AN INVESTIGATION INTO DIFFERENT ARCHITECTURES OF CONVOLUTIONAL NEURAL NETWORKS

Ramitha M A, N. Mohanasundaram*

Abstract

In Deep Learning, a Convolutional Neural Network (CNN) is used to study the various aspects of visual resources. CNN features can be used for advanced tasks like classification and segmentation of images, detection of objects and other complex tasks. The state-of-the-art models consist of stacked convolutional layers. But modern architectures construct convolutional layers by using new ideas. They allow CNN to work more efficiently. This paper analyses the performance of some important CNN architectures in particular applications.

Keywords : CNN,Deep Learning, AlexNet,VGGNet, InceptionNet, ResNet, DenseNet, SENet, ILSVRC

I. INTRODUCTION

The structure and working of human brain are the inspiration behind the neural network and its basic building block is termed as neuron. Neuron represents a mathematical function and provides an output for the input provided. A layered architecture is used to organize neurons.

Convolutional neural network [1][2] is a neural network model and it comprisesan input layer, one or more hidden layers and one output layer. The input layer collectsadequateinformation from the given source. These data are used by the network for processing or learning. The input layer feeds the data into a hidden unit. The hidden unit (convolution layers)extracts the features and transforms it into something that the output unit (fully connected layers) can use.

The information collected by every node in a particular layer is passed to every node in the next layer. Each layermakes changes in the input signal. It is necessary to combine more hidden layers in the existing networkto make the network deeper. Figure 1.1 explains thestructure of CNN. CNN extracts different patterns of a given picture element.

ImageNet project is a visual databasecontaining several hundred images. A software contest named ImageNet Large Scale VisualRecognition Challenge (ILSVRC) has been conducted by Imagenet project. It analysesdifferent algorithms for detection and classificationof images [3]. There are several CNN architectures [4][5] such as AlexNet, VGGNet, ResNetInceptionNet, DenseNet, XceptionNet, and SENet. All these architectures are ILSVRC winners.

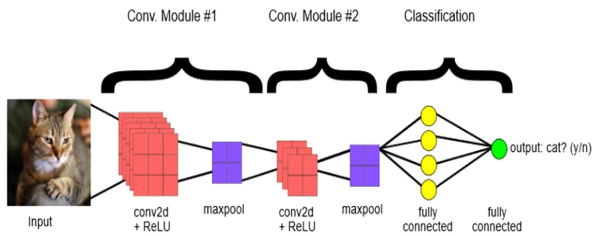

Figure 1.1 Architecture of CNN [6]

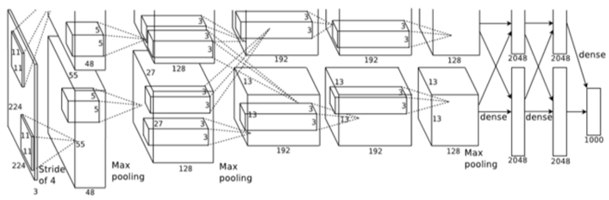

Figure 2.1 AlexNet [7]

This paper compares the performance of the architectures mentioned above and studies the performance in different applications.

II. RELATED WORKS

2.1 ImageNet Classification with Deep Convolutional Neural Networks [6].

Department of CSE, Karpagam Academy of Higher Education, Coimbatore, Tamilnadu, India
*Corresponding Author

AlexNetwas the first deep CNN architecture that offeredremarkable performance in image-recognition and classification.

AlexNet[7] holds eight learned layers which comprise five that are convolutional and three fully connected.It uses 650,000 neurons and incorporates 60 M parameters. Figure 2.1 depicts the detailed model of AlexNetl.

2.2 Very deep convolutional networks for large-scale image recognition [8].

VGGNet has introduced an effective design principle for deep CNN architecture. Instead of11x11 and 5x5filters VGG introduces 3x3 filters. The 2x2 and 3x3 filters are connected back-to-back.These back-to-back filters replace the large size filters.Figure 2.2 representsVGG[9] Net and its layered architecture in detail

Figure 2.2 Architecture of VGG

2.3 Going deeper with convolutions [10].

InceptionNet presents a new concept called Inception block. It uses three filters of different sizes as shown in figure 2.3. These filters are capsulated into one inception block. It exploits the idea termed as 'split, transform, and merge'.

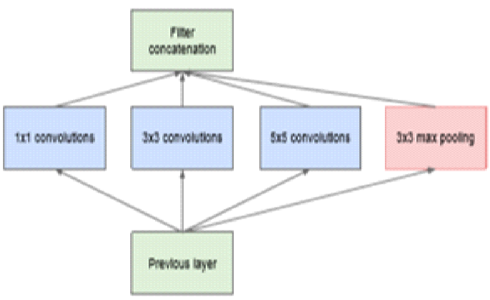

Figure 2.3 An Inception Block [10]

2.4 Deep Residual Learning for Image Recognition [11].

A 152-layered deep CNN is the idea behind ResNet. It is deeper and more accurate than the deep nets proposed before. In CNN, the original mapping is represented as F(x). But in ResNet, it is F(x)+x, where x is an input to the layer. Shortcut connection is used to achieve this result.

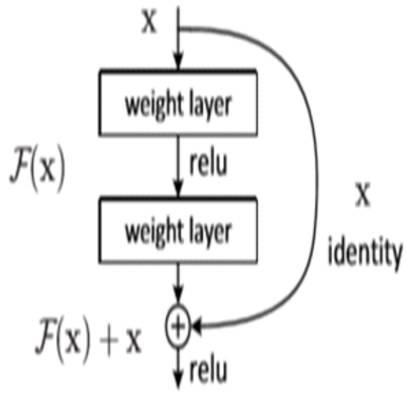

Figure 2.4ResNet architecture.

The shortcut connections are usedto carry out identity mapping. Outputs from the shortcut connections are added to the outputs from stacked layers. Figure 2.4 illustrates the concept of ResNet.

Accuracy gain of Deep residual net,the result an enhanced Inception-ResNet,isdue to an improved version of InceptionNet.

2.5 Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning [12].

Inception-V4 and Inception-ResNet are inspired byInceptionNet.By making the Inception layer deeper and wider, more efficient Inception-V4 is introduced. Thenumber of Inception modules in Inception-V4 is greater compared tothat of Inception modules in Inception-V3.

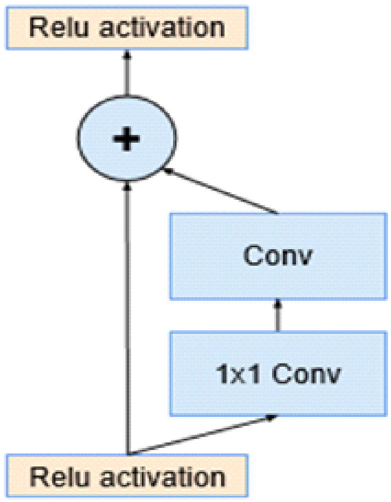

Figure 2.5 Inception-ResNet.

Inception-ResNetis a modified version of ResNet. It comprises residual learning and inception block. Inception ResNet and plain Inception-V4 have the same power. Figure 2.5 describes Inception-ResNet module.

2.6 Xception: Deep Learning with depth wise separable convolutions [13].

Instead of using the inception modules,XceptionNetintroduces depth-wise separable convolution [6]. 1x1 Convolution is used to obtain cross-channel correlations. The spatial correlation of every output channel is obtained by mapping the spatial correlation of each channel.

Each channel of an input goes through spatial convolution independently. Then a point-wise convolution is performed on it. This makes depth-wise separable convolutions. Figure 2.6 shows the architectural details of XceptionNet.

There are 36 convolution layers of 14 modules. All thesemodules have linear residual connections around them.

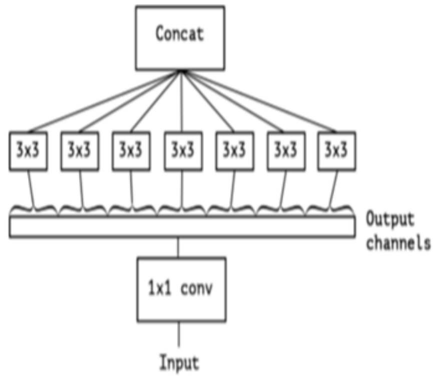

Figure 2.6 Architecture -XceptionNet. [13]

2.7 Aggregated Residual Transformations for Deep Neural Networks [14].

ResNeXtis the combined version of VGG and GoogleNet architecture. One 3x3 filteris used inside a 'split, transform and merge block'.It also adoptsthe idea of residual learning from ResNet.

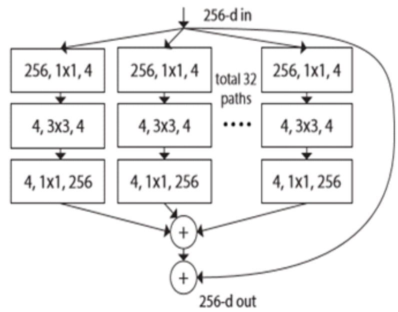

Figure 2.7 ResNeXt Architecture

Figure 2.7 shows the implementation of ResNeXt.

There are 14 modules to incorporate all 36 convolution layers. These 14 modules have linear residual connections. This is the simple architecture which Shows VGG/ResNet strategy.

ResNeXtcarries out a series ofoperations on the input provided. The operations provided by ResNeXt

I) Split: The given vector x undergoes a split operation. The result of this operation is a low-dimensional representation of the given input.

ii) Transform: The low-dimensional representation is transformed into wixi (x stands for the input vector and w for the filter weight).

iii) Aggregate: The wixi in all representations are added together.

2.8 Densely Connected Convolutional Networks [15].

The Densely-Connected feed-forward network collects all feature maps from its previous layer. These feature maps, along with its own feature maps, are provided to the inputs of the next layer as depicted in figure 2.8. The model obtained after training is highly parameter-compliant.

Feature maps of different layers are concatenated together. So, each layer receives input from subsequent layers and improves efficiency.

One of the advantages of using denseNet is that it resolves the problem associated with the vanishing gradient. It ensures the reuse of features and also strengthens its feature-propagation.

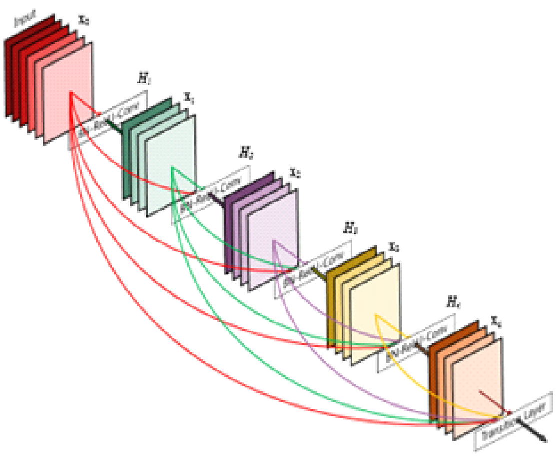

Figure 2.8 DenseNet.

2.9 Squeeze and Excitation networks [16].

SE blocks models' interdependencies between channels. These inter dependencies are used to recalibrate channel-wise feature maps.

For any given transformation consider a function maps the input X to the feature maps U where UϵRH X W X C. A squeeze operation is performed over the features U. The aggregated feature maps are aggregated with their spatial representations (H X W). The result of this operation is a channel descriptor.

With the help of the channel descriptor Excitation operation produces pre-channel modulation weights. These weights areapplied to feature map U, and the output received is considered as the output of the SE block. The detailed description of all these operations is depicted in Figure2.9. Thisoutput is directly fed into the subsequent layers of the network.

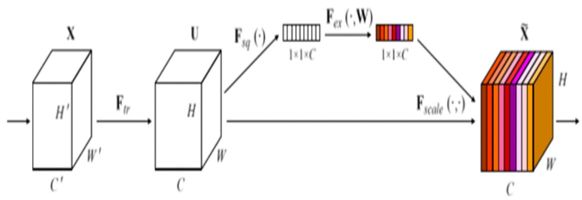

Figure 2.9 SE block

CNN Architecture	Year	Dataset	No: of Parameters	Top-1 Error	Top-5 Error
AlexNet	2012	ImageNet	60 M	40.70%	18.20%
VGG	2014	ImageNet	138 M	24.70%	7.20%
GoogleNet	2014	ImageNet	5 M	32%	6.67%
Inception-V3	2015	ImageNet	24 M	21.20%	5.60%
Res-Net -152	2015	ImageNet	26 M	22.85%	6.71%
Res-Net -200	2015	ImageNet	27 M	21.70%	5.80%
Xception	2016	ImageNet	23 M	21%	5.50%
Inception - V4	2016	ImageNet	43 M	20%	5%
Inception - ResNet- V2	2016	ImageNet	56 M	19.90%	4.90%
DenseNet-121	2016	ImageNet	26 M	25.02%	7.71%
DenseNet-161	2016	ImageNet	27 M	22.33%	6.15%
Res-NeXt-101	2016	ImageNet	25 M	20.40%	5.30%
SENet-154	2017	ImageNet	-	18.68%	3.79%

Figure 3.1 Comparison of different CNN architectures

III. COMPARITIVE STUDY

The overall performance evaluation of each architecture on the ILSVRC Imagenet Competition is compared as shown in Figure 2.10. Study shows that every architecture used the same dataset, ImageNet dataset. It shows that SENet has3.79 top 5-error rate for image classification.

IV. APPLICATIONS OF CNN

CNN can be used to detect Face Anti Spoofing [17] in real life.The dense layers of the ResNet152 model trained repeatedly to obtain better results in face anti-spoofing task.CNN shows better results in HeritageImage Classification [18].SE-ResNet-50 provides the best performance on heritage image classification. CNN plays a crucial role in medical fields.

Diabetic retinopathy stage classification can be performed using CNN [19][20][21][22][23]. Concatenation of AlexNet, VGG and Inception V3 provides better results for DR image classification. CNN can also be used to detect DR[24]. CNN provides better results in the detectionof defectsin fabrics also [25]. GoogleNte provides 100% accuracy in detecting defects in Fabrics.

Inception-v3 Shows better performance in Flower Classification [26] while AlexNet deep learning model shows better resultsin vegetable classification [27]. It is possible to classify Categorical images using RESNET [28].

V. CONCLUSION ANDFUTURE ENHANCEMENT

At present CNN plays a vital role in the classification and predictionof images. With the advancement in technology different CNN architecture has evolved over time. Figure 3.1 describes the comparison of different architectures. The study shows that the architectures after AlexNet classify images with reasonable accuracy. The architecture that classifies images with a top-5 error rate of 3.79% is SENet. A combination of two or more concepts may help people produce better results.

REFERENCES

[1] TianmeiGuo, JiwenDong ,Henjian LiYunxingGao"Simpleconvolutional neural network on Image classification" 2017 IEEE 2ndInternational conference on Big DataAnalysis.pp: 721- 724978- 1-5090-3619-6/17/$31.00 ©2017 IEEE.

[2] IfuAniemeka"A Friendly Introduction to Convolutional Neural Networks "August 22, 2017.

[3] Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, SanjeevSatheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, AdityaKhosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei) ImageNet Large Scale Visual Recognition Challenge.IJCV, 2015

[4] SaadAlbawi ,Tareq Abed Mohammed ,Saad AL-ZA "Understanding of a Convolutional Neural Network".International Journal of Innovative Technology and Exploring Engineering."ISSN:2278-3075 volume-9 issue-1, November 2019.

[5] RaimiKarim https://towardsdatascience.com/illustrated 10-cnn architectures-95d78ace614d.

[6] ML Practicum: Image Classification. https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks

[7] AlexKrizhevsky, IlyaSuteskever, Geoffrey E Hinton, ImageNet "Classification with Deep Convolutional Neural Networks", Communications of the ACM| Vol. 60 | NO. 6| June 2017

[8]Karen Simonyan& Andrew Zisserman, "Very Deep Convolutional Networks for Large-scale Image Recognition".arXiv 1409.1556 ,September 2014

[9] https://neurohive.io/en/popular-networks/vgg16/.

[10] ChristianSzegedy, WeiLiu, angqing Jia, Pierre Sermanet, ScottReed, Dragomir Anguelov, DumitruErhan,"Going deeper with convolutions".2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594, June 2015.

[11]Kaiming He Xiangyu Zhang Shaoqing RenJian Sun. "Deep Residual Learning for Image Recognition".2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp.770-778. doi: 10.1109/CVPR.2016.90

[12] Christian Szegedy ,SergeyIoffe Vincent Vanhoucke," Inception-v4, Inception-ResNet and the Impact of Residual connections on Learning".Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-2017),Pages 4278–4284, February 2017.

[13] Francois Chollet."Xception:Deep Learning with Depth wise Separable Convolutions".2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 1800-1807, doi: 10.1109/CVPR.2017.195, November 2017.

[14]SainingXie Ross GirshickPiotr Doll´ arZhuowen Tu1 Kaiming He, "Aggregated Residual Transformations for Deep Neural Networks".2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016 2016,5987-5995, 2016, 2017.

[15]GaoHuang ,Zhuang Liu, Laurens van derMaaten , "Densely Connected Convolutional Networks".2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),Honolulu, HI, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

[16]Jie Hu, Li Shen, Samuel Albanie,GangSun, EnhuaWu,"Squeeze-and-Excitation Networks" in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 8, pp. 2011-2023, 1 Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.Aug. 1 2020.

[17] ChaitanyaNagpal and Shiv Ram Dubeyy," A Performance Evaluation of Convolutional Neural Networks for FaceAnti Spoofing"2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1-8, doi:10.1109/IJCNN.2019.8852422, July 2019.

[18] Manh-Tu VU, Marie BEURTON-AIMAR,Van-Linh LE " Heritage Image Classification by Convolution Networks ",2018 1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR), Ho Chi Minh City, 2018, pp. 1-6, doi: 10.1109/MAPR.2018.8337517,April 2018.

[19] XioliangWang,Yongjin Lu, Yujuan Wang, Wei-Bang Che," Diabetic Retinopathy Stage Classification Using CNN" ,2018 IEEE International Conference on Information Reuse and Integration for Data Science. 2018 IEEE International Conference on Information Reuse and Integration (IRI), 2018, pages:465-471,2018.

[20] Nikhil M N, Angel Rose A,"Diabetic Retinopathy Stage classification using CNN", International Research Journal of Engineering and Technology (IRJET) e-SSN: 2395-0056 Volume: 06 Issue: 05 | May 2019.

[21] Debiao Zhang, Wei Bu, Xiangqian Wu, "DR Classification using Deeply Supervised ResNet",2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/ SCALCOM/UIC/ATC/CBDCom/IOP/SCI) , San Francisco,

CA, 2017, pp. 1-6, doi: 10.1109/UIC-ATC.2017.8397469, June 2018

[22]Sara HosseinzadehKassani,Peyman HosseinzadehKassani,Reza Khazaeinezhad ,Michal J. Wesolowski, Kevin," Diabetic Retinopathy Classification Using a Modified Xception Architecture",2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Ajman, United Arab Emirates, 2019, pp. 1-6, doi: 10.1109/ISSPIT47144.2019.9001846. Dec. 2019.

[23] Muhammad Mateen, Junhao Wen *, Nasrullah, Sun Song and Zhouping Huang." Fundus Image Classification Using VGG-19 Architecture with PCA and SVD".Symmetry2019, 11, 1.

[24] DarshitDoshi, AniketShenoy and Deep Sidhpura, "DR Detection using Deep CNN", 2016 International Conference on Computing,Analytics and Security Trends (CAST) College of Engineering Pune, India. Dec 19-21, 2016.2016, pp. 261-266, doi: 10.1109/CAST.2016.7914977.Dec. 2016

[25] K.K. Sudha, P. Sujatha ,"A Qualitative Analysis of Googlenet and Alexnet for Fabric Defect Detection".International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1, May 2019

[26] XiaolingXia , Cui Xu ,"Inception-v3 for Flower Classification" 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, 2017, pp. 783-787, doi: 10.1109/ICIVC.2017.7984661, June 2017.

[27] Ling Zhu, Zhenbo Li, Chen Li, Jing Wu, Jun Yue, "High performance vegetable classification from images based on AlexNet deep learning mode" Int J Agric&BiolEng, 2018; 11(4): 217-223,Vol. 11 No.4,2018.

[28] Mrs. ArpanaMahajan, Dr. Sanjay ChaudharyCategorical Image "Classification Based On Representational Deep Network (RESNET)".2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India,2019,pp.327-330 doi: 10.1109/ ICECA .2019.8822133. June 2019