IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 1, October 2022

Study of Convolutional Neural Network

Avinash H. Hedaoo
Assistant Professor, Department of Computer Science
Prerna College of Commerce, Nagpur, Maharashtra, India

Abstract: *The success of traditional methods for solving computer vision problems heavily depends on the feature extraction process. But Convolutional Neural Networks (CNN)have provided an alternative for automatically learning the domain specific features. Now every problem in the broader domain of computer vision is re-examined from the perspective of this new methodology. In deep learning, Convolutional Neural Networks (CNN) are found to give the most accurate results in solving real world problems such as several engineering fields, including array processing, wireless communications, medical signal processing, speech processing and biomedical engineering. The availability of a large amount of data and improvement in the hardware technology has accelerated the research in CNNs. This paper intends to provide a detail study over basic concepts of CNN and its applications.*

Keywords: Convolutional Neural Networks, Machine Learning, computer vision, deep leaning, Machine learning

I. INTRODUCTION

The term artificial intelligence (AI) was first coined by John McCarthy in 1956[2].In recent years, AI based applications have rapidly been developed in all fields[3].Machine learning is defined by TM Mitchell:"Machine learning is the study of computer algorithms that allow computer programs to automatically improve through experience."[4].Deep learning algorithms are a subset of Machine Learning algorithms [5][1].

With the recent advancement in digital technologies, the size of data sets has become too large in which traditional data processing and machine learning techniques are not able to cope with effectively [10,11]. However, analyzing complex, high dimensional, and noise-contaminated data sets is a huge challenge [12-14]. To undertake these problems, in recent years, various deep architectures with different learning paradigm are quickly introduced to develop machines that can perform similar to human or even better [15][9].

A convolutional neural network, symbolled as CNN, or ConvNet, most commonly applied to analyse visual imagery [16]. Such networks are characterised by being multi-layered, and so they are large and they are consequently complex. [17]. These networks are also called "shift invariant" type of artificial neural networks and symbolled as SIANN [18][19].

The concept of neural network has already existed since 1950s whenFrank Rosenblatt created the perceptron. Even convolutional neural network itself is not a new conceptat all [20].In 1959, Hubel & Wiesel [28] found that cells in animal visual cortex are responsible for detecting light in receptive fields. Inspired by this discovery, Kunihiko Fukushima proposed the neocognitron in 1980 [29], which could be regarded as the predecessor of CNN. In 1990, LeCun et al . [30] published the seminal paper establishing the modern framework of CNN, and later improved it in [31]. Like other neural networks, LeNet-5 has multiple layers and can be trained with the back propagation algorithm [32][27].The convolutional neural network was firstly introduced in [21] to recognize handwritten ZIP code in1989, and later extended to recognition and classification of various objects such as hand-written digits (MNIST) [22],house numbers [23], traffic signs [24], Caltech-101 [25] and more recently 1000-category ImageNet dataset [26][20].Initially CNN had been widely used for object recognition tasks but now it is being examined in other domains as well like object tracking [48], pose estimation [49], text detection and recognition [50], visual saliency detection [51], action recognition [52], scene labelling [53] and many more [54].[47]

CNNs saw extensive use in the 90s of 20th century, butfell out of fashion with the emergence of SVM and Bayesian models. One important reason is, small datasets in 1990s and early 2000s such has MNISTandCaltech-101[20].However with the availability of big data and advancements in hardware are the main reasons for the recent

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 1, October 2022

success of deep CNNs. Recently, it is shown that different levels of features, including both low and high-level, can be transferred to a generic recognition task by exploiting the concept of Transfer Learning (TL)[73][74][75]. Different architectural designs such as Wide ResNet, ResNeXt, Pyramidal Net, Xception, PolyNet, and many others explore the effect of multilevel transformations on CNNs learning capacity by introducing cardinality or increasing the width [76][77][78][79]. Therefore, the focus of research shifted from parameter optimization and connections readjustment towards the improved architectural design of the network. This shift resulted in many new architectural ideas such as channel boosting, spatial and feature-map wise exploitation and attention-based information processing etc.,[80][81][82][72]. In industry, companies such as Google, Microsoft, AT&T, NEC, and Facebook have developed active research groups for exploring new architectures of CNN [6] [7][1].

This review gives an insight into the basic structure of CNN as well as its historical perspective. This survey will help the readers to develop the theoretical insight into the design principles of CNN and thus may further accelerate the architectural innovations in CNN. The rest of the paper is organized in the following order: Section I develops the systematic understanding of CNN, discusses its resemblance with primate's visual cortex, as well as its contribution to MV. Section II provides history of CNN, Section III discusses about various layers of CNN. Whereas, Section IV discusses the recent innovations in CNN architectures. Section V discusses on applications of CNNs. Finally, the last section concludes.

II. History of CNN

With the discovery of concepts such as convolution and back propagation applied to neural networks, NN got better. With the power of GPUs and more efficient algorithms, CNNs can be applied to real life applications[8].

In 1959, Hubel & Wiesel [28] found that cells in animal visual cortex are responsible for detecting light in receptive fields[27].In 1959, an apparatus was developed by Russell Kirsch along with his colleagues for transforming images into grids of numbers so that the machine can understand the images[110]. The evolutionary history of deep CNN architectures is pictorially represented in Fig. 1. Improvements in CNN architectures can be categorized into five different area that are discussed below.

2.1 Origin of CNN: Late 1980s-1999

In 1998, LeCuN proposed an improved version of ConvNet, which was famously known as LeNet-5 [83][84]. Due to the good performance of CNN in optical character and fingerprint recognition, its commercial use in ATM and Banks started in 1993 and 1996, respectively.

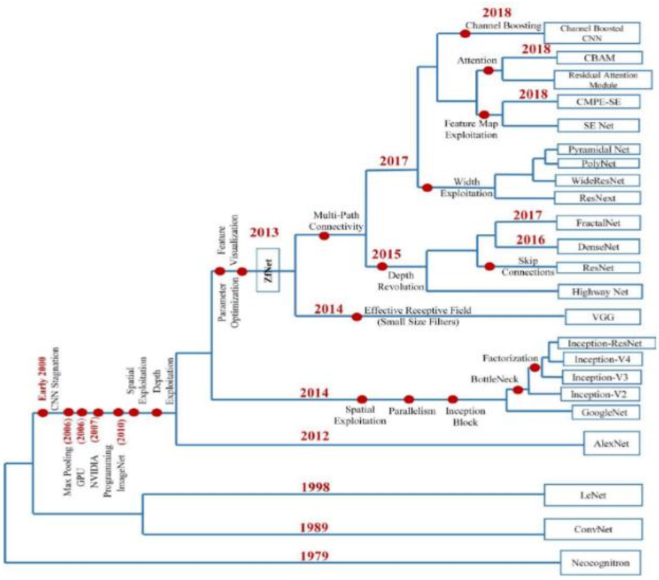

Fig. 1 Evolutionary history of deep CNNs showing architectural innovations from ConvNet till to date architectures.

A. Stagnation of CNN: Early 2000

ISSN (Online) 2581-9429
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, October 2022
Impact Factor: 6.252

As CNN was not effective in converging to the global minima of the error surface it was considered as a less effective feature extractor compared to handcrafted features [85]. In 2003, [86] improved CNN architecture and showed good results compared to SVM on a hand digit benchmark dataset.

2.2 Revival of CNN: 2006-2011

Hinton reported the concept of greedy layer-wise pre-training in 2006. Experimental studies showed that both supervised and unsupervised pre-training could initialize a network in a better way than random initialization. The observation of Bengio and other researchers started the use of activation functions other than sigmoid such as ReLU, tanh etc., [87]. [88]used max-pooling instead of subsampling, which showed good results by learning invariant features [89]. In late 2006, researchers started using graphics processing units (GPUs) to accelerate the training of deep NN and CNN architectures [90][91][92][93]. In 2007, NVIDIA launched the CUDA programming platform, which allows exploitation of parallel processing capabilities of GPU with a greater degree [94] [95]. ImageNet Large Scale Visual Recognition Challenge (ILSVRC)and Neural Information Processing Systems Conference (NIPS) are the two platforms that play a dominant role in strengthening research and increasing the use of CNN and thus making it popular.

2.3 Rise of CNN: 2012-2014

The main breakthrough in CNN performance was brought by AlexNet, which showed exemplary performance in 2012-ILSVRC [96]. The problems of determining filter dimensions, stride, padding, and other hyper-parameters for each layerresolved by the concept of modularity in CNNs made it easy to tailor them for different tasks effortlessly [97][98]. In this connection, a different idea of branching and block within a layer was introduced by the Google group [99].

2.4 Rapid Increase In Architectural Innovations And Applications of CNN: 2015-Present

In 2015, different ideas such as information gating mechanism across multiple layers, skip connections, and cross-layer channel connectivity was introduced [100][101][102]. Most of the famous object detection and segmentation architectures such as Single Shot Multibox Detector (SSD), Region-based CNN (R-CNN), Faster R-CNN, Mask R-CNN and Fully Convolutional Neural Network (FCN) are built on the lines of ResNet, VGG, Inception, etc. [103][104][105]. Applications of deep showed state-of-the-art results on MS COCO-2015 image captioning challenge.

Similarly, in 2016, it was observed that the stacking of multiple transformations not only depth-wise but also in parallel fashion showed good learning for complex problems [76][77]. Different researchers used a hybrid of the already proposed architectures to improve deep CNN performance [102]. In 2017, the focus was on designing of generic blocks In 2018, a new idea of channel boosting was introduced by [81].

The solutions to the problems of high computational cost and memory requirement are discussed in [106][107][108][109]. From 2012 up till now, many improvements have been reported in CNN architectures[72].

III. CNN LAYERS

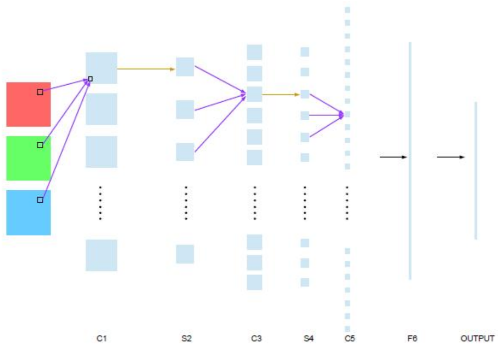

Figure 1: Illustration of LeNet-5.

Fig 2 : Illustration of LeNet-5

Convnets are very similar to normal neural networks which can be visualized as a collection of neurons arranged as an acyclic graph. The main difference from a neural network isthat a hidden layer neuron is only connected to a subset of neurons in the previous layer. Because of this sparse connectivity it is capable to learn features implicitly. The deep architecture of the network results in hierarchical feature extraction [47].

The modern convolutional neural networks proposed by LeCun [113] is a 7-layer (excluding the input layer)LeNet-5structure. It has the following structure C1,S2, C3, S4, C5, F6, OUTPUT as shown in Figure 2,where C indicates convolutional layer, S indicates sub sampling layer, and F indicates fully-connected layer.

Among different structures, they share fourkey features including weight sharing, local connection, pooling, and the use of many layers [112]. There are some commonly used layers such as convolutional layers, sub sampling layers (pooling layers), and fully connected layers. Usually, there is a convolutional layer after the input. The convolutional layer is often followed by a subsampling layer. This combination repeats several times to increase the depth of CNN. The fully connected layers are designed as the last few layers in order to map from extracted features to labels. [111]

As fundamental building blocks of CNNs, CNN layers have shown their variety and flexibility both in their designed structures and the connections. The efforts in modifying layer structures and connections have led to the possibility of training a CNN faster and of making it perform better. In following subsections, we will introduce common layers of modern CNNs and their functions.

3.1 Activation function

Activation function serves as a decision function and helps in learning of intricate patterns. The activation function for a convolved feature-map is defined in equation (3).

$$\mathbf{T}_l^k = g_a(\mathbf{F}_l^k) \quad (3)$$

In the above equation, $\mathbf{F}_l^k$ an output of a convolution, which is assigned to activation functionga(.) that adds non-linearity and returns a transformed output for lth layer $\mathbf{T}_l^k$.[72]There are several types of activation functions[72][20][27] [47][111].

3.2 Pooling

Feature motifs, which result as an output of convolution operation, can occur at different locations in the image. Oncefeatures are extracted, its exact location becomes less important as long as its approximate position relative to others ispreserved. Pooling or down-sampling is an interesting local operation. It sums up similar information in theneighborhood of the receptive field and outputs the dominant response within this local region (Lee et al. 2016).[47]Pooling is an important step to further reduce thedimensions of the activation map, keeping only the importantfeatureswhile also reducing the spacial invariance. This inturn reduces the number of learnable features for the model.Thishelps to resolve the problem of overfitting. Poolingallows CNN to incorporate all the different dimensions of animageso that it successfully recognises the given object evenif its shape is skewed or is present at a different angle.[121]

3.3 Convolutional Layer

The convolutional layer is composed of a set of convolutional kernels where each neuron acts as a kernel. However, if the kernel is symmetric, the convolution operation becomes a correlation operation. Convolutional kernel works by dividing the image into small slices, commonly known as receptive fields. The division of an image into small blocks helps in extracting feature motifs. Kernel convolves with the images using a specific set of weights by multiplying its elements with the corresponding elements of the receptive field[72].

Convolution Layer basically convolves or multiplies the pixel matrix generated for the given image or object to produce an activation map for the given image. The main advantage of activation map is that it stores all the distinguishing features of a given image while at the same time reducing the amount of data to be processed. The matrix with which the data is convolved is a feature detector which basically is a set of values with which the machine is compatible. Different versions of image are generated using different values of feature detector. The convoluted model is also trained with back propagation in order to ascertain minimal error in each layer. According to the lowest error set, depth and padding is set.

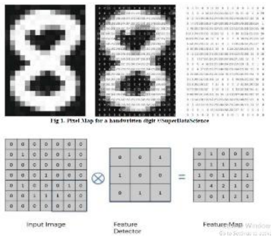

Fig 3. Convolution to produce an Activation Map ©SuperDataScience

Fig. 3 above shows how convolution works. This step involves convoluting the matrix containing the image data and then feature detector which gives us an activation map or a feature map. What happens in convolution is that the values on identical positions in the data and feature map i.e. values having value 1 or more than 1 are kept while rest areremoved. The matrix from the image data is compared 3x3 at a time. The size of feature detector varies with the type of CNN used. For example there are versions of CNN which use 5x5 or even7x7 scale filters for convolution. Convolution follows which aims to show how one function modifies the shape of the other [121].

The spatial extend of sparse connectivity between the neurons of two layers is a hyper paramter called receptive field. The hyper parameters that control the size of the output volume are the depth(number of filters at a layer), stride(for filter movement) and zero padding(to control spatial size of output). The ConvNets are trained with Back propagation and the backward pass as well involves convolution operation but with spatially flipped filters. One of the variants of the traditional CNN is "Network In Network"(NIN) proposed by Lin et al. [114] [115]. To alleviate saturation problem, many non-saturated activations are proposed such as Rectified Linear Unit (ReLU) [116], Leaky ReLU [117], Parametric ReLU (PReLU) [118], Randomized Leaky ReLU (RReLU) [119], and Exponential Linear Unit (ELU) [120][111].

Convolution filter in basic CNNs is a Generalized Linear Model (GLM) for the underlying local image patch. It works well for abstraction when instances of latent concepts are linearly separable. There are some types of convolution which can be sued to enhance CNN's representation abilitylike, Tiled Convolution, Transposed Convolution, Dilated Convolution, Network in Network, Inception Module[27].

3.4 Fully Connected (FC)

The fully connected layer is the last layer of CNN architecture. Neurons in this layer are fully connected to all neurons in the previous layer, as in a regular Neural Network. High level reasoning is done here. The neurons are not spatially arranged(one dimensional) so there cannot be a conv layer after a fully connected layer [47]. In a typical CNN, the feature maps of last convolutional layer are vectorized and fully connected with output units, which are followed by a softmax loss layer [122]. This FC layer takes input from other layers and transforms them into specific no. of classes which are already decided by the network. The output layer of the FC layer is computed for error calculation. Then, a loss function(SVM/Softmax) is defined to compute the gradient of error. These errors are propagated backwardly to update weights and bias in back propagation neural network. In one forward and backward pass, one cycle is completed for training[110].

3.5 Loss Layer

The last fully connected layer serves as the loss layer that computes the loss or error which is a penalty for discrepancy between desired and actual output. For predicting a single class out of K mutually exclusive classes Softmax loss isused. It is the commonly used loss function. Basically it is multinomial logistic regression. It maps the predictions tonon- negative values and also normalized to get probability distribution over classes. Large margin classifier, SupportVector Machine, is trained by calculating Hinge loss. For regressing to real-valued labels Euclidean loss can be used. [47]Four representative losses are Hinge loss, Softmax loss, Contrastive loss, Triplet loss.

IV. CNN ARCHITECTURES

4.1 LeNet

This was one of the pioneering work in Convolutional Neural Networks by LeCun et al. In 1990 [56] and later improved it in 1998 [57]. Itfinds application in reading zip codes, digits, etc. [47]. LeNet exploited the underlying basis of the image that the neighboring pixels are correlated to each other and feature motifs are distributed across the entire image. Therefore, convolution with learnable parameters is an effective way to extract similar features at multiple locations with few parameters. Learning with sharable parameters changed the conventional view of training where each pixel was considered as a separate input feature from its neighborhood and ignored the correlation among them. LeNet was the first CNN architecture, which not only reduced the number of parameters but was able to learn features from raw pixels automatically [72].The final input that isprovided to the FCNis of dimension 120x1x1. No. of parameters taken into account are approximately 60,000[121].

4.2 AlexNet

This architecture developed by Alex Krizhevsky, Ilya Sutskever and Geoff Hinton [58] is credited as the first work in Convolutional Networks to popularize it in the field of computer vision and was published in 2012. The network was similar to LeNet but instead of alternating conv layers and pooling layers, AlexNet had all the conv layers stacked together [47]. Input provided was about 15million RGB image data of dimensions 3x224x224 image(which was later corrected to 227x227) from about 22,000categories. This Deep CNN consists of 7 layers and 60million parameters. The operations consists of 11x11,5x5,3x3 convolutions, 3x3 max pooling ending with 2 fully connected layers of 4096 layers each. AlexNet was able to win the ILSVRC-2012 competitions achieving top-1and top-5 error rates on test dataset[121].To address challenge of over fitting, Krizhevesky et al. (2012) exploited the idea of Hinton (Dahl et al. 2013; Srivastava et al. 2014)[72].

4.3 GoogleNet

This ConvNet architecture from Szegedy et al.[59] from Google won the ILSVRC 2014 competition. They have proposed a new architecture called Inception (v1) that gives more utilization of the computing resources in the network[47].The GoogLeNet also followed the block pattern as followed by VGG 16 and each block is called Inception Module. GoogLeNet uses total of 9 inception module. This network 22 layers deep but compensated by relatively much lesser no. of parameters of about 5 million compared to runner up VGG's 138 million parameters. [121] In GoogleNet, conventional convolutional layers are replaced in small blocks similar to the idea of substituting each layer with micro NN as proposed in Network in Network (NIN) architecture (Lin et al. 2013). The exploitation of the idea of split, transform, and merge by GoogleNet, helped in addressing a problem related to the learning of diverse types of variations present in the same category of images having different resolutions. GoogleNet regulates the computations by adding a bottleneck layer of 1x1 convolutional filter, before employing large size kernels. Other regulatory factors applied were batch normalization and the use of RmsProp as an optimizer (Dauphin et al. 2015). GoogleNet also introduced the concept of auxiliary learners to speed up the convergence rate. However, the main drawback of the GoogleNet was its heterogeneous topology and a representation bottleneck[72].

4.4 VGG(16) Net

Karen Simonyan and Andrew Zisserman have done a thorough analysis of the depth factor in a ConvNet, keeping all other parameters fixed [47].VGG stands for Visual Geometric Group. This network contains about 138 million parameters. Thereare total 16 convolutional layers in VGG 16, distributed in 3blocks containing 2layers of 3x3 convolutions followed by2x2 max pooling and 2 blocks containing 3 layers of 3x3convolutions followed by 2x2 max pooling. The architecture isfinally finished by 2 fully connected layers of 4096 hidden layers each. [121] The use of the small size filters provides an additional benefit of low computational complexity by reducing the number of parameters. VGG regulates the complexity of a network by placing 1x1 convolutions in between the convolutional layers, which, besides, learn a linear combination of the resultant feature-maps. For the tuning of the network, max-pooling is placed after the convolutional layer, while padding was performed to maintain the spatial resolution (Ranzato et al. 2007)[123].VGG was at 2nd place in the 2014-ILSVRC competition but, got fame due to its simplicity, homogenous topology, and increased depth. The main limitation associated with VGG was the use of 138 million parameters, which make it computationally expensive and difficult to deploy it on low resource systems. [72]

4.5 ResNet

Kaiminget. al. [60] have presented a residual learning framework where the layers learn residual functions with respect to the inputs received instead of learning unreferenced functions[47]. ResNet or Deep Residual Network was the winner for ImageNet challenge for 2015 surpassing the human accuracy error for the first time with an error rate of about 3.6%. The network is very very deep and the one presented at challenge was 152 layers deep. In order to counter Vanishing Gradient Problem, ResNet introduced a feature of skip connection. The concept behind resnet was that even if thenet work was deep, the training of the network was similar to that of shallow network by skipping after every 2 layer. In order to compute, the input and output both were copied to the next layer basically learning the residual of previous computation. No. of parameters computed were about 65 million. Some layers also have bottleneck starting and ending by 1x1convolution. Batch Normalization was used after each convolution.[121]In order to address the problems faced during training of deep networks, ResNet exploited the idea of bypass pathways used in Highway Networks (He et al. 2015a). Residual links (shortcut connections) speed up the convergence of deep networks, thus giving ResNet the ability to avoid gradient diminishing problems. [72]To overcome the problem of high depth in residual networks, authors Zagoruyko et al. [21] proposed wide residual networks (WRNs) used for mitosis detection in breast histology images [22]. Larsson et al. [23] proposed Fractalnet [123].

4.6 ResNeXt

ResNeXt, also known as Aggregated Residual Transform Network, is an improvement over the Inception Network (Xie et al. 2017). Cardinality is an additional dimension, which refers to the size of the set of transformations (Han et al. 2018; Sharma and Muttoo 2018). ResNeXt utilized the deep homogenous topology of VGG and simplified GoogleNet architecture. ResNeXt used multiple transformations within a split, transform and merge block and defined these transformations in terms of cardinality. The complexity of ResNeXt was regulated by applying low embedding’s (1x1 filters) before 3x3 convolution, whereas training was optimized by using skip connections (Larsson et al. 2016).[72]

V. CNN APPLICATION

Because of CNN popularity, it has many applications and used in extensive manner. Some of them are discussed below:

5.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) converts language into a presentation that can easily be exploited by any computer. CNNs have also been utilized in NLP based applications such as language modeling, and analysis, etc. Especially, language modelling or sentence modelling has taken a twist after the introduction of CNN as a new representation learning algorithm.[72] CNNs are usually utilized in natural language processing [61]. CNN models are useful for numerous natural language processing issues and achieved glorious results in text categorization, semantic parsing [62], search query retrieval [63], and classification [64], prediction [65], text categorization [66], diversified traditional natural language processing [61][47] , Statistical Language Modelling, Text Classification[27], Sentence Modelling, Twitter Sentiment Analysis and Semantic Role Labeling tasks[8].

5.2 Computer Vision and Related Applications

Computer vision (CV) focuses on developing an artificial system that can process visual data, including images and videos and can effectively understand, and extract useful information from it .Farfade, et al. proposed deep CNN for detecting faces from different poses as well as from occluded faces [67]. In another work, Zhang et al. performed face detection using a new type of multitasking cascaded CNN [68]. Zhang’s technique showed good results when compared to state-of-the-art techniques [69].Experimental results have shown that Wang’s technique outperforms other activity recognition based techniques [70, 71]. Similarly, another three dimensional CNN based action recognition system is proposed by Ji et al. [72]. [47] [72] Deep Pose is the first application of CNNs to human pose estimation problem [27].

5.3 Object Detection and Segmentation

Object detection focuses on identifying different objects in images. Recently, R-CNN has been widely used for object detection[47] . Recently, AE based CNN architectures have shown success in segmentation tasks. [72] One of the most famous object proposal based CNN detector is Region-based CNN (R-CNN) [33]. Spatial pyramid pooling network (SPP net) [34] is a pyramid-based version of R-CNN [33], which introduces an SPP layer to relax the constraint that input images must have a fixed size. Fast RCNN [35] improves SPP net by using an end-to-end training method. Later, Faster R-CNN [35] introduces a region proposal network (RPN) for object proposals generation and achieves further speed-up. More recently, YOLO [36] and SSD [37] allow single pipeline detection that directly predicts class labels.[27]

5.4 Image Classification

CNN has been widely used for image classification. One Recently, Spanhol et al. used CNN for the diagnosis of breast cancer images, and results are compared against a network trained on a dataset containing handcrafted descriptors [47], Wahab et al. 2017, Cireşan et al. 2012]. The different pixel neighborhoods can be evaluated with the help of spatial dependencies of cell and non-cell pixels. In (Su, Liu, Xie, et al. 2015), proposed method is used to segment a high resolution image (1000*1000) done in only 2.3 seconds[38].

Subcategory classification is another rapidly growing subfield of image classification. Along this way, Branson et al .[39] propose a method which detects parts and extracts CNN features from multiple pose-normalized regions. Zhanget al . [40] propose a part-based R-CNN which can learn whole-object and part detectors. Lin et al . [41] incorporate part localization, alignment, and classification into one recognition system which is called Deep LAC. Krause et al . [42] use the ensemble of localized learned feature representations for fine-grained classification.

5.5 Object Tracking

There are several attempts to employ CNNs for visual tracking. Fan et al . [43] use CNN as a base learner. In [43], the authors design a CNN tracker with a shift-variant architecture Li et al. [44] propose a target-specific CNN for object tracking, where the CNN is trained incrementally during tracking with new examples obtained online. Li et al . [44] use a relatively small number of filters in the CNN within a framework equipped with a temporal adaptation mechanism. In [45], a CNN object tracking method is proposed to address limitations of handcrafted features and shallow classifier structures in object tracking problem. In [46] propose a visual tracking algorithm based on a pre-trained CNN[27].

VI. CONCLUSION

CNN has made remarkable progress, especially in image processing and vision-related tasks, and has thus revived the interest of researchers in ANNs. In this context, several research works have been carried out to improve CNN's performance on such tasks. Due to the advantages of CNNs, such as local connection, weight sharing, and down sampling dimensionality reduction, CNN has been widely deployed in both research and industry projects. The advancements in CNNs can be categorized in different ways, including activation, loss function, optimization, regularization, learning algorithms, and innovations in architecture. This paper studies advancement in the CNN architectures, especially based on the design patterns of the processing units and thus has proposed the taxonomy for recent CNN architectures. In addition to the categorization of CNNs into different classes, this paper also covers the history of CNNs, its applications. The learning capacity of CNN is significantly improved over the years by exploiting depth and other structural modifications. It is observed in recent literature that the main boost in CNN performance has been achieved by replacing the conventional layer structure with blocks. Even though convolutions possess many benefits and have been widely used, we reckon that they can be refined further in terms of model size, and security. Moreover, there are lots of problems that convolution is hard to handle, such as low generalization ability, lack of equivariance, and poor crowded-scene results, so that several promising directions are pointed. Although CNNs have achieved great success in experimental evaluations, there are still lots of issues that deserve further investigation. Firstly, they require large-scale dataset and massive computing power for training. Manually collecting labeled dataset requires huge amounts of human efforts. Thus, it is desired to explore unsupervised learning of CNNs. Meanwhile ,to speed up training procedure, it is still worth to develop effective and scalable parallel training algorithms. It is important to investigate how to reduce the complexity and obtain fast-to-execute models without loss of accuracy. Furthermore, one major barrier for applying CNN on a new task is that it requires considerable skill and experience to select suitable hyper parameters. These hyper-parameters have internal dependencies which make them particularly expensive for tuning. Finally, the solid theory of CNNs is still lacking. Meanwhile, it is also worth exploring how to leverage natural visual perception mechanism to further improve the design of CNN. I hope that this paper not only provides a better understanding of CNNs but also facilitates future research activities and application developments in the field of CNNs.

REFERENCES

[1]. S. Majid Rezaee, A Review On Convolutional Neural Networks And Its Applications, https://www.researchgate.net/publication/353922229, July 2021

[2]. John McCarthy. what is artificial intelligence? Computer Science Department Stanford University.www-formal.stanford.edu/jmc/. 2007 Nov 12.

[3]. RaffaeleCioffi, Marta Travaglioni, GiuseppinaPiscitelli, AntonellaPetrillo, andFabio De Felice . Artificial Intelligence and Machine Learning Applications in Smart production: Progress, Trends, and Directions. Sustainability 2020, 12, 492; doi:10.3390/su12020492.

[4]. Tom M. Mitchell. Machine Learning Definition, McGraw-Hill Science, Engineering, Math; (March 1, 1997), Page 1.

[5]. Adrian Carrio, Carlos Sampedro, Alejandro Rodriguez-Ramos, and PascualCampoy. Review Article A Review of Deep Learning Methods and Applications for Unmanned Aerial

[6]. Deng L, Yu D, Delft B— (2013) Deep Learning: Methods and Applications Foundations and Trends R in Signal Processing. Signal Processing 7:3–4. doi: 10.1561/2000000039.

[7]. Asifullah Khan, AnabiaSohailUmmeZahoora, and Aqsa Saeed Qureshi. A Survey of the Recent Architectures of Deep Convolutional Neural Networks. Published in Artificial Intelligence Review, DOI: https://doi.org/10.1007/s10462-020-09825-6.

[8]. D. T. Mane, SGGS IE&T, Nanded, India, U. V. Kulkarni, A Survey on Supervised Convolutional Neural Network and Its Major ApplicationsInternational Journal of Rough Sets and Data Analysis, Volume 4 , Issue 3, 2017

[9]. Xizhao Wang1 • Yanxia Zhao1 • Farhad Pourpanah2, Recent advances in deep learning, International Journal of Machine Learning and Cybernetics (2020) 11:747–750

[10]. Wang X, Joshua HZ (2015) Uncertainty in learning from big data. Fuzzy Sets Syst 258:1–4

[11]. Rezvani S, Wang X, Pourpanah F (2019) Intuitionistic fuzzy twin support vector machines. IEEE Trans Fuzzy Syst 27(11):2140–2151

[12]. Wang Z, Wang X (2018) A deep stochastic weight assignment network and its application to chess playing. J Parallel DistribComput 117:205–211

[13]. Wang Z, Wang X (2018) A deep stochastic weight assignment network and its application to chess playing. J Parallel DistribComput 117:205–211

[14]. Pourpanah F, Lim CP, Wang X, Tan CJ, Seera M, Shi Y (2019) A hybrid model of fuzzy minmax and brain storm optimization for feature selection and data classification. Neurocomputing 333:440–451

[15]. Sengupta S, Basak S, Saikia P, Paul S, Tsalavoutis V, Atiah F, Ravi V, Peters A (2020) A review of deep learning with special emphasis on architectures, applications and recent trends. Knowl-Based Syst https ://doi.org/10.1016/j.knosy s.2020.10559 6

[16]. I. Krizhevsk, Alex Sutskever and G. E. Hinton, "ImageNetClassificationWithDeepConvolutionalNeural Networks," Adv. Neural Inf. Process. Syst., 2012.

[17]. T. Chilimbi, Y. Suzue, J. Apacible, and K. Kalyanaraman "ProjectAdam: Building An Efficient And Scalable DeepLearning Training System," 11th USENIX Symp. Oper. Syst. Des. Implement., pp. 571–582, 2014.

[18]. W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, "Parallel distributed processing model with local spaceinvariant interconnections and its optical architecture," Appl. Opt., 1990.

[19]. Ali FadhilYaseen, A Survey on the Layers of Convolutional Neural Networks, IJCSMC Journal, Vol.7 Issue.12, December- 2018, pg. 191-196
[20]. Chenyou Fan Indiana University Bloomington, IN, Survey of Convolutional Neural Network
[21]. Y. LeCun, et al.,Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4):541–551, 1989.
[22]. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradientbased learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998
[23]. P. Sermanet, S. Chintala, and Y. LeCun. Convolutional neural networks applied to house numbers digit classification. In Pattern Recognition (ICPR), 2012 21st International Conference on, pages 3288–3291. IEEE, 2012.
[24]. P. Sermanet and Y. LeCun. Traffic sign recognition with multi-scale convolutional networks. In Neural Networks (IJCNN), The 2011 International Joint Conference on, pages 2809–2813. IEEE, 2011
[25]. L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. Computer Vision and Image Understanding, 106(1):59–70, 2007.
[26]. I. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
[27]. JiuxiangGu et al., Recent Advances in Convolutional Neural Networks, arXiv:1512.07108v6 [cs.CV] 19 Oct 2017
[28]. D. H. Hubel, T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, The Journal of physiology (1968) 215–243.
[29]. K. Fukushima, S. Miyake, Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition, in: Competition and cooperation in neural nets, 1982, pp. 267–285.
[30]. *B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, Handwritten digit recognition with a back propagation network, in: Proceedings of the Advances in Neural Information Processing Systems (NIPS), 1989, pp. 396–404.*
[31]. Y. LeCun, et al., Gradient-based learning applied to document recognition, Proceedings of IEEE 86 (11) (1998) 2278–2324.
[32]. R. Hecht-Nielsen, Theory of the backpropagation neural network, Neural Networks 1 (Supplement-1) (1988) 445–448.
[33]. R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 580–587
[34]. K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 37 (9) (2015) 1904–1916.
[35]. S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 39 (6) (2017) 1137–1149.
[36]. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
[37]. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, Ssd: Single shot multibox detector, in: Proceedings of the European Conference on Computer Vision (ECCV), 2016, pp. 21–37.
[38]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al., Imagenet large scale visual recognition challenge, International Journal of Conflict and Violence (IJCV) 115 (3) (2015) 211–252.
[39]. S. Branson, G. Van Horn, P. Perona, S. Belongie, Improved bird species recognition using pose normalized deep convolutional nets, in: Proceedings of the British Machine Vision Conference (BMVC), 2014.

[40]. N. Zhang, J. Donahue, R. Girshick, T. Darrell, Part-based r-cnns for fine-grained category detection, in: Proceedings of the European Conference on Computer Vision (ECCV), 2014, pp. 834–849.

[41]. Lin, X. Shen, C. Lu, J. Jia, Deep lac: Deep localization, alignment and classification for fine-grained recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1666–1674.

[42]. J. Krause, T. Gebru, J. Deng, L.-J. Li, L. Fei-Fei, Learning features and parts for fine-grained recognition, in: Proceedings of the International Conference on Pattern Recognition (ICPR), 2014, pp. 26–33.

[43]. J. Fan, W. Xu, Y. Wu, Y. Gong, Human tracking using convolutional neural networks, IEEE Trans. Neural Networks(TNN) 21 (10) (2010) 1610–1623.

[44]. H. Li, Y. Li, F. Porikli, Deeptrack: Learning discriminative feature representations by convolutional neural networks for visual tracking, in: Proceedings of the British Machine Vision Conference (BMVC), 2014.

[45]. Y. Chen, X. Yang, B. Zhong, S. Pan, D. Chen, H. Zhang, Cnntracker: Online discriminative object tracking via deep convolutional neural network, Appl. Soft Comput. 38 (2016) 1088–1098.

[46]. S. Hong, T. You, S. Kwak, B. Han, Online tracking by learning discriminative saliency map with convolutional neural network, in: Proceedings of the International Conference on Machine Learning (ICML), 2015, pp. 597–606.

[47]. Neena Aloysius and Geetha M, A Review on Deep Convolutional Neural Networks, International Conference on Communication and Signal Processing, April 6-8, 2017, India

[48]. J. Fan, W. Xu, Y. Wu, and Y. Gong, Human tracking using convolutionalneural networks, Neural Networks, IEEE Transactions, 2010.

[49]. I. Toshev and C. Szegedy, Deep -pose: Human pose estimation viadeepneural networks, in CVPR, 2014.

[50]. M. Jaderberg, A. Vedaldi, and A. Zisserman, Deep features for textspotting, in ECCV, 2014.

[51]. R. Zhao, W. Ouyang, H. Li, and X. Wang, Saliency detection by multicontextdeep learning, in CVPR, 2015.

[52]. J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T.Darrell, Decaf: A deep convolutional activation feature for generic, 2014

[53]. *Farabet, C. Couprie, L. Najman, and Y. LeCun, Learning hierarchicalfeatures for scene labeling, PAMI, 2013.*

[54]. Nithin, D Kanishka and Sivakumar, P Bagavathi, Generic Feature Learningin Computer Vision, Elsevier, Vol.58, Pages202-209, 2015.

[55]. Matthew D. Zeiler, Rob Fergus, Visualizing and Understanding Convolutional Networks, 13th European Conference, Zurich, Switzerland,2014, Proceedings, Part I, Volume 8689, pages:818-833.

[56]. B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hub- bard, and L. D. Jackel, Handwritten digit recognition with a back- propagation network, in NIPS, 1990.

[57]. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE,1998

[58]. Alex Krizhevsky, Sutskever I, and Hinton G.E. Imagenet classificationwith deep convolutional neural networks. In NIPS, 2012.

[59]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, Going deeper withconvolutions, CoRR, 2014.

[60]. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for imagerecognition, arXiv preprint arXiv:1512.03385, 2015.

[61]. B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, Handwritten digit recognition with a backpropagation network, in NIPS. Citeseer, 1990.

[62]. Y. LeCun, et al., Gradient-based learningapplied to document recognition, Proceedings of the IEEE, 1998.

[63]. Alex Krizhevsky, Sutskever I, and Hinton G.E, Imagenet classificationwith deep convolutional neural networks. In NIPS, 2012.

[64]. K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In International Conference on Computer Vision, pages 21462153. IEEE, 2009.

[65]. I. Krizhevsky. Convolutional deep belief networks on cifar-10. Unpublished manuscript, 2010.

[66]. L. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hub- bard, and L. D. Jackel, Handwritten digit recognition with a backpropagationnetwork, in NIPS. Citeseer, 1990

[67]. Zeiler M., Taylor G., and Fergus R. Adaptive deconvolutional networks for mid and high level feature learning, In ICCV, 2011.

[68]. S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies,2001.

[69]. A. Krizhevsky. Convolutional deep belief networks on cifar-10. Unpublished manuscript, 2010. G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R.

[70]. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580, 2012.

[71]. Fisher Yu and VladlenKoltun, Multi-Scale Context Aggregation by Dilated Convolutions, ICLR,2016

[72]. Asifullah Khan, AnabiaSohail, UmmeZahoora, and Aqsa Saeed Qureshi, A Survey of the Recent Architectures of Deep Convolutional Neural Networks, Artificial Intelligence Review, DOI: https://doi.org/10.1007/s10462-020-09825-6

[73]. Qiang Yang, Pan SJ, Yang Q, Fellow QY (2008) A Survey on Transfer Learning. IEEE Trans Knowl Data Eng 1:1–15. doi: 10.1109/TKDE.2009.191

[74]. Qureshi AS, Khan A (2018) Adaptive Transfer Learning in Deep Neural Networks: Wind Power Prediction using Knowledge Transfer from Region to Region and Between Different Task Domains. arXivPrepr arXiv181012611

[75]. Qureshi AS, Khan A, Zameer A, Usman A (2017) Wind power prediction using deep neural network based meta regression and transfer learning. Appl Soft Comput J 58:742–755. doi: 10.1016/j.asoc.2017.05.031

[76]. Zagoruyko S, Komodakis N (2016) Wide Residual Networks. Procedings Br Mach Vis Conf 2016 87.1-87.12. doi: 10.5244/C.30.87

[77]. Han D, Kim J, Kim J (2017) Deep Pyramidal Residual Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 6307–6315

[78]. Xie S, Girshick R, Dollar P, et al (2017) Aggregated Residual Transformations for Deep Neural Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 5987–5995

[79]. Zhang X, Li Z, Loy CC, Lin D (2017) PolyNet: A Pursuit of Structural Diversity in Very Deep Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 3900–3908

[80]. Wang F, Jiang M, Qian C, et al (2017a) Residual Attention Network for Image Classification. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 6450–6458

[81]. Khan A, Sohail A, Ali A (2018a) A New Channel Boosted Convolutional Neural Network using Transfer Learning. arXivPrepr arXiv180408528

[82]. Woo S, Park J, Lee JY, Kweon IS (2018) CBAM: Convolutional block attention module. Lect Notes ComputSci (including SubserLect Notes ArtifIntellLect Notes Bioinformatics) 11211 LNCS:3–19. doi: 10.1007/978-3-030-01234-2_1

[83]. LeCun Y, et al., 1998, Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324

[84]. LeCun Y, Jackel LD, Bottou L, et al (1995) Learning algorithms for classification: A comparison on handwritten digit recognition. Neural networks Stat MechPerspect 261:276

[85]. Schmidhuber J (2007) New millennium AI and the convergence of history. In: Challenges for computational intelligence. Springer, pp 15–35

[86]. Simard PY, Steinkraus D, Platt JC (2003) Best practices for convolutional neural networks applied to visual document analysis. In: null. p 958

[87]. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp 249–256

[88]. Ranzato M, Huang FJ, Boureau YL, LeCun Y (2007) Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, pp 1–8

[89]. Giusti A, Ciresan DC, Masci J, et al (2013) Fast image scanning with deep max-pooling convolutional neural networks. In: 2013 IEEE International Conference on Image Processing. IEEE, pp 4034–4038
[90]. Oh K-S, Jung K (2004) GPU implementation of neural networks. Pattern Recognit 37:1311–1314
[91]. Strigl D, Kofler K, Podlipnig S (2010) Performance and scalability of GPU-based convolutional neural networks. In: Parallel, Distributed and Network-Based Processing (PDP), 2010 18th Euromicro International Conference on. pp 317–324
[92]. Cireşan DC, Meier U, Masci J, et al (2011) High-Performance Neural Networks for Visual Object Classification. arXivPrepr arXiv11020183
[93]. Nguyen G, Dlugolinsky S, Bobák M, et al (2019) Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. ArtifIntell Rev 52:77–124. doi: 10.1007/s10462-018-09679-z
[94]. Nickolls J, Buck I, Garland M, Skadron K (2008) Scalable parallel programming with CUDA. In: ACM SIGGRAPH 2008 classes on - SIGGRAPH '08. ACM Press, New York, New York, USA, p 1
[95]. Lindholm E, Nickolls J, Oberman S, Montrym J (2008) NVIDIA Tesla: A Unified Graphics and Computing Architecture. IEEE Micro 28:39–55. doi: 10.1109/MM.2008.31
[96]. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks. Adv Neural Inf Process Syst 1–9. doi: 10.1061/(ASCE)GT.1943-5606.0001284
[97]. Simonyan K, Zisserman A (2015) Very Deep Convolutional Networks For Large-Scale Image Recognition. ICLR 75:398–406. doi: 10.2146/ajhp170251
[98]. Amer M, Maul T (2019) A review of modularization techniques in artificial neural networks. ArtifIntell Rev 52:527–561. doi: 10.1007/s10462-019-09706-7
[99]. Szegedy C, Wei Liu, YangqingJia, et al (2015) Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 1–9
[100]. Srivastava RK, Greff K, Schmidhuber J (2015a) Highway Networks. doi: 10.1002/esp.3417
[101]. He K, Zhang X, Ren S, Sun J (2015a) Deep Residual Learning for Image Recognition. Multimed Tools Appl 77:10437–10453. doi: 10.1007/s11042-017-4440-4
[102]. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. Proc - 30th IEEE ConfComput Vis Pattern Recognition, CVPR 2017 2017-Janua:2261–2269. doi: 10.1109/CVPR.2017.243
[103]. Lin TY, Dollár P, Girshick R, et al (2017) Feature pyramid networks for object detection. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 201
[104]. Cai Z, Vasconcelos N (2019) Cascade R-CNN: High Quality Object Detection and Instance Segmentation. IEEE Trans Pattern Anal Mach Intell. doi: 10.1109/tpami.2019.2956516
[105]. Pang J, Chen K, Shi J, et al (2020) Libra R-CNN: Towards Balanced Learning for Object Detection
[106]. Chen W, Wilson JT, Tyree S, et al (2015) Compressing neural networks with the hashing trick. In: 32nd International Conference on Machine Learning, ICML 2015
[107]. Han S, Mao H, Dally WJ (2016) Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. In: 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings
[108]. Wu J, Leng C, Wang Y, et al (2016) Quantized convolutional neural networks for mobile devices. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition
[109]. Frosst N, Hinton G rey (2018) Distilling a neural network into a soft decision tree. In: CEUR Published in Artificial Intelligence Review, DOI: https://doi.org/10.1007/s10462-020-09825-6
[110]. MadhusmitaSahu and Rasmita Dash, A Survey on Deep Learning: Convolution Neural Network (CNN), https://www.researchgate.net/ publication/343969393, 2021
[111]. QianruZhanga, MengZhanga,_, TinghuanChenb, ZhifeiSuna, YuzheMab, Bei Yu, Recent Advances in Convolutional Neural Network Acceleration, Neurocomputing, 2018
[112]. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[113]. Y. Lecun, et al., , Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.

[114].M. Moczulski, M. Denil, J. Appleyard, N. de Freitas, ACDC: A structured e_cient linear layer, in: International Conference on Learning Representations (ICLR), 2016.

[115].M. Jaderberg, A. Vedaldi, A. Zisserman, Speeding up convolutional neural networks with low rank expansions, in: British Machine Vision Conference (BMVC), 2014.

[116].V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: International Conference on Machine Learning (ICML), 2010, pp. 807–814.

[117].A. L. Maas, A. Y. Hannun, A. Y. Ng, Rectifier nonlinearitiesimprove neural network acoustic models, in: InternationalConference on Machine Learning (ICML), 2013.

[118].K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers:Surpassing human-level performance on imagenet classification,in: IEEE International Conference on Computer Vision(ICCV), 2015, pp. 1026–1034.

[119].L. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectifiedactivations in convolutional network, in: InternationalConference on Machine Learning Workshop, 2015.

[120].D.-A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accuratedeep network learning by exponential linear units (elus), in:International Conference on Learning Representations (ICLR),2016.

[121]. ArohanAjit, Koustav Acharya, AbhishekSamanta A Review of Convolutional Neural Networks, 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) February 2020

[122]. Chenyou Fan, Indiana University Bloomington, IN Survey of Convolutional Neural Network

[123]. AzeddineElhassouny, FlorentinSmarandacheTrends in deep convolutional neural Networks

[124].architectures: a review, IEEE/ICCSRE2019, 2019, Agadir, Morocco