International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 96-103

Prospects of Convolutional Neural Networks: A study

Jeevan G V[1], Mamatha G[2]
[1]*(Student, Department of ISE, JSS Academy of Technical Education, Bangalore, India)*
[2]*(Assistant Professor, Department of ISE, JSS Academy of Technical Education, Bangalore, India)*

Abstract- *An essential network in Machine Learning is the Convolutional Neural Network. In the past few decades, CNN has drawn the attention of both industry and academia as CNN has achieved impressive results in quite a lot of area such as, Emotion recognition in text, speech, face, Computer Vision (CV), Image processing, Natural Language Processing (NLP), etc. Some reviews examine CNN from common perception. While existing research highlight CNN's purpose in a variety of circumstances, there is no thoughtfulness of CNN as a complete. Through our study, we wanted to present perspective of CNN on its fast-growth right from one-dimension to multidimensional convolutions, summarize on various methods of convolution, CNN's applications in various fields and a quick look into CNN generation-Next.*

Keywords- *Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Neural Networks, Convolutional Neural Networks (CNNs).*

I. Introduction

A Neural Network (NN) is a metaphor for the Computer-based system that mimics the biological neural networks ofthe Brain. Neural network also known as neural net tries to find relationships in data through algorithms that mimic brain functions.

Convolutional Neural Networks (CNN) has attained outstanding achievements as problem solving using AI/ML and Deep Learning uses CNN. Through Computer Vision, CNN technology has facilitated people to accomplish things which were formerly unimaginable such as, human interaction with robots, unmanned vehicles, virtual assistants, etc. our study provides an overview of classic models, applications, and discusses some possible uses of CNN. By doing so, we give readers an understanding of how modern CNNs can provide efficient service. Normally, CNNs are utilised to categorize images, in recognize tumours in scanned images, Facial emotion and speech emotion recognition, etc. moreover, CNNs are good at mainstream tasks like signal processing, image segmentation, etc.

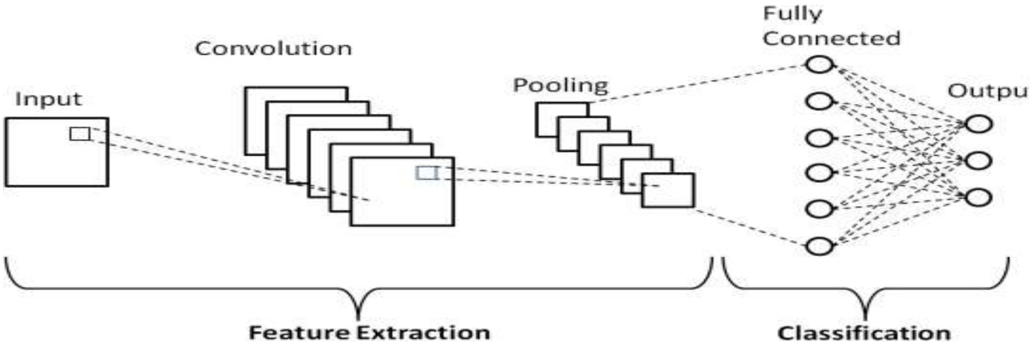


Fig.1 General Structure of Convolutional Neural Network (CNN)

II. Role of Convolutional Layer

The input image, as shown in Fig. 2 is an example of a possible Input. Input can be an image, (image from CIFAR-10 with 32x32 pixels and 3 channels for RGB) or a Video, a greyscale video where the height and width are frame figures and the depth is same as shown in fig. 3, or even an experimental video with wide and high sensors whichhave multiple depths. Assuming raw pixels are the input to the network, then the output layer should have 32*32*3 weights connecting to CIFAR-10 in the Multi-Layer perceptron.

The image in Fig. 4 shows the actions that happen when we manually adjust the connection weight within 33 windows to express the consequence of the convolution matrix, which works similarly to traditional image processing filters. TheCNN initializes the shape filters; followed by training method shape filters that are suited to the job at hand. In order to make this approach more helpful, further layers can be added following the input layer.

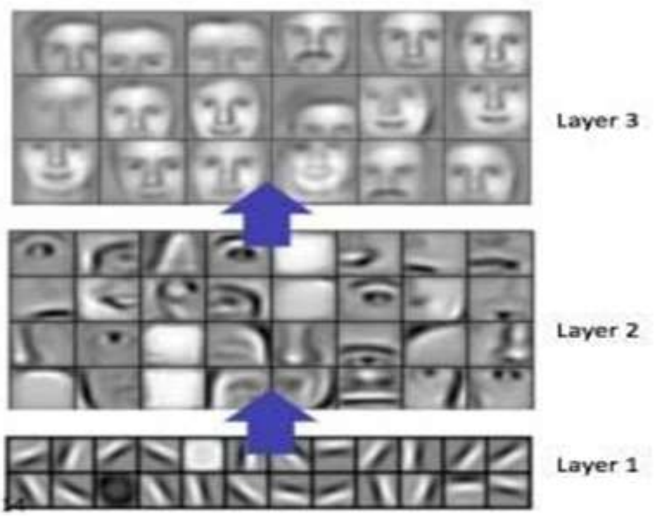


Fig 2: Input image set to Convolution Layers

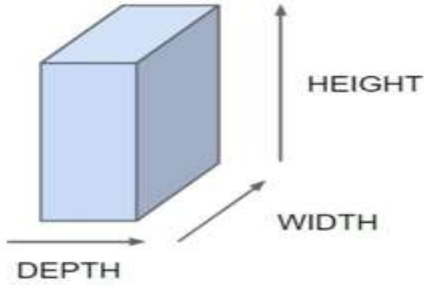
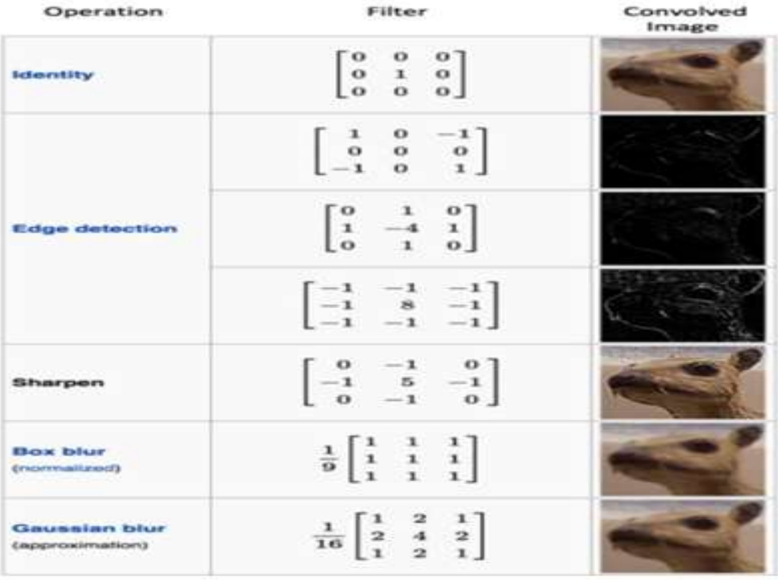


Fig. 3 Three-dimensional representation ofInput data for CNN

Diverse filters can be allocated to every layer, resulting in dissimilar characteristics being extracted from the generated image. Fig. 2 illustrates the connections among the different layers. every layer has its own filter, which results in different characteristics being extracted from the image.



Operation	Filter	Convolved Image
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Edge detection	$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$	
	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	
Gaussian blur (approximation)	$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	

Fig .4 Effect of different Filters on Image

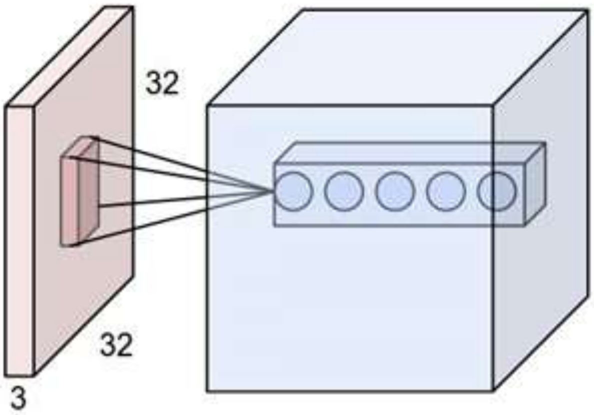
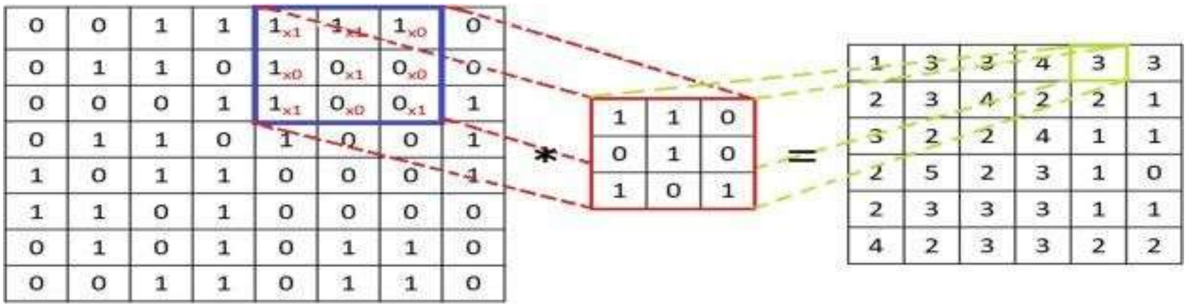


Fig. 5 Multiple layers corresponding to unrelatedfilter

In convolutional layers, a set of filters are combined with the input volume, each filter has its own set of parameters, The result is a neuron-based activation map, when they are applied to the inputs, the filter slide across the input's width and height, and the dot product between the input and filter is computed at each location. That means, each neuron's receptive field is thin and equivalent to the size of the filter because the Height and Width of every filter are less than the input. Convolutional layers produce their output by stacking the activation maps of all filters along the depth dimension. It is believed that local connection is motivated by the receptive fields of neurons that are as narrow as in the animal visual cortex. A CNN layer with local connectivity permits the network to guide filters that maximize their response to a section of input, allowing it to utilize the input's spatial correlation (image pixels are correlated to their neighbouring pixels than non-neighbouring pixels).Moreover, by convolutioning the filter parameters with the input, the activation map is generated, which reduces the number of parameters required for optimal learning, expressing, and generalization.

Kernel filters may have the same dimension as the input images, as but with fewer constant parameters thanthe convolutional layer, which extracts the fundamental features of input images. If one wanted to compute a 35 x 35 x 2, 2D scalogram image, the filter size would be f x f x 2, where f can be 3, 5, 7, 9 and so on. It must, however, be smaller than that of input image. Filter mask moves throughout the entire input image iteratively, assessing the kernel filter weights and pixel values of input image, the result of this is a 2D activation map; using this, CNN will recognize the visual features. Fig. 6 shows an example of how to calculate a 2D activation map.



0	0	1	1	$1_{\times1}$	$1_{\times1}$	$1_{\times0}$	0
0	1	1	0	$1_{\times0}$	$0_{\times1}$	$0_{\times0}$	0
0	0	0	1	$1_{\times1}$	$0_{\times0}$	$0_{\times1}$	1
0	1	1	0	1	0	0	1
1	0	1	1	0	0	0	1
1	1	0	1	0	0	0	0
0	1	0	1	0	1	1	0
0	0	1	1	0	1	1	0

1	1	0
0	1	0
1	0	1

1	3	3	4	3	3
2	3	4	2	2	1
3	2	2	4	1	1
2	5	2	3	1	0
2	3	3	3	1	1
4	2	3	3	2	2

Fig. 6 Computing Activation map

III. Pooling

Pooling is a way to reduce the difficulty of succeeding layers by down-sampling, in context of image processing; it is equivalent to reduction of the resolution. Pooling does not influence the quantity of filters. Maxpooling partition the image input into smaller rectangles and returns only the greatest value of every sub-region, as shown in Fig. 7. Pooling is based on stride-2, but stride-1 is available to avoid down-sampling, which is rare. Down-sampling do not retain the location of the information; hence, its use is restricted in situations where existence of information is crucial (than spatial). To increase efficiency Pooling can be used with unequal filters and strides.

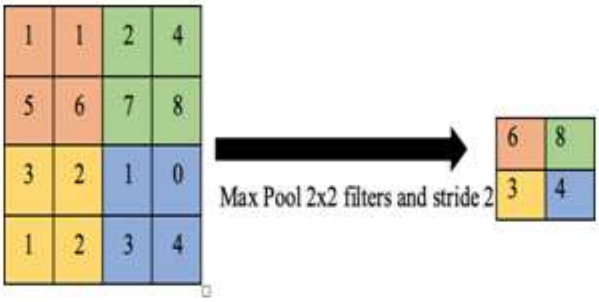


Fig. 7 Max-pooling for Down-sampling

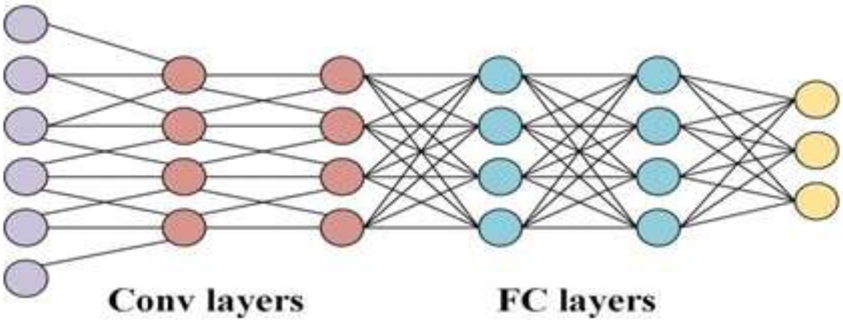


Fig. 8 Interconnection of Convolution layers and Fully Connectedlayers

IV. Several Convolutions

CNNs use convolutional structures in order to extract features from data. Unlike classical feature extraction techniques, CNNs put an end to the need for manual feature extraction. CNN kernels in essence represent diverse receptors that can handle various kinds of information; their activation functions simulate neural electrical impulses that precede the next neuron.

CNNs have advantages and benefits from Fully Connected network (FC), interconnection are as shown in fig. 8,

1) Local connections - every neuron need not be linked to preceding layer neurons, so parameters get reduced and convergence speed up
2) Weight sharing - the same weight can be shared across links, further reducing parameters.
3) Using down-sampling, dimension reduction can be achieved by reducing minor features, as well as reducing the number of parameters; a pooling layer will be able to retrieve.

Deformable Convolution network

The typical grid sampling points in a standard convolution are offset by 2D in deformable convolutions, as shown infig. 9. It allows the sampling grid to be deformed in any way it wants. The offsets of the last feature map are learned using additional CNN layers.

Group Convolution

Function of a Grouped Convolution is to unite a number of convolutions into a single layer; resultant is a many channel outputs per layer. This so formed larger network assists design network in learning a different range of properties from basic to high level. Example, The objective of using Grouped Convolutions in AlexNet is to stretch the model across several GPUs as a cost effective measure. on the other hand, it was demonstrated that this module may be utilised to improve classification accuracy with models such as ResNeXt. Accuracy can be improved by expanding cardinality byexposing a new dimension through GC.

Steerable Convolutional Networks

To improve the network's resistance to data geometry transformations and to decrease over fitting, Steerable CNN applies prior knowledge of transformation invariance or equi-variance. Equivariant networks are preferred above invariant networks. When images are rotated in one-dimension space, the invariant network will no longer perform accurate detection on anomalous faces, for example abstract paintings are recognized as real positive samples. As an outcome, equi-variance is required to predict the linear transformation of input images. Filters can react to changes of both position and pose, just as a usual CNN

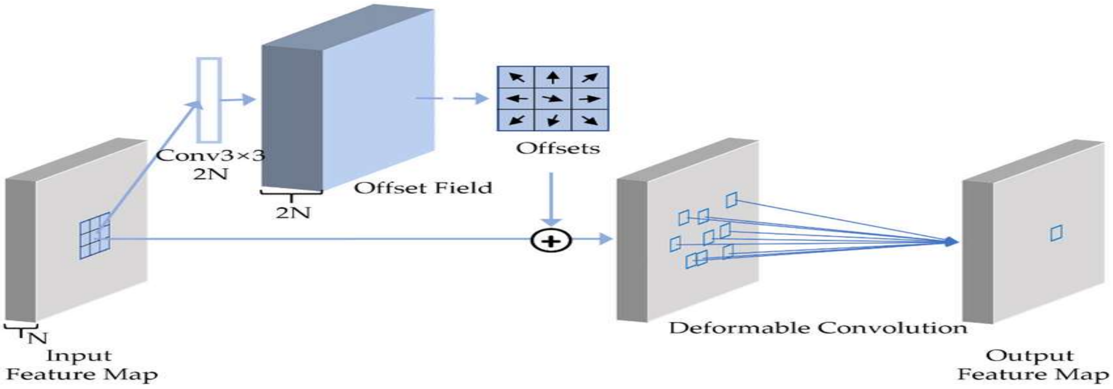


Fig. 9 Deformable convolution network

Graph Convolutional Networks (GCNs)

There is a semi-supervised learning strategy called Graph Convolutional Networks (GCNs) that works on graph structures. These networks are built as a variation of CNN that works directly with graphs. The neural networks are motivated by an estimated first-order estimate of spectral graph convolutions; the model represents both local graph structure and node attributes. GCN scales linearly with graph edges.

V. CNNS Popular Architectures

LeNet-5 architecture

Seven layers are present in LeNet-5 architecture, three are convolutional layers, two are sub-sampling layers, and another two are fully linked layers. Architecture of LeNet-5 is as shown in fig. 10, Input is the first layer.

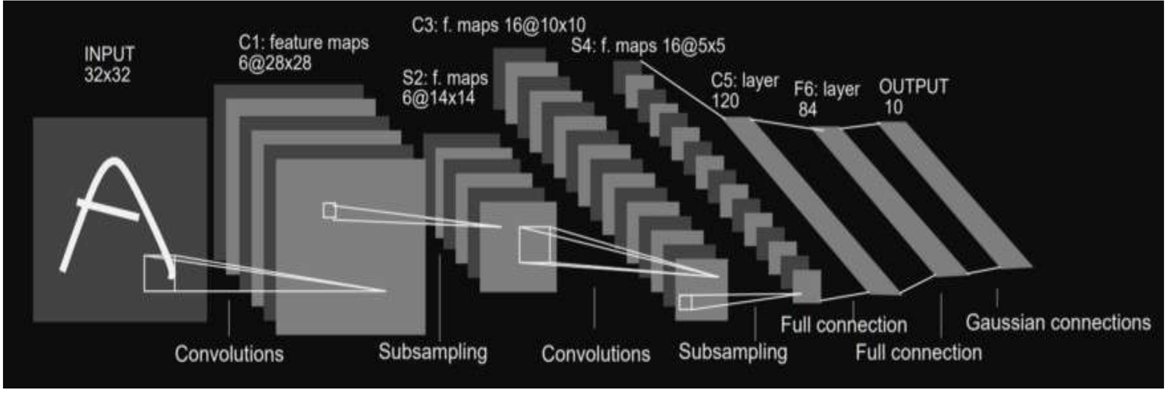
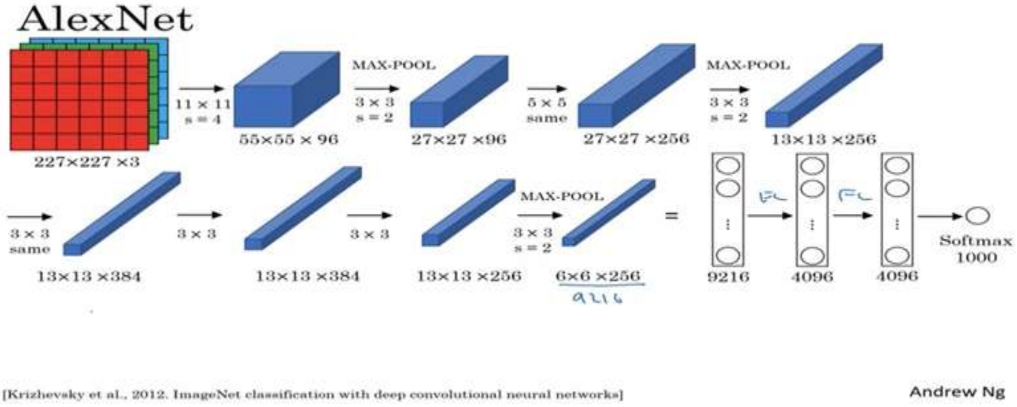


Fig.10 LeNet-5 architecture

Input images should be as large as 32x32 pixels, since this is the size of the images that will be transferred tosubsequent layers. In MNIST dataset, images are 28x28 pixels large, it is necessary to pad the 28x28 images to make them comply with the input layer's requirements. Normalizing the gray scale images between -0.1 and 1.175 willreduce training time by ensuring a Mean of 0 and a SD (standard deviation) of 1. Pixel values of the photos are normalised at the time of image classification in LeNet-5.

Convolutional layers and subsampling layers are the 2 types of layer constructs used in the design of LeNet-5. An example of the layer identification approach can be seen in the fig. 10, which shows convolutional layers andsubsampling layers are denoted by the letters "Cx" and "Sx," respectively. The letter "Fx" indicates fully connected layers. The kernel/ filter are the windows containing the weight value that are used to convolution the weight valuewith the input values. Each unit within a convolutional layer has a local receptive field of size 5*5. First convolution layer C1 provides 6 feature maps as output. Sub-sampling layer that follows the first layer also generates six feature maps, each with a dimension that corresponds to the input feature maps from the previous layer; this down-sampling layer is numbered as 'S2'.

- **AlexNet** architecture



[Krizhevsky et al., 2012. ImageNet classification with deep convolutional neural networks]

Andrew Ng

Fig.11 AlexNet architecture

AlexNet, a CNN network with eight layers, has won the ImageNet Large Scale Visual Recognition Challenge in 2012. Other networks demonstrated manually designed features which were not as effective as learning-based features, for breaking the computer vision paradigm.

Similar design ideas and architectures are shared by AlexNet and LeNet, however they still have significant variances. ReLU is used as activation function in AlexNet, sigmoid, is much smaller than LeNet5. AlexNet has eight layers. The first 5 are convolution layers, the second two are hidden layers and the third is the output layer. Fig. 11 shows AlexNet architecture, 1^{st} layer has a convolution window of 11*11, since most ImageNet images are greater than 10 times (height x width) than images of MNIST; therefore a bigger convolution window is needed to hold each object.In the 2^{nd} and subsequent layers the convolutional window is reduced to 5*5 and to 3*3, The network includes large number of pooling layers with windows of size 3*3 and a stride of 2 after 1^{st} , 2^{nd}, and 5^{th} convolution layers, AlexNet have as high as 10 times more number of convolution channels as LeNet, following the final convolution layer are two number of fully-connected layers with outputs of 4096. Based on early GPUs with limited memory, AlexNet used split data stream architecture, so that each GPU could store in and compute half of the model only. These days GPU comes with relatively large these days, hence we rarely have to divide models crosswise among GPUs.

VI. Dataset Collection

RMAF accuracy percentage for Datasets, when used with different CNN models is as shown in table 1.

The RMAF (ReLU-Memristor-like Activation Function) is a activation function used in NN (Neural n/w) to make use of -ve values. For making AF smooth, non-monotonous, and non-linear; RMAF adds to the function, a constant parameter and a threshold parameter this results in RMAF outperforming Rectified Linear Unit (ReLU) and other AFs on datasets and complex models.

Models of CNN	**Dataset**	**Activation function-RMAF(Accuracy %)**
ResNet 50 AlexNet DenseNet 121 SqueezeNet	ImageNet	87.60 80.02 86.25 85.37
ResNet 50 AlexNet DenseNet 121 SqueezeNet	MNIST	99.73 92.28 98.58 97.64
ResNet 50 AlexNet DenseNet 121 SqueezeNet	CIFAR-100	79.82 69.97 75.58 68.78

ResNet 50	CIFAR-10	98.77
AlexNet		84.58
DenseNet 121		79.82
SqueezeNet		87.31

Table 1. RMAF of CNN models for datasets

VII. Applications of CNNS

In Deep Learning, CNN is considered as one of the most essential ideas. During the reign of Big-data, CNN can access large quantities of data to provide promising results, in contrast with the previously used methods. This has led to a plethora of applications of CNN, along with the processing of 1-Dimension and 2-Dimension pictures.

1-D CNN areas of Application

Compact 1-D CNN is preferred over the deep 2-D equivalents, in a number of applications due to their inherent ability to combine feature-extraction with categorization. Some of the application that uses 1-D CNN for their design are listedbelow

- **Automatic speech recognition**

during 20th century, Automated Speech Recognition (ASR) has strived to perform real-time and accurate translation of human speech into (written) text, The main purpose of ASR is to translate Speech into Text on a one-to-one basis, logically, this is considered as a complicated process, because individual's speech signals differ widely among speaker, and conditions become considerably more hard when external noise or differences in speech styles are present (e.g. dialect of the similar language).

- **ECG monitoring in real-time**

Electrocardiogram (ECG) is frequently used to monitor and measure health of our Heart; cardiac abnormalities are identified and diagnosed based on the electrical activity generated by the heart and the same is collected during the ECG test.

- **Vibration-based structural damage detection in Civil structure**

Civil engineering structures need to be monitored for structural degradation to ensure their long-term health, service ability, integrity, and safety. Monitoring of the damage occurrence and spread, greatly affects the capability of a structure to perform effectively.

2-D CNN areas of Application

Unlike previous classification algorithms and in comparison with 1-D CNN, 2-D CNN does not demand to extract features. In addition, it is possible to tune 2-D CNN with large databases, thus increasing accuracy and resiliency.

- **Image Classification**

The categorization of images is known as Image classification, CNN is credited as the pioneer program to categorize digits written by hand; AlexNet lead the way for CNN-based classifications. As a result of emphasis on depth in CNNs,VGGNets and GoogleNet are some of the deeper network architectures with greatly improved classification accuracy.

- **Object Detection**

Object detection involves classification of images; in addition to that, system must define a bounding box around images. Examples of one-stage algorithms are YOLO, CornerNet and SSD, examples of two-stage methods are faster R-CNN, fast R-CNN and R-CNN

- **Image Segmentation**

The course of dividing an image into different segment is known as Picture segmentation, the notion of fully convolutional networks applies CNN architecture to segment images in pictures.

Review of few literatures using CNN for Facial Emotion Recognition

Akriti [3] they used Deep Learning open library "Keras" made available by Google for Face Emotion Recognition, CNN is used for image classification, they trained the CNN network with two datasets and evaluated the validation and loss accuracy.

- Saad [2] major issue discussed in this paper is related to Convolutional Neural Network, and how each of the parameters affects the network's performance. Biggest layer in the NN is the convolution layer, which consumes most of the network's resources.
- Pranav [4] their model provides a two-layer convolution network that classifies five different facial emotions. The model shows a similar training and validation accuracy, this demonstrates that model is a good fit with data.
- Dong [5] they proposed a method that combines pre-processing based on CNN and MFVT to generate HR NA images by improving clearness and resolution of WA input images before view-transformation process. The MFVT technique enhances the description of NA image by balancing the deficiency of pixel density of the WA images at the outer edges.
- Milad [6] Based on the JAFFE database, the system has analyzed 213 images of Japanese girls. Dataset includes 7- facial emotion expressions, a normal face and six ways of expressing emotion. The proposed approach has shown 86% of accuracy, by using conventional CNN. following 25 epochs, the [6] method attains 97% accuracy and the opponent have achieved only 90% of accuracy.
- Zewen [1] they discussed on some hardware implementation schemes for CNN
- Bochen [10] they proposed and implemented a random deep neural network based algorithm for extracting image features.

VIII. Conclusion

The objective of this paper is to provide a broad outline of CNNs, together with their rationale, few interesting convolutions, classic networks, related functions, real world application, and potential prospects. Convolutional is the most essential layer in CNN, since it contributes for most of the network's processing power. Number of stages in the network has great impact on its performance, when the number of layers increases; subsequently the amount of training and testing time too increases. In today's world, CNN is considered as a means of improving Machine Learning for wide range of applications, such as face recognition, man-machine interaction, virtual reality, etc. Different dimensions of convolutions must be constructed for diverse problems. Prospects of CNN and its variants are high in the near future,and in this regard a continuous study on CNN is essential.

REFERENCES

[1] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2021.3084827.

[2] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[3] A. Jaiswal, A. Krishnama Raju and S. Deb, "Facial Emotion Detection Using Deep Learning," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154121.

[4] E. Pranav, S. Kamal, C. Satheesh Chandran and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 317-320, doi: 10.1109/ICACCS48705.2020.9074302.

[5] D. Y. Choi, J. H. Choi, J. W. Choi and B. C. Song, "CNN-based pre-processing and multi-frame-based view transformation for fisheye camera- based AVM system," 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 4073-4077, doi: 10.1109/ICIP.2017.8297048.

[6] M. M. Taghi Zadeh, M. Imani and B. Majidi, "Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters," 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), 2019, pp. 577-581, doi: 10.1109/KBEI.2019.8734943.

[7] J. -M. Guo, P. -C. Huang and L. -Y. Chang, "A Hybrid Facial Expression Recognition System Based on Recurrent Neural Network," 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2019, pp. 1-8, doi: 10.1109/AVSS.2019.8909888.

[8] Sabrina Begaj, Ali Osman Topal, Maaruf Ali, “Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network(CNN)”, 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)

[9] S. Liu, D. Li, Q. Gao and Y. Song, "Facial Emotion Recognition Based on CNN," 2020 Chinese Automation Congress (CAC), 2020, pp. 398-403,doi: 10.1109/CAC51589.2020.9327432.

[10] B. Yang, "Image Feature Extraction Algorithm based on Random Deep Neural Network," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 863-867, doi: 10.1109/ICICV50876.2021.9388588.