How to Cite:
Singh, K., Singh, D., & Mishra, N. (2022). Review: Convolutional neural networks and its architecture. *International Journal of Health Sciences*, *6*(S1), 9183–9190.
https://doi.org/10.53730/ijhs.v6nS1.7074

Review: Convolutional neural networks and its architecture

Kavya Singh
Computer Science & Engineering, Galgotias University, Greater Noida
Email: kavyasingh.1899@gmail.com

Deepanshu Singh
Computer Science & Engineering, Galgotias University, Greater Noida
Email: arushrajsingh27@gmail.com

Dr Nitin Mishra
Computer Science & Engineering, Galgotias University, Greater Noida
Email: Drnitinmishra10@gmail.com

***Abstract*---**Deep Learning is-one of the machine learning areas, applied in recent areas. Various techniques have been proposed depends on varieties of learning, including unsupervised, semi-supervised, and supervised-learning. Some of the experimental results proved that the deep learning systems are performed well compared to conventional machine learning systems in image processing, computer vision and pattern recognition. This paper provides a brief survey, beginning with Deep Neural Network (DNN) in Deep Learning area. The survey moves on-the Convolutional Neural Network (CNN) and its architectures, such as LeNet, AlexNet, GoogleNet, VGG16, VGG19, Resnet50 etc. We have included transfer learning by using the CNN's pre-trained architectures. These architectures are tested with large ImageNet data sets. The deep learning techniques are analyzed with the help of most popular data sets, which are freely available in web. Based on this survey, conclude the performance of the system depends on the GPU system.

***Keywords*---**LeNet, AlexNet, ResNet, DenseNet, VGGNet.

Introduction

Convolution Neural Network are basically a special class of Artificial Neural Network that we see in regular neural network which expect images as input. They are designed to work on images mostly to handle computer vision problems like ANN also have weights ,neurons and bias unit and the weights in these CNNs

are also eliminated by optimizing an appropriate objective function because input to these networks or images it allows for two things one is fast connection as well as parameter sharing because the images are used as input these two concept are possible basically it is possible they have sparse connections as well as sharing of its weights between the output neurons in the network. CNNs take images as a input so, there are different types of images grey scale as well as RGB (Red, Green, Blue). As, the size of input images is increase ANNs do not scale very well. Another aspect why we should not use because ANNs takes an input vector in that process we lose the spatial structure of the data .But, the images have spatial structure which we do not want to exploit .In CNNs there is parameter sharing which again reduces the number of weights and the parameter sharing which in turns enables us to exploit the local connecting of neural.

Convolution neural network

Sajja Tulasi Krishna, Hemantha Kumar Kalluri, in her research paper mentioned six layers to perform CNNs operation.

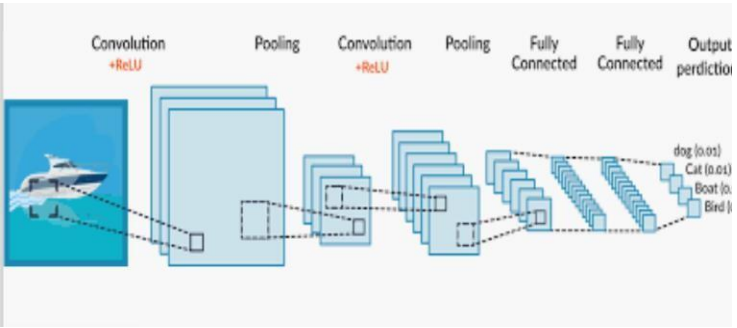

Fig 1. Convolution Neural Network

Input layer: Input layer accepts the raw images and forwarded to other layers for extracting the features of the images.

Convolution Layer: Every output is connected a small neighborhood in the input through weight matrix(filter or kernel) .we can define multiple kernels for every convolution layer each giving rise to an output. Each filter is moved around the input giving rise to one 2-D outputs. The outputs corresponding to each filter are stack giving rise to the output volume.

ReLu (Rectified Linear Unit): After performing the convolution layer the next layer is ReLu. The ReLu functions replaces the the all negative number of the convolution layer with zero which decreases the effective time taken to train the model.

Pooling: Provides translational invariance by subsampling .It reduces the size of the feature map. Average pooling and max pooling are commonly used to perform the pooling.

Fully Connected Layer: The last layer of the model is fully connected which takes all the filtered images and express it into label and categories

Softmax Layer: the Softmax Layer is used just before the output layer which is used to give the decimal probability to every class .These decimal probabilities lied between 0 and 1.

Convolution neural network architectures

To perform image recognitions by CNN we have several architectures.
LeNet Architecture
It was one of the earliest instances of convolution neural network used for image recognitions it is specific application was for digit recognition and apparently had commercial application where it was used to read millions of checks in banks. LeNet took as input 32 *32 images there are in this network was trained with MNIST dataset .So, the images were about the size 28*28 and after modification like data augmentation it took image 32*32 .Typical architecture is basically input followed by a convolution followed by a pooling layer and this was repeated leading to finally to a couple of connected layer and which is basically one of ten classification[1].

The first layer had 5*5 convolution no zero padding which give rise to 28*28 output followed by 2*2 pooling ,this was an average pooling operation then followed by 5*5 convolution again no padding which give rise to 10*10 output and subsequently another 5*5 gives and then a max pooling which give rise to 16*5*5 maps and we do again 5*5 convolution on that giving rise to 120*1*1 output and this is called fully connected.

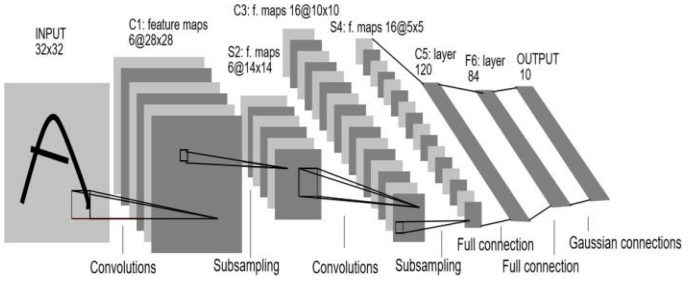

Fig 2. Architecture of LeNet for digit recoginition [Yann Lecun]

AlexNet Architecture

AlexNet was the image net large scale image recognition challenge. This was the Deep Neural Network about seven layers and 60 million parameters. The structure of this network was very similar to LeNets architecture in terms of max pooling and convolution operations but then it had various other convolution in it. The RGB images as input layer of size 224*224*3. So, the first convolution layer had 11*11 filter with the stride of 2 for giving rise to 96 feature map of size 55*55 and the we had max pooling operation using 3*3 kernel and a stride of 2 again followed by 5*5 convolution . All max pooling were 3*3 with stride of 2 and all convolution operation and again 5*5 with stride of 1.

The layer which have without any pooling , they had convolution with the padding one to preserve the size of the network . All intermediate had 3*3 kernel

size and have padding one .For fully connected layers in the end have 4096 times 4096 weights and max pooling following this was 256*13*13 as input with the stride of 2.[2]

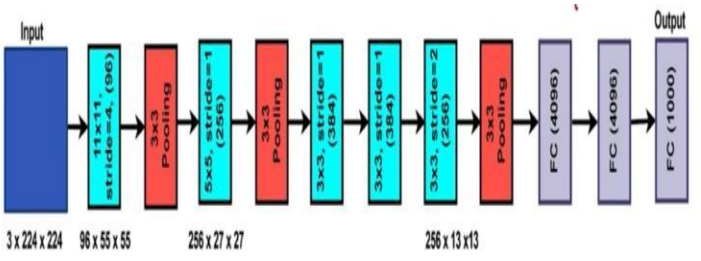

Fig 3. AlexNet architecture [A. Krizevsky, I. Sutskever, G.Hinton imageNet classification with deep neural network 2012]

ResNet Architecture

Introduced by [He et al at the ILSVRC 2015] They observed that with network with network depth increasing accuracy gets saturated (which might be unsurprising) and then degrades rapidly. The network comprises of novel approach pathway called skip connection. These connection provide alternate pathway for data and gradients to flow and thus making training possible .This connectivity patterns aids in training network with 152 layers while being less complex than VGGNet[3]

DenseNet Architecture

DenseNet architecture it improves gradient propagation by connecting all layers directly with each other, So, if we have L number of layers there will be L connections .However in Dense Net there will be [L*(L+1)]/2 connections. Here is a particular incarnation of dense net the input here will be k input maps for an RGB images . The first layer creates a k feature maps in the below fig 4 (k=4). As we go deeper in the network if we go to the second set of layer as it takes input not only from the previous layer but also from the input layer as we go to the next layer it takes input of all the preceding layers.

We notice that as we go deeper into the network this become kind of unsustainable. There is a problem of feature which is max explosion. To overcome from the they fix the number of output maps from each of the layer and also created these, so called Dense blocks. Each dense block contains a fixed number of layers inside them and among those layers feature maps are shared and output from particular block is given from Transition Layer. Transition Layer allows formats pooling which typically leads to a reduction in the size of the feature map.[7]

Fig 4 DenseNet Architecture [G.Huang]
Fig 4 is an example of dense block where in each convolution block receives input from the preceding layer. The memory explosion is circumvented by setting the number of feature map learnt in convolution layer to a small value (k=4)

VGGNet Architecture

VGGNet is Visual Geometry Group at Oxford university .This particular architecture was entered into the 2014 ImageNet challenge .It had 16 weights layer .Design is very similar to LeNet and Alexnet architecture .As we go deeper into the network we increase the size of the depth or the number of the feature map increased So, tha =t was incorporated they stuck to one filter size 3*3 filtered throughout all the layers .There are around 130 to 140 million parameters depending upon which network we use[6]

GoogleNet Architecture

The basic building block of the GoogleNet is Inception module .Inception module incorporates both this concept since that every layer has all possible filter size .So, they build clock convolution block which had multiple filter size and let the network backprop the learning to which limitless .Backprop decides which weights to updates based on objective function.

[]in his research paper proposed two incepton module first one is called naïve implementation and the Fig5 is actual implementation .We have input feature map 1*1 ,3*3, 5*5 convolutional layer and 3*3 max pooling also applied to the feature map the output of each of them are taken and then concatenated .But in this module there are lots of numeric computation to reduce this they use 1*1 convolution layer befor each convolution and poling operation in naive implementation of inception model it reduces the depth of the feature map.

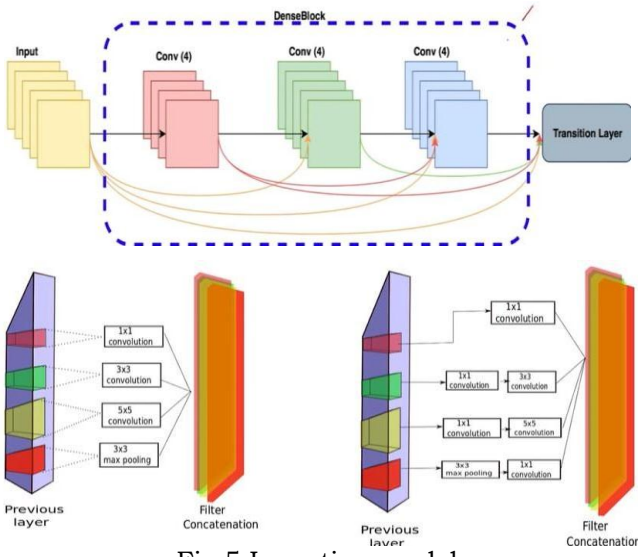

Fig 5 Inception model

In GoogleNet architecture there was initial set of convolutional layer and max pooling which reduces the size of the input to 28*28*192 .So, the input used as usual was 224*224*3 . These reduce size of input 2828192 was used as input of the inception layer sequence of inception layer followed by max pooling again and so on till we have typical output with one thousands activations.

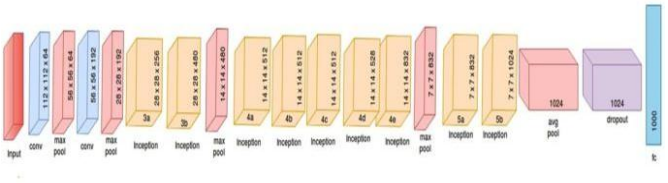

Fig 6 GoogleNet architecture

Sajja Tulasi Krishna, Hemantha Kumar Kalluri in his research paper given a table of error rate of each architecture. Below

Table 1.

S.No	ARCHITECTURE	ERROR RATE%
1.	LeNet[1]	28.2
2.	AlexNet[2]	16.4
3.	VGGNet	7.30
4.	GoogleNet	6.70
5.	ResNet[5]	3.57
6.	DenseNet[6]	3.4

Table 2
CNN architecture with number of parameter

	DISCOVERED YEAR	DISCOVERED BY	NUMBER OF PARAMETER
LeNet	1998	Yann LeCun et al.	60k
AlexNet	2012	Alex k et al.	62.3 million
VGGNet	2014	Simmonyan, Zisserman	138 million
GoogleNet	2014	Google	4 million
ResNet	2015	Kaiming He	25 million
DenseNet	2015	G. Huang et al	48 million

Results and Discussion

Krizhevsky et al [2] proposed a network having five convolution layer and three fully connected layer . The researchers got best top 1 and top error rate 37.5% and 17% respectively .The researchers nearly took four to five days to train the model on GPU. Transfer Learning is used in VGGNet , GoogleNet , ResNet to get the better image classification accuracy by using data augmentation which provide best result when the when the training data is less. After reviewed all the architectures of image recognition we saw that ResNet architecture gives the low error rate in all of them .But in ResNet there is a problem of map explosion so we use DenseNet architecture which give slightly more error than ResNet but there is no map explosion because in DenseNet there is a feature of Dense block each dense block contains a number of fixed layer inside them and among those layers feature map are shared.

Transfer Learning

One of the strategy used to trained the deep neural network when the number of datapoints available for training for a particular task is very low If training datapoint is very less then it is not sufficient to train a deep neural network so to train less training data we take network like AlexNet or VGG or Inception. We forward pass through a pre trained network and store the embedding system. Use the embedding to train the traditional machine learning algorithms such as SVM or Logistic regression to classify data appropriately.

In transfer learning we can take the network with pre trained weights. Then modify the classification layer from 1000 neurons to number of classes in the new data set and then train the network .The process in which we train the network is called Fine tuning or in general is called Transfer Learning . [4]

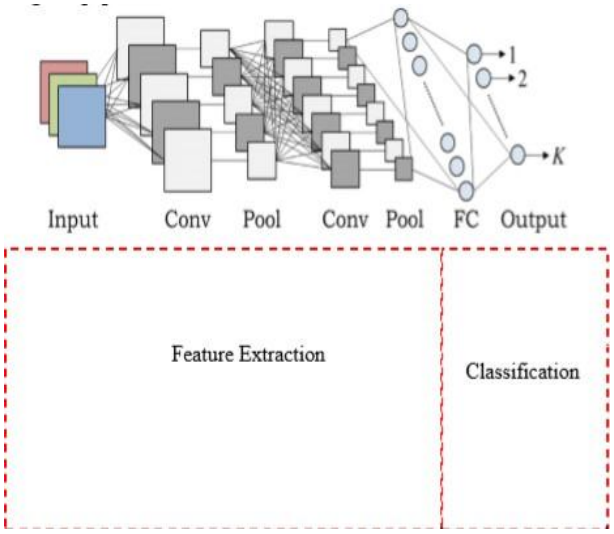

Fig 7 Transfer Learning [4]

Conclusion

In this review paper, I review the convolutional neural network which is used for image pattern recognition .I reviewd the architecture of the CNNs and some pre trained model .I found that ResNet architecture have less error rate and accuracy is better than other architecture but there is some problem in ResNet architecture which is map explosion to resolve this DenseNet architecture is used.

References

1. LeCun, &Yann, (1998). "Gradient-based learning applied to document recognition", Proceedings of the IEEE, IEEE. Vol. 86.11, pp. 2278-2324.
2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012).”Imagenet classification with deep convolutional neural networks”, In Advances in neural information processing systems, pp. 1097-1105.
3. He, K., Zhang, X., Ren, S., &Sun, J. (2016). “Deep residual learning for image recognition”. In Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE (CVPR), pp. 770-778.
4. Sajja Tulasi Krishna, Hemantha Kumar Kalluri
5. Hana D., Qigang Liu, &Weiguo Fan. (2017), “A New Image Classification Method Using CNN transfer learning and Web Data Augmentation”, Expert Systems with Applications, Elsevier, Vol. 95, pp. 43- 56.
6. Simonyan, Karen, & Andrew Zisserman (1998)," Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv, pp. 1409.1556.
7. G.huang,Z.Liu,L.V.D matten et al”Densely connected convolutional network” 2017 IEE