May - June 2020
ISSN: 0193-4120 Page No. 3636-3644

Taxonomy and Complete Overview from AI to Deep Convolutional Neural Networks in Computer Vision

[1]Vijayalaxmi Chiniwar, [2]Mohamed Abdul Kader Jailani

[1,2]Assistant Professor, School of CSA, REVA University, Bengaluru, India
[1]chiniwarvijaya@reva.edu.in, [2]mohamadjilani@reva.edu.in

Article Info
Volume 83
Page Number: 3636-3644
Publication Issue:
May - June 2020

Abstract

Homo sapiens are the only species that top among all the other species on the planet earth and also known for their three important qualities such as learning (patana), thinking (chinthana) and analysis (mamthana). Perhaps this is the only reason, many researchers try to mimic the human intelligence in machine which is still continued till date in the domain known as Artificial Intelligence. The paper focuses on Non-symbolic AI technique which is one of the two main important approaches such as symbolic and non-symbolic AI technique used in solving a problem or finding a solution. The Non-symbolic technique which is either principle-based or similar to that of biologically inspired way of solving problem through distributed and connected Neurons. In this way the paper narrow down from AI to Neural networks to Deep convolutional neural networks and its variations. The major contribution of this paper as it take journey from AI to Deep CNN, the understanding of the concept especially in the application domain such as computer vision is been captured in the form of taxonomy.

Article History
Article Received: *19 August 2019*
Revised: *27 November 2019*
Accepted: *29 January 2020*
Publication: *12 May 2020*

Keywords: *Artificial Intelligence, Neurons, Neural Network, Deep Convolution Neural Network, Taxonomy.*

1. Introduction

Homo sapiens are the only species that top among all the other species on the planet earth. Humans stands first not because of muscle power but for their intelligence. Human Intelligence is the cognitive mental ability of the human to adapt to new environment by learning, thinking and analysis capability. It is evident from the history that the human intelligence has constantly evolved over the period by learning from past situations and adapting to new situations. This kind of natural human intelligence is missing in the machines due to the lack of learning/ thinking ability. Artificial Intelligence is the branch of Computer science which deals with the study of making the machine to think like human being [1]. The learning by machines is been tried to mimic human learning for solving problem, planning, learning, reasoning as well as for recommending. AI is carried out either through Symbolic or Non-Symbolic techniques. To fetch precise details and high-performance returns in domains such as computer vision and natural language processing etc. techniques like mathematical model and algorithms is not feasible and not recommended, as it is too difficult to characterize. As a result, Non-Symbolic AI techniques (Machine Learning, Neural Network, Deep Learning, and Evolutionary Algorithms [10] such as Genetic Algorithms) has been a boon to provide tangible new benefits. The Deep Learning is one such branch of AI and also a type of Non-symbolic AI technique which made a huge impact in recent years especially in the world of computer vision. Computer vision is the ability of computer to perceive images based on unsupervised deep learning methods to simulate on par with human vision. Due to the complexity and parallel nature of exponential computing, extensive CPU bound processing, a deep layered architecture of highly interconnected neurons which are to be processed by GPU’s are vital for processing this kind of unparalleled computing task. In recent years, instead using dense fully connected neural network with an extremely large amount of parameters to

May - June 2020
ISSN: 0193-4120 Page No. 3636-3644

be trained using back propagation and also the cause of vanishing gradients (gradient becomes zero) or exploding gradients (gradient changes abruptly) , a deep convolutional neural networks play vital role which resolve the complex nature of fully connected neural network to the simple convolutional neural network. Due to the simplicity of applying the convoluted and pooled kernels, a wide range of deep convolutional neural networks have come up to resolve and simplify the task of image classification or image recognition. LeNet [8], AlexNet[6] , GoogleNet [4], ResNet[7], and CapsuleNet are all recently designed Convolution Neural Network for image classification. The CNN which made a huge success in the world of 2D image classification among other types of fully connected deep neural network learning algorithm such as Feedforward neural Network, Radial Basis neural Network, Multilayer Perceptron, Recurrent neural Network and Modular neural Network.

Further CNN is classified into traditional CNN such as LeNet, AlexNet, GoogleNet and ResNet and Non-traditional CNN such as CapsuleNet. In this paper, the taxonomy and complete comprehensive study of various convolutional neural networks for image classification is analyzed using various parameters.

A. Inspiration from Biological Neuron

Learning is the fundamental capability of living organisms to acquire, learn, modify and create innovative knowledge / Intelligence in order to fine tune its existing knowledge. As a human being, one learn from environment , learn from example, learn from experts and self-learning, through the different natural sensory organs like eyes, fingers, ears, nose, and tongue etc. which plays enormous role in our day to day learning related to the mundane task of vision, kinesthetic, sound, taste and smell and sixth sense called rational thinking. This kind of inherent parallel learning takes place from our birth to till our death, which transform us from unlearned person to learned person. This inherent learning helps us in critical thinking process to choose right (complex) decisions based on the knowledge that is acquired in our day to day life through learning. The human learning curve starts from his birth till death. The sensory receptors which continually receives data from the environment, learn and adapt to changes according to the result of learning. This inspiration of biological learning theory is adapted into machine learning using artificial neurons called ANN (Artificial Neural Network) [17]. Computer machines (Super computers with parallel processing) which process numerical data with the speed of millions of floating point operations per second [FLOPS], but lag in learning images due to the inefficient learning algorithms. Human Excel machines in learning, due to the complex rational thinking process. Historically, the decision support systems that had been developed over the years much resemble a rote kind of learning. The paradigm shift took place from the rote learning to the so-called machine learning in which machines can learn to find the best possible set of parameters that map the input into the output using the statistical analysis techniques. There are two types of learning called supervised and unsupervised learning. Supervised learning [15] is kind of machine learning where a label is associated with the data (called as teacher input) which is vital for learning. But for the unsupervised [16] kind of learning, one has to find the relationship among the training data without labels , which is more complex form of learning than supervised learning. Semi supervised learning [16] is a form of learning from both label and unlabeled data. There are other new learning methods such as reinforcement learning, which is used to learn actions that maximizes the cumulative reward or minimizes the penalty like games and automation.

Human Nervous System

The block diagram [Fig 1] of human nervous system is depicted in the figure 1.

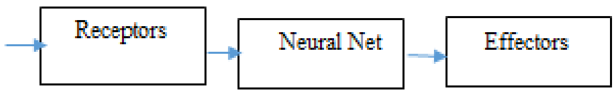

Figure 1: Block diagram of human nervous system

The nervous system of human being is centered around the brain which is the backbone of the biological neural network which takes input such vision, audio, smell, touch from the environment and apply learning and makes appropriate decisions to be carried out by the effectors. In brain, there are different size of anatomical elements that range from minuscule to large scale which take appropriate decisions through learning [17]. These anatomical elements are massively interconnected at various levels which is the unique characteristic of the brain. Molecules and ions are at the lowest level. Neural microcircuit which represent assembly of synapses (Synapses depends upon the molecules and ions for their action) which lie at the next levels are organized into connectivity patterns to perform function. The microcircuits form group into dendritic subunits. Local circuits are the next level hierarchy which made up of neurons with similar or different properties. These local circuits are responsible for the operations of localized region of the brain. Interregional circuits are made up of pathways, columns, and topographic maps, which involve multiple regions located in different parts of the brain. At last, these group of organized topographic maps are responsible for different sensory information such as visual, auditory, smelling, hearing and touching senses. The neurons in the brain are innumerable in numbers which are massively connected together which takes lower order day to day decisions into complex strategic higher order decisions.

B. Artificial Neural Networks

Artificial neural networks are non-symbolic AI technique which are modelled after biological neurons which gains momentum in the field of artificial intelligence. ANN is either hardware / software designed to learn and follow layered architecture like network architecture where in output of one layer is passed as input to subsequent layers. There exist three kind of layers namely input layer, hidden layer and output layer. Input layer is used to read the input and pass the input into hidden layers which extract features from the input and pass it into output layer. Each layer has innumerous neurons and these neurons are massively interconnected. The number of layers and the number of neurons in each layer are dependent upon the nature and complexity of the problem to be solved. Each layer neurons extract its own local / lower level features from the input parameters and pass it next higher layer which groups features into global /higher level features.

The best of Artificial neural network (ANN) is in representing conceptual model or biological inspired model through neural network architecture and learning algorithms. The simplest ANN architecture is composed of basic building blocks referred as neurons which are interconnected and interrelated to form network referred as Neural Network. The way in which it is connected give rise to the different forms of neural network architecture.

The learning algorithms which formed on learning rule such as delta rule, Hebbian rule, competitive learning rule, etc helps in determining the weight update of neurons. The learning algorithm is further classified to three types: firstly, supervised learning algorithm giving direct comparison between actual output and expected output. Secondly, unsupervised learning providing correlation between input data. Thirdly, Reinforcement learning algorithm which is special type to first form verifies whether actual output is correct or not.

The Convolution neural network is yet another type of neural network such as Feed forward neural network, Multilayer perceptron neural network, Radial basis neural network, recurrent neural network, Recursive neural network, Modular Neural Network, etc.

Among all these forms of neural networks Feed forward neural network, recurrent neural network and Convolution neural layer is most commonly used for image extraction and image classification. Feed forward neural network is a form of neural network where all the data traverse from network of input to network of output referred as feed-forward. The strength of this network lies in the fact that it can represent more complex functions very easily. However, these networks have a limitation of overfitting. Recurrent neural network which forms a directed graph and demonstrate the power of holding values for long temporal sequence because of its loop nature. But the limitation of this network is the vanishing gradient / exploding gradient. As a result, to overcome the above problems, convolution neural network was designed and further many modified models came into existences. This paper provides a broad overview on various types of convolution models. In Fig2, the taxonomy of AI to CNN is depicted.

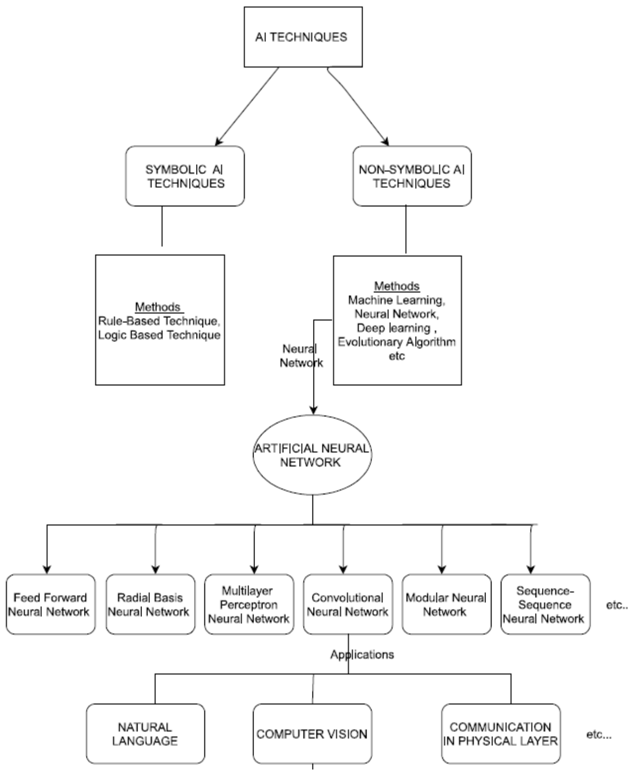

Figure 2.1: Taxonomy of Neural network

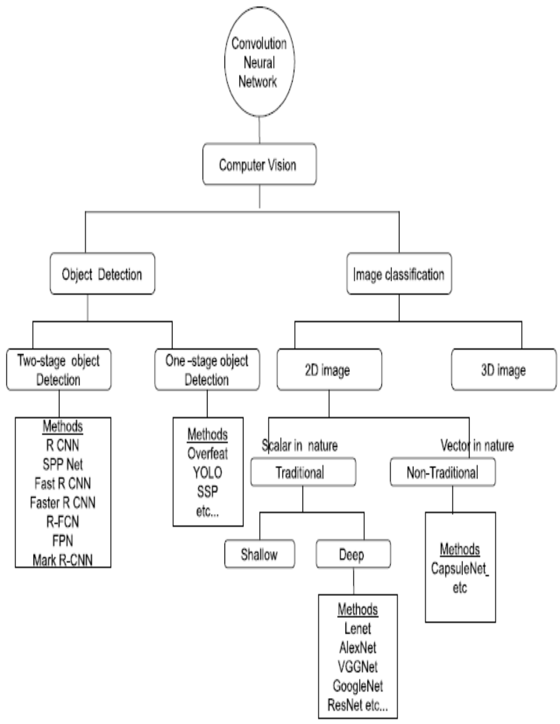

Figure 2.2: Taxonomy of CNN

May - June 2020
ISSN: 0193-4120 Page No. 3636-3644

CNN has sliding window function called convolution filter/ convolution kernel applied to sub block of the overall matrix representation of the data. For example, if the image data of size (100*100) pixels are taken as input then approximately 100 million parameters need to be trained in a fully connected deep neural network. Convolutional neural network reduces these parameters by zooming in on specific bits (sub block) of image data by each neuron of successive layer which aggregate the input into next higher-level features. Pooling layers in CNN are responsible for grouping lower level features into higher level features and also mitigate overfitting via subsampling procedure. Pooling also make CNN to recognize features independent of the location called the location invariant features. Another important aspect of CNN is the batch normalization process which is used to mitigate the vanishing or exploding gradients. Back propagating the errors to train the neurons to fine tune its parameters fail, if the gradients are vanishing or exploding. (The kind of neural network which propagates its gradient to update the weights of the parameters are called as Back Propagation Neural Network). Gradient may become zero from the initial MSE (Mean Square Error) and subsequently it stop updating the weights of the parameters (Loss value Stop Changing) and result in the output of the neuron without change is called as vanishing gradient. In some cases, gradient changes abruptly and explodes which may lead to slow convergence of the training, which is called as exploding gradients. In order to mitigate the problem of vanishing gradient or exploding gradients, batch normalization is applied which normalize input by means of scaling and shifting (multiply the input data by constant / add constant value to the input. CNN is further classified into traditional CNN and non-traditional CNN. LeNet, AlexNet, GoogleNet, ResNet, are examples of traditional convolutional neural networks and Capsule Net is an example for non-traditional convolutional neural networks. The taxonomy from AI to Deep CNN is depicted it the Fig 2.1 and Fig 2.2

2. Object Detection

To have a proper understanding of any given input as image, understanding of the concepts and location of object is each image becomes important [11]. This process which is referred as object detection, act as foundation for semantic /scene understanding of images and video/video frame in many computer vision applications such as image classification, human behavior analysis, face recognition, driverless vehicles, etc.

The biggest challenge in the object detection [12] is to identify the object of the location referred as object localization and also to identify the type of object referred as object classification. Object classification is categorized into two types: traditional and nontraditional object detection. The following table 1 & 2 further list and explain briefly the details associated with it.

Table 1: Traditional Object Detection

Traditional OD models	**Purpose**	**Methods**
Informative Region selection	Used to identify the position of object	Multi sliding window
Features classification	Used to identify objects and extract features which helps in semantic and robust representation	HOG, Haar-like
Classification	Used to differentiate between target object and other in an image.	SVM, Adaboost, DPMa

Table 2: Non-Traditional Object Detection

Non-Traditional models	**Purpose**	**Methods**
Two -stage proposal object detectors/classifier	It is improvised version of Traditional object detection models which used particular procedure or network to generate Region of interest (RoI)	R-CNN,SPPNet, Fast R-CNN, Faster R-CNN, FPN,Mark R-CNN, etc
One stage regression level object detectors/regressor	This model fine-tuned the limitations of two-stage object detectors which is carried out in two levels to one level to identify of object location. However, location determination of bounding box is matter of concern as a regression problem.	OverFeat, Yolo, SSD

May - June 2020
ISSN: 0193-4120 Page No. 3636-3644

3. Image Classification

The images received as an input for image classification are categorized to 2D images and 3D images. 2D image classification which is further categorized into traditional (shallow and deep) and non-traditional deep convolution neural networks.

The following table: 3 tries to provide all the categories in chronological order based on levels of version in comprehensive way.

Table 3: Comparison of traditional and non-traditional image classifiers

	Types	Description	Methods
Traditional 2D image classification	Shallow image classification	This architecture is same as that of basic ANN architecture since it is composed of basic building block referred as Neurons and network to each other with 3 types of layers. It has only one input layer (accepting multidimensional values) and few hidden layers and one output layer	
	Deep convolution image classification	This is the improved version on shallow convolution layer as it incurred poor performance with a smaller number of layers. To overcome this deep convolution architectures incorporated a greater number of hidden layer by overcoming overfitting problem /gradient dissipation problem	LeNet, AlexNet, VGGNet, DenseNet, GoogleNet, ResNet etc
Non-Traditional 2D image classification	Deep convolution image classification	Traditional Deep Convolution input received and processed in the form of scalar which in returns reduced training parameters and increased the performance by increasing feature extraction and learning process. However, it has weak Global information association capability as it focused more local information. This is overcome through by replacing scalar to vector input and output image representation. Where the magnitude length describes probability distribution and direction describes the properties of entities encapsulated in vector.	CapsuleNet: This type has completely different neuron structure, neural network and the way in which propagation and distribution of information is carried out between layers.

A digital camera cannot substitute for our eyes to understand the feelings or the message conveyed by the image. In order to understand images, first image need to be classified into particular category. Supervised or semi supervised learning method can be applied to classify the images based on the features of the images. Image classification can There are many different types of Image Classification methods such as pixel-wise, sub-pixel wise and object based wise. During 1980s and 1990s Pixel-wise image classification is the basic type of classification under which supervised, unsupervised and hybrid techniques [Zhang et al; 2005; Alajlan et al, 2012] were used and mostly implemented on homogeneous regions than heterogeneous regions. During 1990s as a solution to this fuzzy classification and spectral mixture analysis techniques [Adams et al; 1986; Wang,1990] and then sub-pixels-based analysis techniques [Adams et al.,1986; Roberts at al.,1998; Wu and Murray ,2003] were developed. During late 1990s object-based classification methods have been developed. There exist different image data sets for image understanding, Popular data sets are MNIST, IMAGENET, CIFAR.etc. In table 1, the comparison between traditional and non-traditional image classifiers is shown.

Image Representation

An image is represented digitally using a two-dimensional array of pixel intensity values. It is nothing but a collection of pixels with spectral information. A pixel is the smallest individual element of an image which cannot be further sub divided and is called picture element (pixel). A digital image is a two-dimensional discrete signal. Mathematically, such signals are represented as functions of two independent variables which represent brightness function of the two spatial coordinates of the pixel namely x and y coordinate. A monochrome digital image f(x,y) is a two dimensional array of luminance values.

May - June 2020
ISSN: 0193-4120 Page No. 3636-3644

$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \ldots & f(0,N-1) \\ f(1,0) & f(1,1) & \ldots & f(1,N-1) \\ .. & .. & \ldots & .. \\ f(M-1,0) & f(M-1,1) & \ldots & f(M-1,N-1) \end{bmatrix}$$

With $0 \le f(x,y) \le L$ and typically L=255. Each pixel stores the luminance of the pixel with 8 bit per pixel. A binary image has only two values 0/1 which represent the off/on state of the pixel.

$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \ldots & f(0,N-1) \\ f(1,0) & f(1,1) & \ldots & f(1,N-1) \\ .. & .. & \ldots & .. \\ f(M-1,0) & f(M-1,1) & \ldots & f(M-1,N-1) \end{bmatrix}$$

Where f(x,y) is either 0/1.

A Color digital image is represented by the RGB value of the pixel which store the color values of three primary colors namely red, green and blue component of the pixel. The individual component of color values are universally 8-bit values, resulting in a total of 3 bytes (24 bits) per pixel. This increase storage requirements of color images by threefold of the monochrome image. Naturally, there are different representation of storing image data, such as pixel interleaved (meshed) and color interleaved (planar) formats.

In the pixel interleaved format, every image pixel is represented by a list of three values.

4. Convolutional Neural Network

CNNs are special neural network for processing data which has grid like topology[10]. Example image data. Each CNN have three stage namely convolution, detector and pooling. In the first stage, it employs special kind of linear mathematical operation called convolution in place of matrix multiplication. It is an operation that applied to the arguments of two real valued function. It is mathematically denoted by s(t)=(x*w)(t), here * is called as convolution operator. x(t) and w(t) are input data and its weight function which is called as input and kernel of the convolution. The output is referred as feature map. It help to improve the sparse interactions, parameter sharing and equivariant representations of the deep learning system. Sparse representation largely reduces memory requirements. Parameter sharing supports rather than learning separate set of parameters for each location, learn only one set and also the same parameter is used in more than one function. Equivariant function means that if the input changes, the output changes in same way. In second stage, each linear activation is run through a non linear activation function such as ReLu (rectified linear activation). It is also called as detector stage. In the third stage, it uses pooling function which replaces the output of net at a certain location by the summary statistic of the nearby outputs. Pooling helps to make representation invariant of the small translations in the input.

Types of CNN

There are two types of CNN called traditional CNN and non-traditional CNN.

Traditional CNN

In this paper, the four major CNN are analyzed such as LeNetAlexNet, GoogleNet and ResNet using neuron structure, network structure, the way of propagation and distribution between layers, number of layers, number of convolutional and pooling operations, stride and number of learning parameters to be trained.

A. LeNet

Lenet is the convolutional neural network for hand written digit recognition task, cheque processing in OCR and document recognition. It was developed during the 1998 by Yann Lecun and his team [1]. It is a multi-layered neural network with back propagation and uses gradient descent learning. The following diagram depicts the architecture of LeNet-5. It comprises of seven layers, all of which has trainable weighted inputs. The input image is of 32*32 pixels. It uses three convolutional layers named C1, C3 and C5 and sub sampling layers named S2 and S4. C1 is a convolutional layer having six feature maps. Each unit in the feature map is associated with 5*5 neighborhood. The size of the feature maps is 28*28. C1 has 156 trainable weights and 122,304 connections.

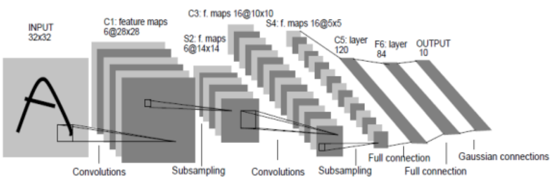

Figure 3: LeNet CNN architecture

Layer S2 is sub sampling layer of size 14*14 with 6 feature maps. Each unit in the feature map is associated with 2*2 neighborhood in Corresponding feature of C1. In S2, the inputs are applied sigmoidal function. Layer S2 has 12 trainable weights and 5,880. C3 is a convolutional layer having sixteen feature maps. Each unit in the feature map is associated with 5*5 neighborhood at corresponding S2 feature maps. C3 has 1516 trainable weights and 151,600 connections. Layer S4 is sub sampling layer of size 5*5 with 16 feature maps. Each unit in the feature map is associated with 2*2 neighborhood in Corresponding feature of C3. Layer S4 has 32 trainable weights and 2,000. C5 is a convolutional layer having 120 feature maps. Each unit in the feature map is associated with 5*5 neighborhood at corresponding S4 feature maps. C3 has 48,120 trainable weights and 151,600 connections. Layer F6 has 84 units and fully connected to C5. It has 10,164 trainable weights. The weighted sum of F6 is applied sigmoid squashing function and then hyperbolic tangent is computed. Finally output layer contain Radial basis function units, for each class of digits for 84 different character recognition. Fig 3, depicts the architecture of Lent

B. AlexNet

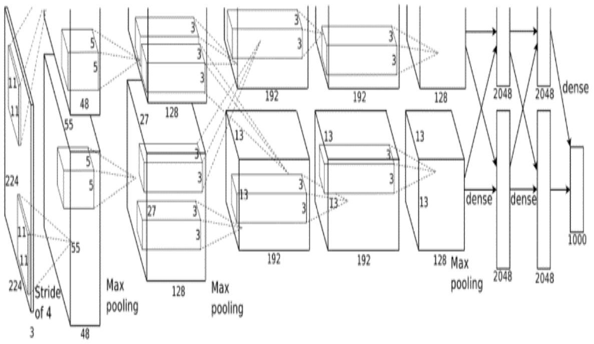

Figure 4: Alexnet CNN architecture

Alexnet is the convolutional neural network which own the IMAGENET dataset challenge ILSVRC (Imagenet Large Scale Visual Recognition Challenge) for image recognition of 1.2 million images labelled with 1000 classes. It was developed during the 2012 by Alex Krizhevsky and his team [1]. It is a GPU (Parallel computer) based multi layered neural network with 60 million parameters and 650,000 neurons. The above diagram (Fig:6) depicts the architecture of Alexnet. It has 8 learned layers and comprises of 5 convolutional layers and 3 fully connected layers. The input image is of 224*224 pixels of RGB channels. The AlexNet unique features are ReLu (Rectified Linear Units), Training on Multiple GPUs, Local Response Normalization and overlapped pooling. The first Convolutional layer filters the 224*224*3 image with 96 kernels of 11*11*3 with a stride of 4 pixels (distance between the adjacent block view). The output of first convolution layer is given as input to the second convolutional layer (where input is normalized and pooled) and filters with 256 kernels of size 5*5*48. The 3,4 and 5 convolutional layers are connected with one another without normalization and pooling. The third convolutional layer has 384 kernels of size 3*3*256 connected to the output of second convolutional layer. The fourth convolutional layer comprises of 384 kernels of size 3*3*192.The fully connected layer consists of 4096 neurons each. The overfitting was eliminated using data augmentation and dropout. It uses stochastic gradient descent learning algorithm. Fig 4, depicts the architecture of AlexNet

Figure 5: Googlenet CNN architecture

C. GoogleNet

GoogleNet is the convolutional neural network which own the IMAGENET dataset challenge ILSVRC14 (Imagenet Large Scale Visual Recognition Challenge) for image recognition of very large images labelled with 1000 classes. It was developed during the 2014 by Christian Szegedy and his team [1]. It has 22 layers. Its intuitive concept is inception architecture which allows increasing the number of units at each state without affecting computational complexity. Also, it adds extra linear layer to adapt and fine tune neural network for other labels. It is designed CPU based but could be trained on high end GPUs. It suggests to move to more promising sparse architectures in future.

D. ResNet

In order to address the problem of degradation in deep neural networks, ResNet was introduced by Kaiming He and his team during the year 2015. It has 152 layers which is deepest CNN for IMAGENET challenge. It has both plain and residual network. A residual network is a one which uses short cut connections which act as highway networks and performs identity function. The outputs of identity function are added to stacked layers which neither add extra parameters nor computation

TEST Engineering & Management

May - June 2020
ISSN: 0193-4120 Page No. 3636-3644

complexity. It is very easy to optimize and reduces error rate. Fig 6, depicts the architecture of ResNet

Non-Traditional CNN

Capsule net

This type has completely different neural network structure and it differs the way in which propagation and distribution of information is carried out between layers [14].

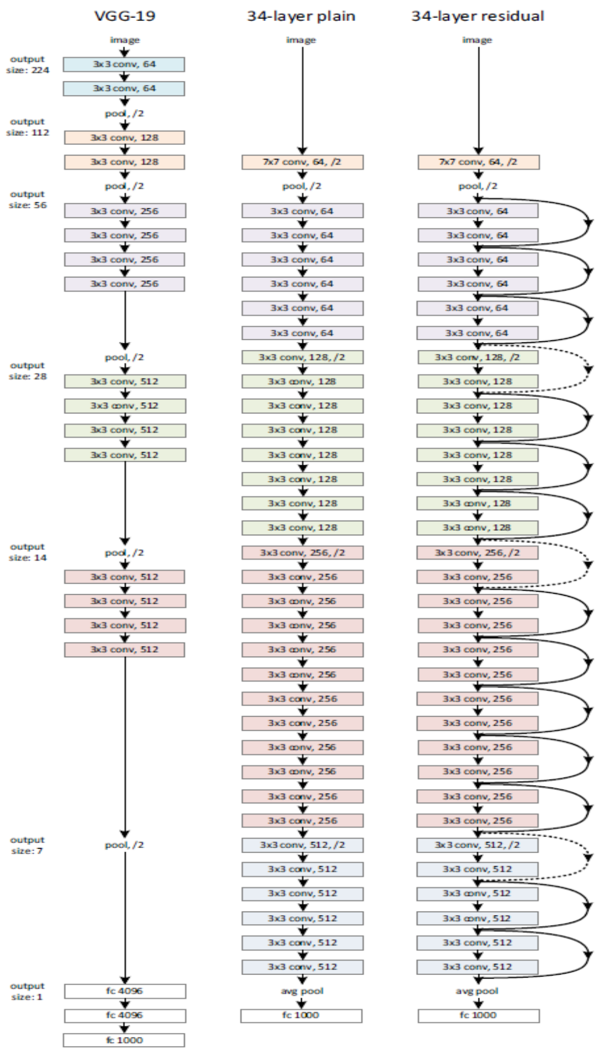

Figure 6: ResNet CNN architecture

5. Conclusion

After the AlexNet, many other neural Network particularly tried to prove that, the number of layers improve the error rate. However, after certain extent as the layers goes on adding which incur with gradient dissipation problem and performance starts either degrading or decline.

This limitation was addressed using other form of Classical Net called ResNet which not only tried to solve gradient dissipation problem with decrease in parameters but also boosted the performance of Neural Network with 152 deep layers and with top-5 error rate to 3.57%.

The limitation of classical CNN is overcome through Capsule Net which changed the learning model completely by considering vector inputs instead of scalar input. In vector input (encapsulation of entity), the length of the vectors is used to describe the probability distribution and direction of vectors represents the property of corresponding entities.

References

[1] Artificial Intelligence by Eliane Rich, Keven Nnight and Shivashankar B Nair, 3rd Edition, McGrawHill Publication

[2] Neha Sharma, Vibhor Jain, Anju Mishra, "An Analysis of CNN for Image Classification, 2018 ICCIDS Elsevier

[3] Karen Simonyen, Andrew Zisserman, 2015 ICLR,Very Deep CNN for Large Scale Image Recognition

[4] Christian Szegedy, Wei liu, Yangqing Jia, Pierre Sermanet, Scoot Reed, Dragomir Anguelov, Dumitru Ergan, Vincent Vanhoucke, Andrew Rabinovich, Going Deeper with convolutions

[5] Miao Li, Shuying Zang, Bing Zhang, Shanshan Li & Changshan Vu, European Journal of Remote Sensing, 2017, A Review of Remote sensing Image Classification techniques: The Role of Spatio-Contextual Information

[6] Alex Krizhevsky One Weird trick for parallelizing convolutional neural networks

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Microsoft Research, Deep Residual Learning for Image Recognition

[8] Yann Lecun, Leon Bottou, Yoshua Bengio and Patrick Haffner, Proceeding of IEEE, 1998, Gradient-based Learning applied to document recognition

[9] O'Shea T, Hoydis J. An introduction to deep learning for the physical layer. IEEE Transactions on Cognitive Communications and Networking. 2017 Dec;3(4):563-75.

[10] Ding S, Li H, Su C, Yu J, Jin F. Evolutionary artificial neural networks: a review. Artificial Intelligence Review. 2013 Mar 1;39(3):251-60.

[11] Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: A review. IEEE transactions on neural networks and learning systems. 2019 Jan 28;30(11):3212-32.

[12] Qian H, Xu J, Zhou J. Object Detection Using Deep Convolutional Neural Networks. In2018 Chinese Automation Congress (CAC) (pp. 1151-1156). IEEE.

[13] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems 2012 (pp. 1097-1105).

[14] Yang F, Li W, Tang W, Wu X. The Analysis Between Traditional Convolution Neural Network and CapsuleNet. In2018 International Conference on Control, Automation and Information Sciences (ICCAIS) 2018 Oct 24 (pp. 210-215). IEEE.
[15] Murphy, Machine Learning
[16] Ian Goodfellow, yoshuva Bengio, and Aaron Courville, MIT press, Deep Learning
[17] Simon Haykin, 3rd Edition, Pearson, Neural Networks and Learning Machines

Authors Profile

Vijayalaxmi P Chiniwar, MCA is currently working as Assistant Professor, School of CSA, REVA UNIVERSITY. She is interested in doing research in the area of Artificial Intelligence,
E-mail: chiniwarvijaya@reva.edu.in

Mohamed Abdul Kader Jailani N M.Tech,MCA Assistant Professor, School of CSA, REVA UNIVERSITY, Bengaluru, India. He is interested in doing research in image processing, machine and deep learning.
Email: mohamadjilani@reva.edu.in