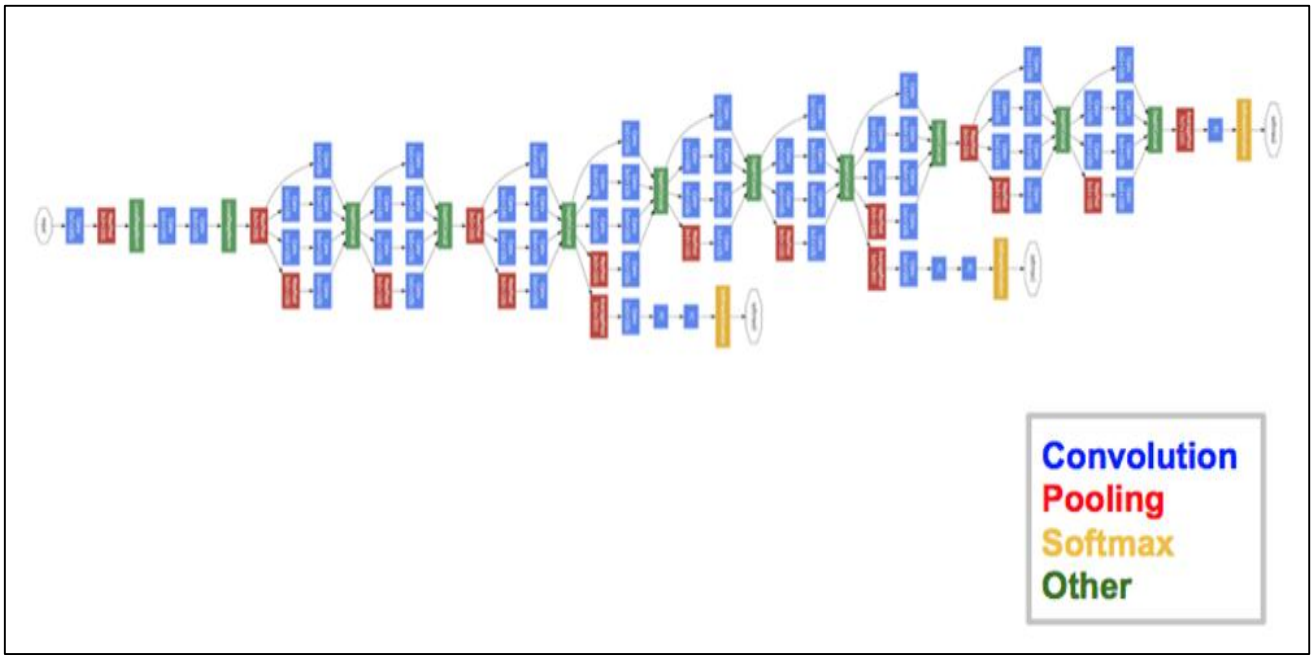

Fig. 8: GoogLeNet Architecture

E. VGGNet

The runner-up in ILSVRC 2014 was the network from Karen Simonyan and Andrew Zisserman that became known as the VGGNet. Its main contribution was in showing that the depth of the network is a critical component for good performance. Their final best network contains 16 CONV/FC layers and, appealingly, features an extremely homogeneous architecture that only performs 3x3 convolutions and 2x2 pooling from the beginning to the end. Their pretrained model is available for plug and play use in Caffe. A downside of the VGGNet is that it is more expensive to evaluate and uses a lot more memory and parameters (140M). Most of these parameters are in the first fully connected layer, and it was since found that these FC layers can be removed with no performance downgrade, essentially diminishing the quantity of vital parameters.[8]

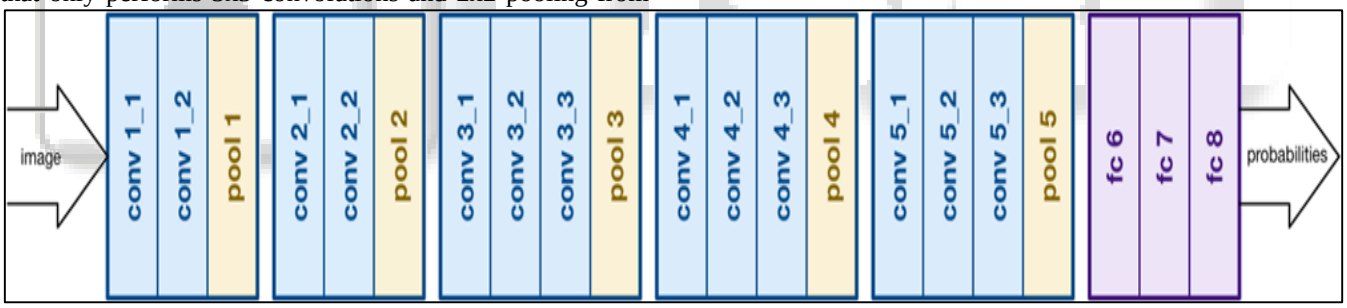

Fig. 9: VGGNet Architecture

F. ResNet

Residual Network developed by Kaiming He et al. was the winner of ILSVRC 2015. It features special skip connections and a heavy use of batch normalization. The architecture is also missing fully connected layers at the end of the network. The reader is also referred to Kaiming's presentation (video, slides), and some recent experiments that reproduce these networks in Torch. ResNets are currently by far best in class Convolutional Neural Network models and are the default choice for using ConvNets in practice (as of May 10, 2016). In particular, also see more recent developments that tweak the original architecture from Kaiming He et al. Identity Mappings in Deep Residual Networks (published March 2016).[9]

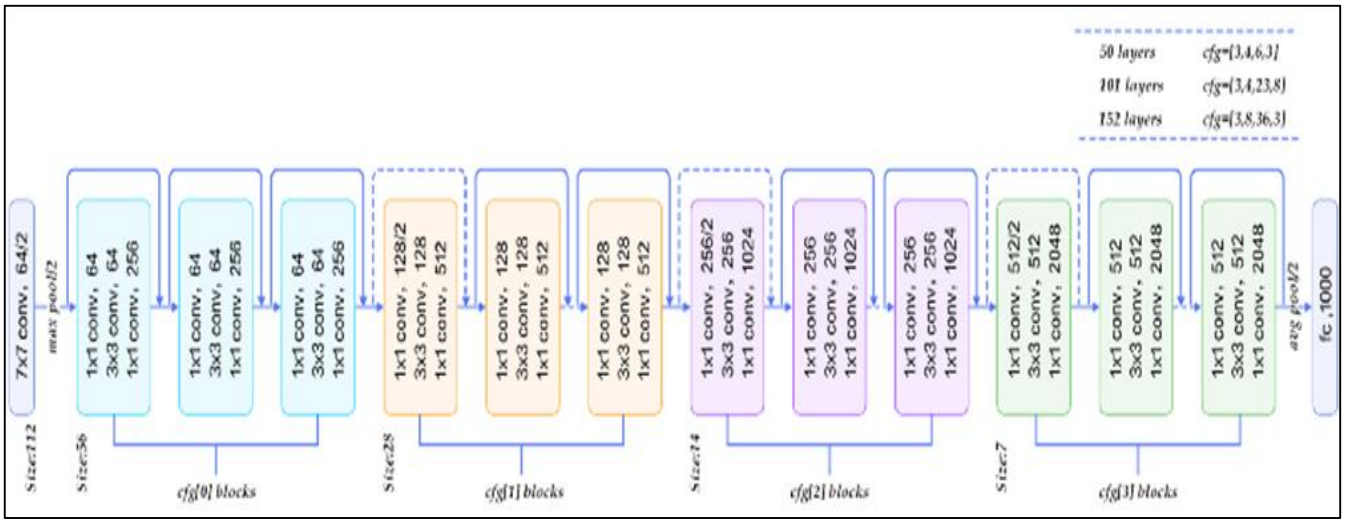

Fig. 10: ResNet Architecture

IV. COMPARISON

Now Here we Present the comparison between the architectures which are discussed above in the form of bar chart, where on x-axis shows the architecture acronym and y-axis shows top 5% error percentage. Along with that we tried to show the layers of each architecture.

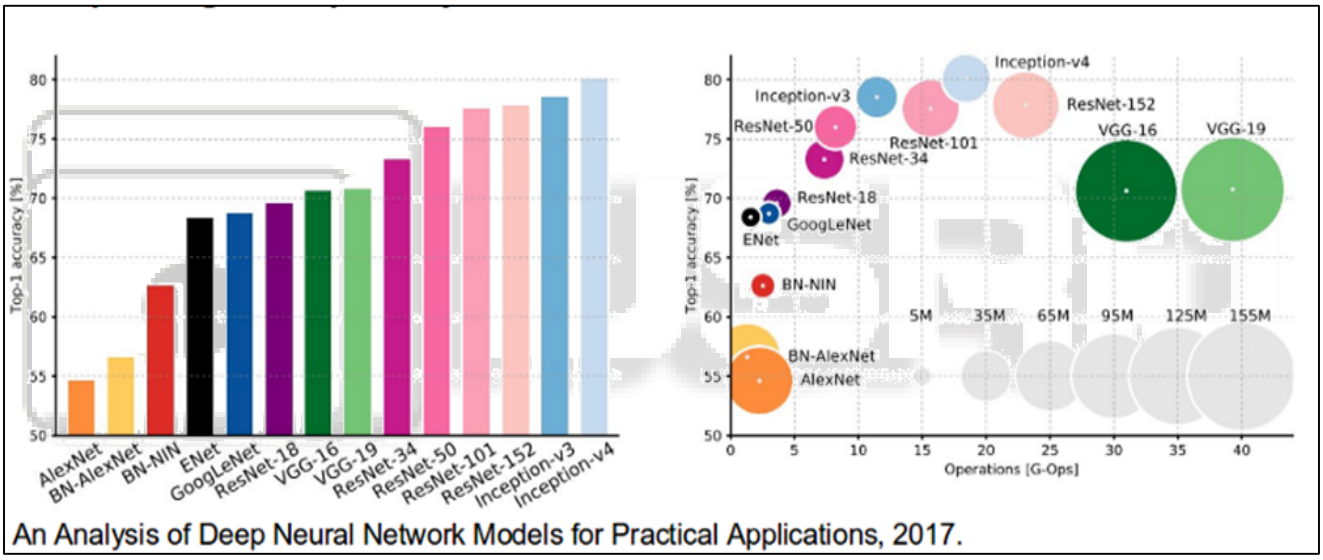

Fig. 11: Accuracy of various Deep CNN Architectures

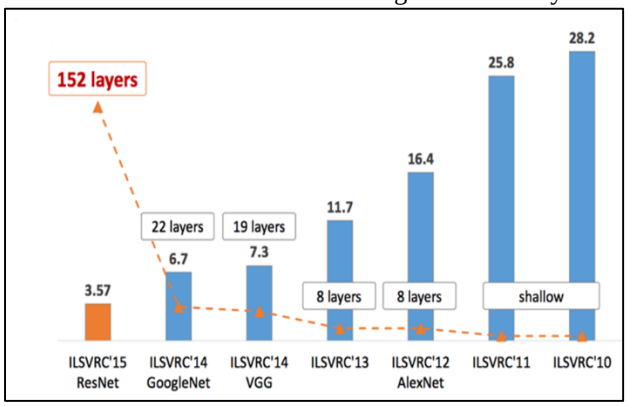

Fig. 12: Comparison of Top-5 error rate of various deep CNN Architectures

V. CONCLUSION

Deep learning is as of now a prominent research direction, the use of convolution neural network convolution layer, pool layer and the whole connection layer and other basic structure, you can let the network structure to learn and extract the relevant features, and to be used. This feature provides many conveniences for many studies, eliminating the need for a very complex modeling process. In addition to that, deep learning is now applied in image classification, object detection, attitude estimation and image segmentation. Deep Learning algorithms are performing quit good in these areas and nowadays people more attracted to deep learning than traditional methods as it gives better performance. At present deep learning algorithms are used in supervised learning as other layers are learning on their own but we calculate the loss for fully connected layers and adjust the parameters, in future we will be applying deep learning in unsupervised learning. Deep Learning is also increasing its application area by getting into semantic analysis, text analysis putting Recurrent Neural Network(RNN) algorithm in action. Deep Learning algorithms are self learning algorithms and that's, what makes them so stunning and attractive. At present, research combined with deep learning

and intensive learning is still in its infancy, but some research in this area has achieved good performance in multi-object recognition tasks and video game learning. Natural Language Processing(NLP) methods are also using the phenomenon of deep learning which helps them to understand the text content more effectively. People now started using depth of learning and simple reasoning techniques in the field of voice and image. if these network features gets more optimized it may work in other application area as well and it can achieve great outcomes.

REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, Nov. 2004.
[2] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition (CVPR)", IEEE Computer Society Conference on, vol. 1, pp. 886-893, 2005.
[3] Yamashita, R., Nishio, M., Do, R.K.G. et al.,"Convolutional neural networks: an overview and application in radiology",Insights Imaging (2018) .
[4] LeCun, Y. & others. LeNet-5, convolutional neural networks. URL http//yann. lecun. com/exdb/lenet20 (2015).
[5] Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. Adv. Neural Inf. Process. Syst.1–9 (2012).
[6] Zeiler, M. D. & Fergus, R. Visualizing and Understanding Convolutional Networks. arXiv Prepr. arXiv1311.2901v330, 225–231 (2013).
[7] Simonyan, K. & Zisserman, A. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. ICLR75, 398–406(2015).
[8] Szegedy, C. et al.Going Deeper with Convolutions. ArXiv:1409.4842(2014).
[9] He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. Multimed. Tools Appl.77, 10437–10453 (2015).

