RESEARCH ARTICLE OPEN ACCESS

CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES-A REVIEW

HARISH MS*,

*(Department of ECE, Government College of Technology, and Coimbatore
Email: cb105ec021@gmail.com)

--************************----------------------------------

Abstract:

Convolution Neural Network (CNN) architectures have been significantly evolved during the last two decades. The field of Deep learning (DL) is significantly improved over time where the new ideas implemented to have better computing power performance for diverged applications especially in Image processing classification and segmentation applications. In this paper the Convolutional Neural Network architectures evolved so far are reviewed. This article helps the beginner to have an over view of CNNs and their modern architectures.

Keywords — **Convolutional Neural Networks, Image Recognition, Parameter Optimization.**

--************************----------------------------------

I. INTRODUCTION

Computer Vision is used to study the human visual science through collection of images or videos for analysis, segmentation, recognition of objects and classification applications. The use of image recognition technology can be used for detection and identification of specific target objects, subjective image classification assessment and other issues. At present image recognition has a great commercial value and application prospects in image search, face recognition analysis. In early image recognition system, feature extraction methods such as scale invariant feature transform and histogram oriented gradients were used. These features are a feature of manual design. For different image classification problems, the extracted features have a direct impact on the performance of the system, so the researchers are in need to design adaptable image recognition systems to improve system performance. In the early 1960s, Hubel and Wiesel, through the study of cat's visual cortical system of cat, proposed the concept of receptive field and found the hierarchical processing mechanism of information in the visual cortical pathway, Nobel Prize in Physiology or Medicine. By the mid-1980s, Fukushima et al., which was based on the concept of receptive field, could be seen as the first realization of Convolution Neural Networks (CNNs) and the first neuron based between the local connectivity and the hierarchical structure of the artificial neural network. The neural cognition machine decomposes a visual pattern into many sub-patterns, and these sub-pattern features are processed by hierarchical cascaded feature planes so that the model is very good even in the case of small targets of the target object Recognition ability. In recent years we have witnessed the birth of various Convolutional Neural Network architectures. These evolved innovations can be categorized as parameter optimization, regularization and structural reformulation. However, it is observed that the main improvement in CNN performance was mainly due to restructuring of processing units and designing of new blocks. In this paper the different CNN architectures which are used in image classification applications are reviewed in their chronological order of innovation.

II. CNN ARCHITECTURES

A) LeNet-5:

LeNet was proposed by LeCuN in 1998. It is famous due to its historical importance as it was the first CNN, which showed state-of-art performance on hand digit recognition tasks. It has the ability to classify digits without being affected by small distortions, rotation, and variation of position and scale. LeNet exploited the idea that image features are distributed across the entire image, and convolutions with learnable parameters are an effective way to extract similar features at multiple locations with few parameters. The average pooling layer as we know is now called a sub-sampling layer. This architecture has about 60,000 parameters. LeNet was the first CNN architecture, which not only reduced the number of parameters and computation but ingeniously learned features. Figure 1 shows the LeNet architecture.[1]

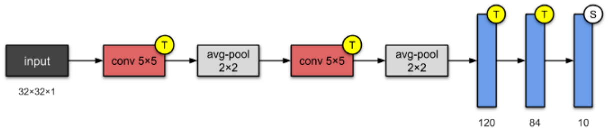

Fig 1 LeNet Architecture

B) AlexNet:

AlexNet is considered as the first deep Convolutional Neural Network architecture, which showed groundbreaking results for image classification and recognition tasks. AlexNet was proposed by Krizhevesky et al. who enhanced the learning capacity of the CNN by making it deeper and by applying a number of parameter optimizations strategies. AlexNet has eight layers- five Convolutional and three fully connected layers. AlexNet has significant importance in the new generation of CNNs and it has started a new era of research in CNNs. Figure 2 shows the AlexNet architecture in detail. They were the first to implement rectified linear unit activation function (RELU).[2]

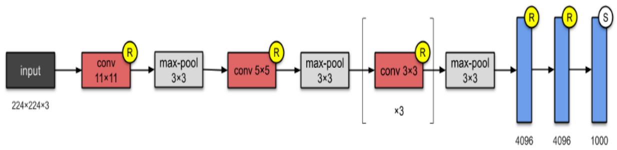

Fig 2 AlexNet architecture

C) VGG16:

[3] With the successful use of CNNs for image recognition, Simonyan and Zisserman proposed a simple and effective design principle for CNNs. This new architecture was termed as VGG and was modular in layers pattern. VGG was made deeper (19 layers) than AlexNet and ZefNet. VGG replaced the 11x11 and 7x7 filters with a stack of 3x3 filters layer and experimentally demonstrated that concurrent placement of 3x3 filters can induce the effect of the large filter. VGG suggested that parallel placement of small size filters make the receptive field as effective as that of large size filters (5x5 and 7x7). CNN complexity in VGG was further regulated by placing 1x1 filters in between the Convolutional layers, which in addition learn a linear combination of the resultant feature maps. Furthermore, tuning of the network is performed by placing max pooling 39 after Convolutional layer and padding was performed to maintain the spatial resolution. VGG16 has 13 fully connected and 3 fully connected layers. It consists of 138 million parameters and takes up about 500MB of storage space. VGG suffered from high computational burden due to the use of about 140 million parameters. Figure 3 shows the VGG16 CNN architecture.

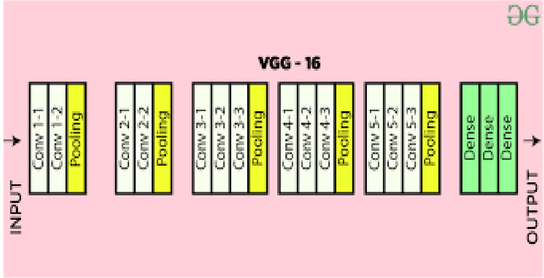

Fig 3 VGG16 Architecture

D) GoogleNet:

In the year 2014-ILSVRC competition was held and GoogleNet was the winner of the competition. GoogleNet is also known as the Inception-VI. It introduced the new concept of inception module block in the CNN, whereby it incorporates the multi-scale convolutional transformations using split, transform and merge idea for feature extraction. The inception block encaspulates filters of various dimensions 1x1, 3x3 and 5x5. to capture spatial information in combination with channel information at different spatial resolutions. In GoogleNet, conventional convolutional layer is replaced by small blocks similar to idea of substituting each layer with micro NN as proposed by Network in Network (NIN) architecture. In addition to improvement in a learning capacity, GoolgleNet focus was to make CNN parameter very efficient. Instead of using a fully connected layer, connection density were reduced by using global average pooling at the last layer. The number of parameters were reduced to five million from 40 million due to parameter tuning. Heterogenous topology was the main drawback of GoogleNet that needs to be customized as per module regulations. Figure 4 shows the blockdiagram of the Inception Block architecture. Another, limitation of GoogleNet was a representation bottleneck that drastically reduces the feature space in the next layer thus sometimes leads to loss of information that is useful. The motivation for inception version two and three is to avoid representational bottlenecks and efficient computations are done by using factorization methods. [4]

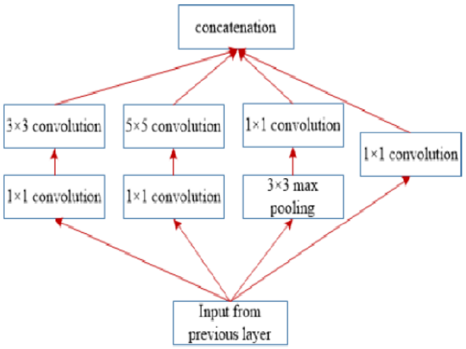

Fig 4 Inception Block architecture

E) ResNet:

ResNet was proposed by He, K, Zhang and it is considered as the continuation of Deeper Nets. The previous CNN architectures have seen just increase in the number of layers but accuracy gets saturated. When compared with AlexNet and VGG16, ResNet is 20 and 8 times deeper respectively.ResNet performed well in image recognition and localization tasks. ResNet CNN architecture is one of the early adopters of batch normalization ResNet50 version is with 26 million parameters. [5]

F) Xception:

Xception CNN architecture is an inspiration from Inception. Xception modified the original inception block by making it wider and replacing different spatial dimensions (1x1, 5x5, 3x3) with a single dimension (1x1 followed by 3x3). Xception makes computation easy by separately convolving each channel across spatial axes, which is followed by pointwise convolution (1x1 convolutions) to perform cross-channel correlation. The novelty introduced in Xception is the introduction of CNN based entirely on depthwise seperable convolutional layers. Xception does not reduce the number of parameters, but it makes learning more efficient and results in improved performance. Figure 5 shows the Xception CNN architecture.[6]

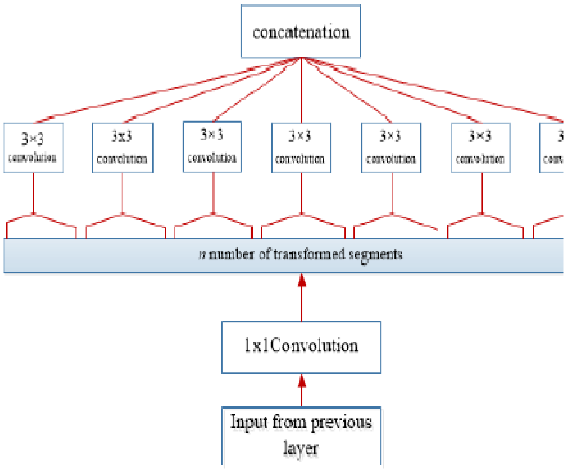

Fig 5 Xception CNN Building Block

G) DenseNet:

DenseNet is one of the CNN architecture to solve the vanishing gradient problem. The problem with ResNet was that it explicitly preserves information through additive identity transformations due to which many layers contribute very little or no information. Moreover, ResNet has a large number of weights as each layer has a separate set of weights. To address this problem, DenseNet used cross-layer connectivity but, in a modified fashion. DenseNet connected each layer to every other layer in a feed-forward fashion, thus feature maps of all preceding layers were used as inputs into all subsequent layers. DenseNet has narrow layer structure; however, it becomes parametrically expensive with an increase in a number of feature maps. The direct admittance of each layer to the gradients through the loss function improves the flow of information throughout the network. This incorporates a regularizing effect, which reduces overfitting on tasks with smaller training sets.[8]

H) ResNext:

RexNext aslo called by the name Aggregated residual transform network. ResNext simplified GoogleNet architecture by fixing spatial resolution using 3x3 filters within split, transform, and merge block. ResNext used multiple transformations within a split, transform and merge block and defined these transformations in terms of cardinality. All transformations were applied at low embedding's using 1x1 filters. Moreover, in ResNext the problem of vanishing gradient and degradation was addressed by adding dropout inbetween layers. Figure 6 shows the ResNext CNN architecture building block.[7]

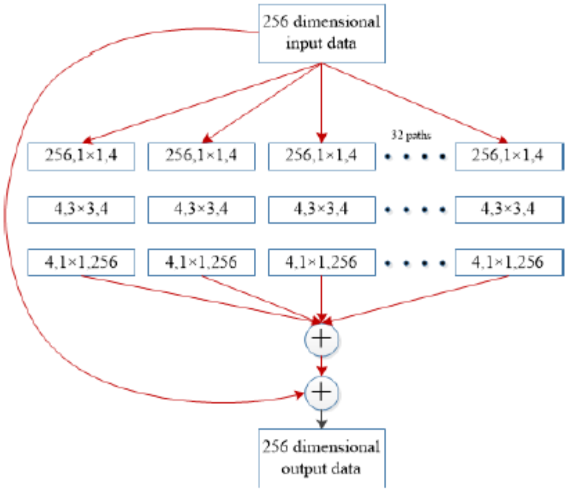

Fig 6 ResNext Building Block

III. CONCLUSION

Convolutional Neural Networks has made a commendable progress in Image processing applications with Deep Learning study. The advancements in CNN architectures can be categorized based on activation function, optimization and learning algorithms. This paper explores and reviews the different Convolutional neural network architectures that are used in most of the computer vision related deep learning applications for Image Processing and explains the different CNN architecture in their order of innovation.

IV. FUTURE DIRECTION

The motivation in terms of future work will be in the simulation implementation of various CNN architectures in Image Processing applications viz in the satellite imagery analysis using CNN and hardware validation of the developed model using GPU (Graphics Processing Unit).

V. REFERENCES

[1] Yan LeCun, Leon Bottu, Yoshua Benigo and Patrick Hafner, Gradient Based Learning Applications to document recognition, Proceedings of the IEEE (1998).

[2]AlexKrizhevsky, Ilya Sutskever, Geoffrey Hinton, ImageNet classification with Deep Convolutional Neural Networks, NeurIPS, 2012.

[3] Karen Simonyan, Andrew Zisserman, Very Deep Convolutional Networks for Large-scale Image recognition, ArXiv e-prints, 2014.

[4] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, Going Deeper with Convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Microsoft, Deep residual learning for image recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[6] François Chollet, Google, Xception: Deep Learning with Depth wise separable Convolution, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[7] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, Kaiming He. University of California San Diego, Facebook Research, Aggregated residual transformation for Deep Neural Networks, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[8] He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image steganalysis. ArXiv e-prints. arXiv1512.03385 1–17 (2017)

VI. AUTHOR PROFILE

Mr Harish MS pursued Bachelor of Technology from Amrita University in Electronics and Communication Engineering and MSc in Telecommunication Engineering from Middlesex University, UK and Master of Engineering in Applied Electronics from Government College of Technology, Coimbatore, India. He is an academician having more than five years of experience in teaching engineering subjects. His major research interests are in Image Processing, Machine and Deep Learning.