International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Comparative Analysis of CNN Architectures

Shrihari Kulkarni[1], Sanketh Harnoorkar[2]

[1,2]*Department of Information Science and Engineering, R V College of Engineering, Bengaluru-560059*

ABSTRACT:

This paper presents with a detailed comparison of 3 popular models of Convolution Neural Networks and their performance of popular datasets. The models compared are LeNet, AlexNet and VGG-16. These models are run on mnist and iris datasets using the same number of epochs, learning rate and sample size in order to bring about a fair comparison.

INTRODUCTION:

Convolution Neural Networks are very powerful tool for analyzing images and drawing out impactful information from them. They are a superior model of artificial neural networks known to give high efficiency in lesser training and testing time. For example, mnist classifier using artificial neural network with categorical cross entropy and Stochastic Gradient Descent (SGD) is 87% while that achieved by simplest of convolution neural network model is greater than 97%. We can clearly see a remarkable difference in accuracy. Similarly, even learning time is also smaller in convolution neural networks than simple artificial neural networks.

CNN uses the pattern recognition feature and hence is able to successfully able to classify and detect many images. Any regular CNN implementation used features like padding, strides, volume operations, pooling and filters.

Padding: refers to the process of adding extra data to the edges on matrix so that the Convolution neural network does not lose much information while learning. A valid padding refers to no padding taking place whereas a same padding refers to padding such that the output dimensions remain same as the input.

Strides: Stride refers to the number of many cells the filter is to be moved in the input so as the get the next set of result.

Pooling: This is used to reduce the size of dimensions and also to speed up the learning process.

Convolution neural networks also have advantage over traditional artificial neural networks in terms of parameter sharing and sparsity of connections.

Over time a lot of architectures have been developed which involve a sensible combination and permutation of above mentioned features.

LITERATURE SURVEY

Exploiting Image-trained CNN Architectures for Unconstrained Video Classification [1] details about the various models as a hybrid which can be used for classification of images. It details about choice, sampling, collection, pooling of all kinds, fusion, fisher vectors. They have used MEP classifier in this paper and fusion performance was observed.

Comparison of Three Different CNN Architectures for Age Classification [2] deals with comparison of CNN architectures for face detection. It compares 6 layer CNN, resnet-18 and resnet-34. It also identified the various rotation types and their impact on performance.

Advanced CNN architectures [3] deals with building blocks of CNN. It explains how residual neural networks work and also gives survey extensions to resnet and other neural architectures. It also covers details on Resnet as implicit ensembles, as learning iterative refinements and also deals with connections with recurrent networks and brain. It deals with architectures such as WaveNet, Inception Resnet and XceptionNet. It also talks about different types of fully connected layers.

ImageNet Classification with Deep Convolutional Neural Networks [4]. This is the paper in which AlexNet was introduced. It used ReLu non linearity and it was trained on multiple GPU's. The paper details the processes undertook to fit the ImageNet dataset and the accuracy it eventually achieved. It details on the data augmentation techniques and also the dropout methods. It tells about the way in which learning was done and also on the various quality evaluations done by the group of researchers is worth noting.

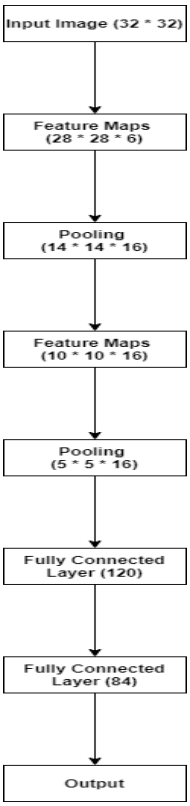
Visualizing and Understanding Convolutional Networks [5]. It details with things such as un pooling, rectification, filtration, feature visualization, feature visualization during training. It also talks about minor details such as architecture selection and occlusion sensitivity. It introduces new methods of feature generalization.

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION [6] paper deals with the model popularly known as VGG is a simple and in deep network architecture which deals with training, configurations and testing on eILSVRC-2012 dataset and succeed in achieving a high accuracy. It also talks about the classification framework to choose and on multi scale evaluation of models.

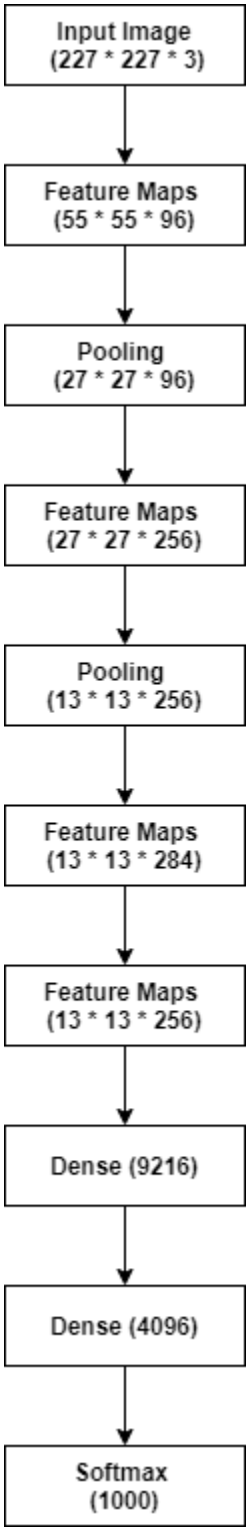
OVERVIEW OF ARCHITECTURES

LeNet:

LeNet-5(simply called as Lenet) is a simple CNN structure proposed by Yann LeCun in 1998.It is a simple network. It has 7 layers. 3 Convolution Layers, 2 pooling layers and 1 fully connected layer. It was one of the early architectures developed on Convolution Neural networks. The choices made while learning may also seem odd when compared to today's standards, but those choices were necessary considering the computational abilities at that time. The input for LeNet-5 may be a 32×32 grayscale image which passes through the primary convolutional layer with 6 filters having size of 5×5 and a stride of 1. The image dimensions changes from 32x32x1 to 28*28*6. The next layer is average pooling layer. At the time this paper was published, average pooling was the norm rather than max pooling



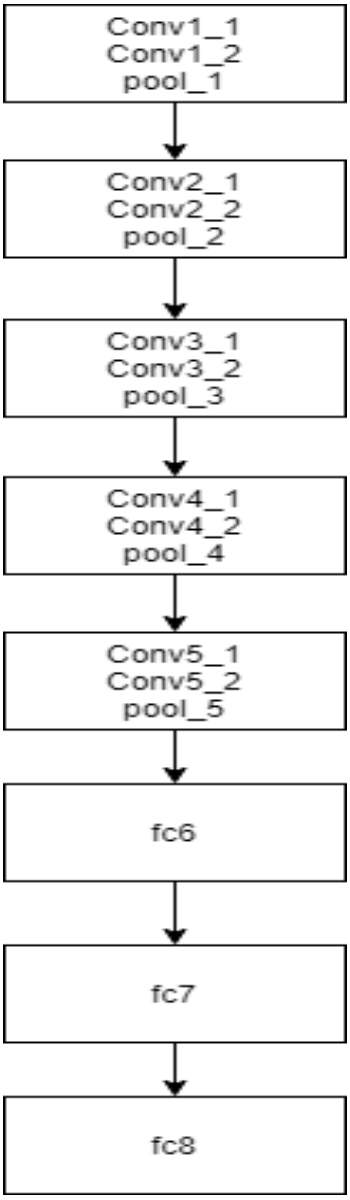
AlexNet:-After a long time after releasing LeNet, AlexNet was introduced was developed in 2012 by Alex Krizhevsky in 2012 as part of ImageNet Competition. The architecture is a larger



Than LeNet. This architecture changed the view of world and research community towards computer vision and brought a new interest in deep learning. AlexNet has a very similar architecture to LeNet, with the added insight of stacking multiple convolution layers before the pooling and activation layers, rather than alternating layers of convolution, activation, then pooling. It is also deeper and bigger. It also had dropout during training which reduced overfitting, and it had data augmentation (rotations, translations, color variation) which increased robustness. Similar networks were used by many others. These typically included repeating a few convolutional layers each followed by max pooling; then a few dense layers. But there was no standard about filter sizes to be used, how many convolutions before a max-pooling, etc.

VGG-16:

Refers to Visual Geometry Group-16 architecture developed in 2016. It has a total of 16 layers and model size of 528MB.It has only 3*3 convolutions it was trained for 2-3 weeks on 4



GPU's It has 138 million parameters. With the introduction of VGG, they brought some standards: it was suggested that all filters to have size of 3x3, max pooling should be placed after each 2 convolutions and the number of filters should be doubled after each max-pooling. And the original proposed VGG network was much deeper than the AlexNet.

DATASETS:

The MNIST database of handwritten digits, available from this page, features a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a bigger set available from NIST. The digits are size-normalized and centered during a fixed-size image. The original black and white (bi-level) images from NIST were size normalized to suit during a 20x20 pixel box while preserving their ratio. The mnist dataset has been made open source and can be easily downloaded. It has been cleanly segregated into train and test datasets and hence it is easy to train and test them.

Iris dataset contains 3 classes of 50 instances. Each class represents to each class of the iris plant. Iris dataset has been useful in many of the present day learning of deep learning and computer vision problems and is used as a benchmark to test any new architecture. Iris along with mnist are two of the most popular public datasets available.

These datasets are chosen because of their popularity and widely available methods to import them directly using a single line of code and for training as well as testing. And also their popularity allows us to benchmark their standards and quality.

RESULTS:

All the three models were run on mnist and iris dataset with the same learning rate and same number of epochs. They were also run on same datasets with no changes being done on the original images.ie: no image augmentation, no filtering initially Scaling was done for all images while using VGG-16 to 224*224 since that is the requirement of the architecture. The results were as follows:

Architecture	MNIST	IRIS
LeNET-5	97.2	96.4
AlexNet	99.2	98.6
VGG-16	98.2	99.2

As we can observe the performance of all the 3 architectures were great with all of them showing a much better accuracy than any artificial neural network possible. With VGG-16 having a better performance in IRIS and AlexNet when it comes to mnist database.

CONCLUSION

Thus this paper provides the detailed analysis of the three models in CNN architectures considering the different parameters taken into consideration while predicting the output of the image. Thus depending upon the requirement of the application the study helps to analyze which model to choose exactly depending upon the pros and cons of each architecture. Convolutional neural systems (CNNs) have achieved shocking accomplishments over an assortment of areas, including clinical research, and an expanding interest has risen in radiology. Albeit profound learning has become a prevailing technique in an assortment of complex errands, for example, picture grouping and article identification, it's anything but a panacea. Being acquainted with key ideas and points of interest of CNN just as restrictions of profound learning is fundamental so as to use it in radiology look into with the objective of improving radiologist execution and, in the long run, quiet consideration.

REFERENCES

1. S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, 2015, pp. 730-734, doi: 10.1109/ACPR.2015.7486599.

2. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review - Scientific Figure on Research Gate. Available from: https://www.researchgate.net/figure/Architecture-of-LeNet-5-LeCun-et-al-1998_fig2_317496930 [accessed 17 May, 2020]

3. Exploiting Image-trained CNN Architectures for Unconstrained Video Classification Shengxin Zha Northwestern University Evanston IL USA szha@u.northwestern.edu Florian Luisier, Walter Andrews Raytheon BBN Technologies Cambridge, MA USA {fluisier,wandrews}@bbn.com

4. A Comparison of CNN-based Face and Head Detectors for Real-Time Video Surveillance Applications Le Thanh Nguyen-Meidine1 , Eric Granger 1 , Madhu Kiran1 and Louis-Antoine Blais-Morin2 1 Ecole de technologie sup ´ erieure, Universit ´ e du Qu ´ ebec, Montreal, Canada ´ lethanh@livia.etsmtl.ca, eric.granger@etsmtl.ca, mkiran@livia.etsmtl.ca 2 Genetec Inc., Montreal, Canada lablaismorin@genetec.com.

5. Li, N., Ye, J., Ji, Y., Ling, H., Yu, J.: Saliency detection on light field. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014)