Convolutional Neural Network

Ajay Singh
Computer Science
S.P.I.T
Mumbai, India
s.ajay1029@gmail.com

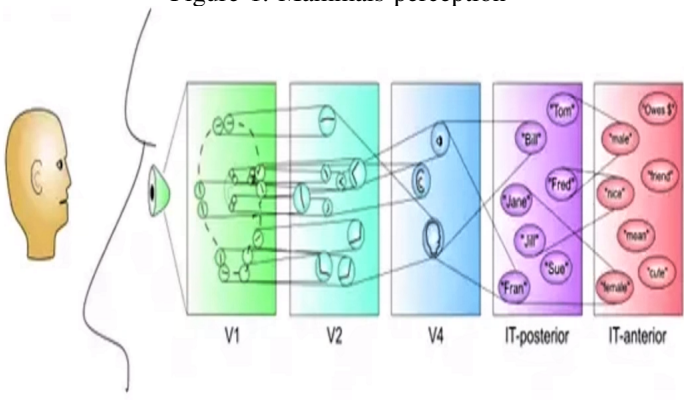
***Abstract*—With every passing day the demands for Artificial Intelligence is sky rocketing. This is due to the capability of AI that provide a machine with human like intelligence. The keynote of Artificial Intelligence is that the machine should view and interpret the world as a human. To give computers a human vision, machine uses various tasks such as image and video recognition, image analysis and classification, etc. This evolution in computer vision is basically over one specific algorithm - a convolutional neural network. This survey basically focuses on all the basic aspects of the convolutional neural network from its history, architecture, evolution, to recent advancement and also analyze the current challenges and its application.**

***Index Terms*—Neural Network,AI,Image Recognition,Face detection,Pattern recognition,Convolution.**

I. Introduction

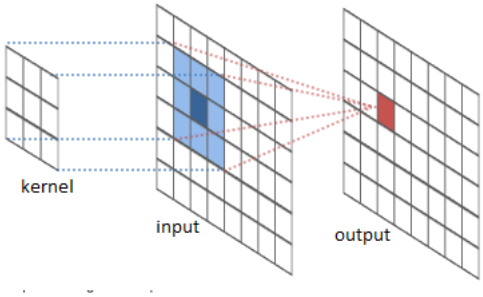
The base of CNN gains an insight from the biological model of how a mammal visually recognize the environment around using multiple layers of neuron(cells in brain) which altogether helps to recognize a pattern.The very first work done on CNN was as a neural network model for for visual pattern recognition known as the Neocognitron. The Model was self organizing i.e no teaching was needed and was able to recognize pattern irrespective of their positioning [30].

Figure 1. Mammals perception



In computer vision term ”convolutional” means that some layers of the neural network are composed of several groups of units which is collectively called the feature maps [91]. CNN is a special type of neural feed-forward networks comprising a number of convolutional layers. This convolutional layer consists of learnable filters(also known as kernels) that, when translated with a 2D input, respond to certain features, produces a 2D filtered yield [6].

Figure 2. Kernel



II. Architecture

CNN Architecture consists of ’n’ Convolution layer followed by pooling layer. Each convolution layer is assoicated with acivation function and the end is connected to Fully connected layer.The Basic Architecture of CNN is shown in the fig.3.

A. Convolutional Layer

Convolution layer consists of Convolutional kernels(every neuron resembles a kernel). These kernel are linked with a very small area in the image which is a receptive field.It works in two step
i. Dividing the image in smaller chunks/block. ii.Assigning t+he weight to it. Convolution operation can expressed as follows:

$$f_l^k = (I_x, y * K_l^k) \qquad ...(1)$$

Where, the input image is represented by $I_x y$,xy shows spatial locality and K_l^k represents the l^th convolutional kernel of the k^th layer.

B. Activation Function

Activation function serves as a decision function and helps in learning a complex pattern. Selection of an appropriate activation function can accelerate the learning process. Activation function for a convolved feature map is defined in equation (2).

$$T_l^k = f_A(F_1^k) \qquad ...(2)$$

In above equation,F_1^k is an output of a convolution operation, which is assigned to activation function;$f_A(.)$ that adds nonlinearity and returns a transformed output T_l^k for k^th layer. In literature, different activation functions such as sigmoid, tanh, maxout, ReLU, and variants of ReLU such as leaky ReLU, ELU, and PReLU are used to inculcate nonlinear combination of features. However, ReLU and its variants are preferred

over others activations as it helps in overcoming the vanishing gradient problem

C. Pooling Layer

Feature map, which result as an output of convolution operation can occur at different locations in the image. Once features are extracted, its exact location becomes less important as long as its approximate position relative to others is preserved. Pooling or downsampling like convolution, is an interesting local operation. It sums up similar information in the neighborhood of the receptive field and outputs the dominant response within this local region.

$$Z_l = f_p(f_x y^l) \quad ...(3)$$

Equation (3) shows the pooling operation in which Z_l represents the l^th output feature map, $f_x y^l$ shows the l^th input feature map, whereas $f_p(.)$ defines the type of pooling operation.

D. Fully Connected Layer

Fully connected layer is mostly used at the end of the network for classification purpose. Unlike pooling and convolution, it is a global operation. It takes input from the previous layer and globally analyses output of all the preceding layers. This makes a non-linear combination of selected features, which are used for the classification of data.

III. Evolution

CNN history begins from the neurobiological experiments conducted by Hubel and Wiesel (1959, 1962). Their work provided a platform for many cognitive models, almost all of which were latterly replaced by CNN. Over the decades, different efforts have been carried out to improve the performance of CNNs. This history is pictorially represented in Fig. 4. These improvements can be categorized into five different eras and are discussed below

A. Late 1980s-1999: Origin of CNN

CNNs have been applied to visual tasks since the late 1980s. In 1989, LeCuN et al. proposed the first multilayered CNN named as ConvNet, whose origin rooted in Fukushima's Neocognitron. LeCuN proposed supervised training of ConvNet, using Backpropagation algorithm in comparison to the unsupervised reinforcement learning scheme used by its predecessor Neocognitron. LeCuN's work thus made a foundation for the modern 2D CNNs. In 1998, ConvNet was improved by LeCuN and used for classifying characters in a document recognition application. This modified architecture was named as LeNet-5, which was an improvement over the initial CNN as it can extract feature representation in a hierarchical way from raw pixels.

B. Early 2000: Sluggish CNN

In the late 1990s and early 2000s, interest in NNs reduced and less attention was given to explore the role of CNNs in different applications such as object detection, video surveillance, etc. Use of CNN in ML related tasks became dormant due to the insignificant improvement in performance at the cost of high computational time. At that time, other statistical methods and, in particular, SVM became more popular than CNN due to its relatively high performance . It was widely presumed in early 2000 that the backpropagation algorithm used for training of CNN was not effective in converging to optimal points and therefore unable to learn useful features in supervised fashion as compared to handcrafted features. Meanwhile, different researchers kept working on CNN and tried to optimize its performance.

C. 2006-2011: Revival of CNN

NNs have generally complex architecture and time intensive training phase that sometimes spanned over weeks and even months. In early 2000, there were only a few techniques for the training of deep Networks. Additionally, it was considered that CNN is not able to scale for complex problems. These challenges halted the use of CNN in ML related tasks. To address these problems, in 2006 many interesting methods were reported to overcome the difficulties encountered in the training of deep CNNs and learning of invariant features. Hinton proposed greedy layer-wise pre-training approach in 2006, for deep architectures, which revived and reinstated the importance of deep learning . The revival of a deep learning , was one of the factors, which brought deep CNNs into the limelight. Huang et al. (2006) used max pooling instead of subsampling, which showed good results by learning of invariant features. In 2010, Fei-Fei Li's group at Stanford, established a large database of images known as ImageNet, containing millions of labeled images. This database was coupled with the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competitions, where the performances of various models have been evaluated and scored.

D. 2012-2014: Rise of CNN

Availability of big training data, hardware advancements, and computational resources contributed to advancement in CNN algorithms. Renaissance of CNN in object detection, image classification, and segmentation related tasks had been observed in this period. However, the success of CNN in image classification tasks was not only due to the result of aforementioned factors but largely contributed by the architectural modifications, parameter optimization, incorporation of regulatory units, and reformulation and readjustment of connections within the network. With CNN becoming more of a commodity in the computer vision (CV) field, a number of attempts have been made to improve the performance of CNN with reduced computational cost. Therefore, each new architecture try to overcome the shortcomings of previously proposed architecture in combination with new structural reformulations. In year 2013 and 2014, researchers mainly focused on parameter optimization to accelerate CNN performance in a range of applications with a small increase in computational complexity. In 2013, Zeiler and Fergus defined a mechanism to visualize learned filters of each CNN layer. Visualization approach was used to improve the feature extraction stage by reducing the size of the filters. Similarly, VGG

architecture proposed by the Oxford group, which was runner-up at the 2014-ILSVRC competition, made the receptive field much smaller in comparison to that of AlexNet but, with increased volume. In VGG, depth was increased from 9 layers to 16, by making the volume of features maps double at each layer. In the same year, GoogleNet that won 2014-ILSVRC competition, not only exerted its efforts to reduce computational cost by changing layer design, but also widened the width in compliance with depth to improve CNN performance. GoogleNet introduced the concept of split, transform, and merge based blocks, within which multiscale and multilevel transformation is incorporated to capture both local and global information.

E. 2015-Present: Rapid increase in Architectural Innovations and Applications of CNN

It is generally observed the major improvements in CNN performance occurred from 2015-2019. The research in CNN is still on going and has a significant potential of improvement. Representational capacity of CNN depends on its depth and in a sense can help in learning complex problems by defining diverse level of features ranging from simple to complex. Multiple levels of transformation make learning easy by chopping complex problems into smaller modules. However, the main challenge faced by deep architectures is the problem of negative learning, which occurs due to diminishing gradient at lower layers of the network. To handle this problem, different research groups worked on readjustment of layers connections and design of new modules. In earlier 2015 used the concept of cross-channel connectivity and information gating mechanism to solve the vanishing gradient problem and to improve the network representational capacity . This idea got famous in late 2015 and a similar concept of residual blocks or skip connections was coined . Residual blocks are a variant of cross-channel connectivity, which smoothen learning by regularizing the flow of information across blocks. This idea was used in ResNet architecture for the training of 150 layers deep network . The idea of cross-channel connectivity is further extended to multilayer connectivity by Deluge, DenseNet, etc. to improve representation.

IV. Types of Nets

A. LeNet

LeNet was proposed by LeCuN in 1998. It is famous due to its historical importance as it was the first CNN,which showed state-of-the-art performance on hand digit recognition tasks.It has the ability to classify digits without being affected by small distortions, rotation, and variation of position and scale. LeNet is a feed-forward NN that constitutes of five alternating layers of convolutional and pooling, followed by two fully connected layers. In early 2000,GPU was not commonly used to speedup training, and even CPUs were slow.The main limitation of traditional multilayer fully connected NNwas that it consider seach pixel as a separate input and applies transformation on it, which was a huge computational burden, specifically at that time. LeNet exploited the underlying basis ofimage that the neighboring pixels are correlated to each other and are distributed across the entire image. Therefore,convolution with learnable parameters isan effective way to extract similar features at multiple locations with few parameters. This changedthe 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 and computation but was able to automatically learn features

B. AlexNet

LeNet though begin the history of deep CNNs but at that time,CNN was limited to hand digit recognition tasks,and didn't scale well to all classes of images. AlexNet is considered as the first deep CNN architecture,which showed ground breaking results for image classification and recognition tasks.AlexNet was proposed by Krizheveskyet al.,who enhanced the learning capacity of the CNN by making it deeper and by applying a number of parameter optimizations strategies[21]. Basic architectural design of AlexNet is shown in Fig.5.In early 2000, hardware limitations curtailed the learning capacity of deep CNN architecture by restricting them to small size. In order to get benefit of the representational capacity of CNN, Alexnet was trained in parallel on two NVIDIA GTX 580 GPUsto overcome shortcomings of the hardware.In AlexNet, feature extraction stages were extended from 5 (LeNet) to 7 to make CNN applicable for diverse categories of images.Despite the fact that generally, depth improves generalization for different resolutions of images but,the main drawback associated with increase in depth is overfitting. To address this challenge, Krizheveskyet al.(2012) exploited the idea of Hinton whereby their algorithm randomly skips some transformational units during training to enforce the model to learn features that are more robust. In addition to this, ReLUwas employed as a non-saturating activation function to improve the convergence rate by alleviating the problem of vanishing gradient to some extent.Overlapping subsampling and local response normalization were also applied to improvethe generalization by reducing overfitting.Other adjustments made were the use of large size filters (11x11and 5x5)at the initial layers, compared to previously proposed networks. Due to efficient learning approach of AlexNet, it has a significant importance in the new generation of CNNs and has started a new era of research in the architectural advancements of CNNs

C. ZefNet

Learning mechanism of CNN,before 2013, was largely based on hit-and-trial, without knowing the exact reason behind the improvement. This lack of understanding limited the performance of deep CNNs on complex images. In 2013, Zeilerand Fergus proposed an interesting multilayer Deconvolutional NN(DeconvNet), which got famousas ZefNet. ZefNet was developed to quantitatively visualize network performance. The idea of the visualization of network activity

was to monitor CNN performance by interpreting neuron's activation. In one of the previous studies exploited the same idea and optimized performance of DeepBelief Networks (DBNs)by visualizing hidden layers'feature. In the same manner evaluated the performance of deep unsupervised auto encoder (AE) by visualizing the image classes generated by the output neurons.DeconvNetworks in the same manner as the forward pass CNN but,reverses the order of convolution a land pooling operation. This reverse mapping projects the output of convolutional layer back to visually perceptible image patterns consequently gives the neuron-level interpretation of the internal feature representation learned at each layer. The objective of ZefNetwas to monitor the learning scheme during training and thus use the findings in diagnosing a potential problem associated with the model. This idea was experimentally validated on AlexNet using DeconvNet,which showed that only a few neurons were active, while other neuronswere dead(inactive)in the first and second layer of the network.Moreover, it showed that the features extracted by the second layer exhibited aliasing artifacts. Based on these findings,Zeiler and Fergus adjusted CNN topology and performed parameter optimization. Zeiler and Fergus maximized the learning of CNN by reducing both the filter size and stride to retain maximum number of features in the first two convolutional layers.This read just in CNN topology resulted in performance improvement,which suggested that features visualization can be used for identification of design shortcomings and for timely adjustment of parameters

D. VGG

With the successful use of CNNs for image recognition, Simonyanet al.proposed a simple and effective design principle for CNN architectures. Their architecture named as VGG was modular in layers pattern. VGG was made 19 layers deep compared to AlexNet and ZefNetto simulate the relationship of depth with the representational capacity of the network.ZefNet, which was a front line network of 2013-ILSVRC competition,suggested that small size filters can improve the performance of the CNNs. Based on these findings, VGG replaced the 11x11 and 5x5filters with a stack of 3x3 filters layer and experimentally demonstrated that concurrent placement of 3x3 filters can induce the effect of the large size filter(receptive field as effective as that of large size filters(5x5 and 7x7)).Use of the small size filters provide an additional benefit of low computational complexity by reducing the number of parameters. These findings set a new trend in research to work with smaller size filters in CNN.VGG regulates complexity of network by placing 1x1 convolution in between the convolutional layers, which in addition,learn a linear combination of the resultant feature maps. For the tuning of the network,max pooling is placed after the convolutional layer,while adding was performed to maintain the spatial resolution. VGG showed good resultsboth for image classification and localization problems. Although, VGG was not at the top place of2014-ILSVRC competition but,got fame due to its simplicity, homogeneous topology,and increased depth. The main limitation associated with VGG was that of high computational cost. Even with the use of small size filters, VGG suffered from high computational burden due to the use of about 140 million parameters

E. GoogleNet

GoogleNet was the winner of the 2014-ILSVRC competition and is also known as Inception-V1. The main objective of the Google Net architecture was to achieve high accuracy with a reduced computational cost. It introduced the new concept of inception block in CNN, whereby it incorporates multi-scale convolutional transformations using split, transform,and merge idea. The architecture of inception block is shown in Fig.6. This block encapsulates filters of different sizes (1x1, 3x3, and 5x5) to capture spatial information at different scales (both at fine and coarse grain level).In GoogleNet, conventional convolutional layers are replaced in small blocks similar to the idea of substituting each layer with micro NNas proposed in Network in Network (NIN) architecture. 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 different images.In addition to the improvement in learning capacity, GoogleNet focus was to make CNN parameter efficient. GoogleNet regulates the computation by adding a bottle neck layer with a1x1 convolutional filter, before employing large size kernels. It used sparse connections (not all the output feature maps are connected to all the input feature maps), to overcome the problem of redundant information and reduced cost by omitting feature maps (channels)that were not relevant. Furthermore, connection's density was reduced by using global average pooling at the last layer,instead of using a fully connected layer. These parameter tuning caused a significant decrease in the number of parameters from 40 million to 5 million parameters. Other regulatory factors applied were batch normalization and use of Rms Prop as an optimizer[129]. 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 that needs to be customized from module to module. Another, limitation of GoogleNet was a representation bottleneck that drastically reduces the feature space in the next layer and thus sometimes may lead to loss of useful information

F. ResNet

ResNet is considered as a continuation of deepNets. ResNet revolutionized the CNN architectural race by introducing the conceptof residual learning in CNN and devised an efficient methodologyfor training of deep Nets. Similar to Highway Networks, it is also placed under the Multi-Path based CNNs, thus its learning methodology is discussedin Section 4.3.2.ResNet proposed 152-layers deep CNN, which won the 2015 ILSVRC competition. Architecture of the residual block of ResNet is shown in . ResNet, which was 20 and 8 times deeper than AlexNetand VGG respectively,showed less computational complexity than previously proposed Nets

empirically showed that ResNet with 50/101/152 layers has lesser roron image classification task than 34 layer explain Net. Moreover, ResNet gained 28% improvement on the famous image recognition benchmark dataset named as COCO. Good performance of ResNet on image recognition and localization tasks showed that depth is of central importance for many visual recognition tasks.

G. ResNext

ResNext, also known as Aggregated Residual Transform Network, is an improvement over the Inception Network exploited the concept of the split, transform and merge in a powerful but simple way by introducing a new term cardinality . Cardinality is an additional dimension, which refers to the size of the set of transformations . Inception network has not only improved learning capability of conventional CNNs but also makes a network resource effective. However, due to the use of diverse spatial embedding's (such as use of 3x3, 5x5 and 1x1 filter) in the transformation branch, each layer needs to be customized separately. In fact, ResNext derives characteristic features from Inception, VGG, and ResNet. ResNext utilized the deep homogenous topology of VGG and simplified GoogleNet architecture by fixing spatial resolution to3x3 filters with in the split, transform, and merge block. It also uses residual learning. Building block for ResNext is shown in Fig.8.ResNext used multiple transformations within a split, transform and merge block and defined these transformations in terms of cardinality (2017) showed that increase in cardinality significantly improves the performance. The complexity of ResNext was regulated by applying low embedding's (1x1 filters) before 3x3 convolution. Whereas training was optimized by using skip connections.

H. DenseNets

In continuation of Highway Networks and ResNet, DenseNet was proposed to solve the vanishing gradient problem The problem with ResNet was that it explicitly preserves information through additive identity transformations due to which many layers may contribute very little or no information. To address this problem, DenseNet used cross-layer connectivity but,in a modified fashion.DenseNet connected each layer to every other layer in a feed-forward fashion, thus feature maps of all preceding layers were used as inputs into all subsequent layers. This establishes connections in DenseNet,as compared to connections between a layer and its preceding layer in the traditional CNNs. It imprints the effect of cross-layer depth wise convolutions. As DenseNet concatenates the previous layers features instead of adding them, thus, the net work may gain the ability to explicitly differentiate between information that is added to the network and information that is preserved. DenseNet has narrow layer structure;however, it becomes parametrically expensive with an increase in a number of feature maps. The direct admittance of each layer to the gradients through the loss function improves the flow of information throughout the network. This incorporates a regularizing effect, which reduces overfitting on tasks with smaller training sets.

I. WideResNet

It is concerned that the main drawback associated with deep residual networks is the feature reuse problem in which some feature transformations or blocks may contribute very little to learning. This problem was addressed by WideResNet. Zagoruykoand Komodakis suggested that the main learning potential of deep residual networks is due to the residual units, whereas depth has a supplement any effect. WideResNet exploited the power of the residual blocks by making ResNet wide rather than deep. WideResNet increased width by introducing an additional factor k,which controls the width of the network. WideResNet showed that the widening of the layers may provide a more effective way of performance improvement than by making the residual networks deep. Although,deep residual networks improved representational capacity,but they have some demerits such as time intensive training, inactivation of many feature maps (feature reuse problem), and gradient vanishing and exploding problem. Here addressed feature reuse problem by incorporating dropout in residual blocks to regularize network in an effective way. Similarly introduced the concept of stochastic depth by exploiting dropouts to solve vanishing gradient and slow learning problem. It was observed that even fraction improvement in performance may require the addition of many new layers. An empirical study showed that WideResNet was twice the number of parameters as compared to ResNet,but can be trained in a better way than the deep networks. Wider residual network was based on the observation that almost all architectures before residual networks, including the most successful Inception and VGG, were wider as compared to ResNet. In WideResNet, learning is made effective by adding a dropout in-between the convolutional layers rather than inside a residual block

J. PyramidalNet

In earlier deep CNN architectures such as AlexNet, VGG,and ResNet, due to the deep stacking of multiple convolutional layers, depth of feature maps increase in subsequent layers. However, the spatial dimension decreases, as each convolutional layer is followed by a sub-sampling layer . Therefore, argued that in deep CNNs,enriched feature representation is compensated by a decrease in feature map size. The drastic increase in the feature map depth and at the same time the loss of spatial information limits the learning ability of CNN.ResNethas shown remarkable results for image classification problem. However in ResNet,the deletion of residual block,where dimension of both spatial and feature map (channel)varies (feature map depth increases, while spatial dimension decreases),generally deteriorates performance.In this regard, stochastic ResNet,improved the performance by reducing information loss associated with the droppingof the residual unit[105]. To increase the learning ability of ResNet,pproposed,Pyramidal Net. In contrast to the

drastic decrease in spatial width with an increase in depth by ResNet, Pyramidal Net increases the width gradually per residual unit. This strategy enables pyramidal Net to cover all possible locations instead of maintaining the same spatial dimension within each residual block until down-sampling occurs. Because of gradual increase in the depth of features map in a top-down fashion, it was named as pyramidal Net.

V. Applications

The Basic purpose of CNN is to say what the the object is? and where is it?. Application of CNN are everywhere from camera in your mobile to various other devices which we daily encounter.Some of the domain where CNN is used are-
i. Mobile Phone - Gesture control ,Camera all make use of CNN for their Functioning.
ii. Surveillance - CNN in Surveillance is used for object recognition ,object detection ,people detection etc. Consider scenario of Airport, here we need CNN to keep a track of who and what is coming in and going out of the airport.
iii. Automotive industry - CNN plays a very important part in auto driving car.It provides the onboard system with the live feed of the outside world i.e. Roads, Sideways, Pedestrian based on the information the system will direct the car.
iv. AR-VR - Google lens that we use is a CNN based Application which can grab any kind of information from an image. It can also measure the dimension of the room ,can create a virtual object in the room which can be moved.= all using CNN.

VI. Issues

A portion of the open issues in the territory of Convolutional Neural Systems are talked about here yet most issues are managed in numerous continuous works. One of the significant issues is that preparation of CNN requires tuning of an enormous number of parameters driving to experimentation of the model engineering.

VII. Conclusion

In this survey, we examined the evolution of convolutional neural networks from its basic components to various types. Today, CNN sees many applications such as face detection and image, video recognition and voice recognition as a powerful full tool within machine learning.

References

[1] O. Abdel-Hamid, A. Mohamed, H. Jiang, and G. Penn. Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4277–4280, March 2012.

[2] S. S. Ahranjany, F. Razzazi, and M. H. Ghassemian. A very high accuracy handwritten character recognition system for farsi/arabic digits using convolutional neural networks. In *2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*, pages 1585–1592, Sep. 2010.

[3] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Transactions on Medical Imaging*, 35(5):1207–1216, May 2016.

[4] Bao-Qing Li and Baoxin Li. Building pattern classifiers using convolutional neural networks. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)*, volume 5, pages 3081–3085 vol.5, July 1999.

[5] V. Bevilacqua, G. Tattoli, D. Buongiorno, C. Loconsole, D. Leonardis, M. Barsotti, A. Frisoli, and M. Bergamasco. A novel bci-ssvep based approach for control of walking in virtual environment using a convolutional neural network. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 4121–4128, July 2014.

[6] A. Bilal, A. Jourabloo, M. Ye, X. Liu, and L. Ren. Do convolutional neural networks learn class hierarchy? *IEEE Transactions on Visualization and Computer Graphics*, 24(1):152–162, Jan 2018.

[7] Caihua Liu, Jie Liu, Fang Yu, Yalou Huang, and Jimeng Chen. Handwritten character recognition with sequential convolutional neural network. In *2013 International Conference on Machine Learning and Cybernetics*, volume 01, pages 291–296, July 2013.

[8] J. A. Calderon-Martinez and P. Campoy-Cervera. A convolutional neural architecture: an application for defects detection in continuous manufacturing systems. In *2003 IEEE International Symposium on Circuits and Systems (ISCAS)*, volume 5, pages V–V, May 2003.

[9] H. Cecotti and A. Graser. Convolutional neural networks for p300 detection with application to brain-computer interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):433–445, March 2011.

[10] Chao Yan, B. Zhang, and F. Coenen. Driving posture recognition by convolutional neural networks. In *2015 11th International Conference on Natural Computation (ICNC)*, pages 680–685, Aug 2015.

[11] J. Chen, X. Kang, Y. Liu, and Z. J. Wang. Median filtering forensics based on convolutional neural networks. *IEEE Signal Processing Letters*, 22(11):1849–1853, Nov 2015.

[12] Y. Chen, T. Krishna, J. S. Emer, and V. Sze. Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks. *IEEE Journal of Solid-State Circuits*, 52(1):127–138, Jan 2017.

[13] K. Cheng, Y. Chen, and W. Fang. Improved object detection with iterative localization refinement in convolutional neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9):2261–2275, Sep. 2018.

[14] J. Chung and K. Sohn. Image-based learning to measure traffic density using a deep convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, 19(5):1670–1675, May 2018.

[15] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber. Convolutional neural network committees for handwritten character classification. In *2011 International Conference on Document Analysis and Recognition*, pages 1135–1139, Sep. 2011.

[16] M. Coşkun, A. Uçar, Ö. Yildirim, and Y. Demir. Face recognition based on convolutional neural network. In *2017 International Conference on Modern Electrical and Energy Systems (MEES)*, pages 376–379, Nov 2017.

[17] H. Deng, G. Stathopoulos, and C. Y. Suen. Error-correcting output coding for the convolutional neural network for optical character recognition. In *2009 10th International Conference on Document Analysis and Recognition*, pages 581–585, July 2009.

[18] C. Desai, J. Eledath, H. Sawhney, and M. Bansal. De-correlating cnn features for generative classification. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 428–435, Jan 2015.

[19] C. Ding and D. Tao. Trunk-branch ensemble convolutional neural networks for video-based face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):1002–1014, April 2018.

[20] J. Ding, B. Chen, H. Liu, and M. Huang. Convolutional neural network with data augmentation for sar target recognition. *IEEE Geoscience and Remote Sensing Letters*, 13(3):364–368, March 2016.

[21] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P. Heng. Automatic detection of cerebral microbleeds from mr images via 3d convolutional neural networks. *IEEE Transactions on Medical Imaging*, 35(5):1182–1195, May 2016.

[22] L. Du, Y. Du, Y. Li, J. Su, Y. Kuan, C. Liu, and M. F. Chang. A reconfigurable streaming deep convolutional neural network accelerator for internet of things. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 65(1):198–208, Jan 2018.

[23] M. Elhoseiny, S. Huang, and A. Elgammal. Weather classification with deep convolutional neural networks. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 3349–3353, Sep. 2015.

[24] J. Fan, W. Xu, Y. Wu, and Y. Gong. Human tracking using convolutional neural networks. *IEEE Transactions on Neural Networks*, 21(10):1610–1623, Oct 2010.

[25] B. Fasel. Facial expression analysis using shape and motion information extracted by convolutional neural networks. In *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pages 607–616, Sep. 2002.

[26] B. Fasel. Head-pose invariant facial expression recognition using convolutional neural networks. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pages 529–534, Oct 2002.

[27] B. Fasel. Robust face analysis using convolutional neural networks. In *Object recognition supported by user interaction for service robots*, volume 2, pages 40–43 vol.2, Aug 2002.

[28] L. Fedorovici, R. Precup, F. Dragan, R. David, and C. Purcaru. Embedding gravitational search algorithms in convolutional neural networks for ocr applications. In *2012 7th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 125–130, May 2012.

[29] L. Fedorovici, R. Precup, F. Dragan, and C. Purcaru. Evolutionary optimization-based training of convolutional neural networks for ocr applications. In *2013 17th International Conference on System Theory, Control and Computing (ICSTCC)*, pages 207–212, Oct 2013.

[30] Kunihiko Fukushima. *Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position*. 1980.

[31] D. Garbin, E. Vianello, O. Bichler, Q. Rafhay, C. Gamrat, G. Ghibaudo, B. DeSalvo, and L. Perniola. Hfo2-based oxram devices as synapses for convolutional neural networks. *IEEE Transactions on Electron Devices*, 62(8):2494–2501, Aug 2015.

[32] C. Han, C. Hsieh, Y. Chen, G. Ho, K. Fan, and C. Tsai. License plate detection and recognition using a dual-camera module in a large space. In *2007 41st Annual IEEE International Carnahan Conference on Security Technology*, pages 307–312, Oct 2007.

[33] N. Hatipolu and Gökhan Bilgin. Segmentation of histopathological images with convolutional neural networks using fourier features. *2015 23rd Signal Processing and Communications Applications Conference, SIU 2015 - Proceedings*, pages 455–458, 06 2015.

[34] Ho-Joon Kim, J. S. Lee, and J. Park. Dynamic hand gesture recognition using a cnn model with 3d receptive fields. In *2008 International Conference on Neural Networks and Signal Processing*, pages 14–19, June 2008.

[35] J. Huang, J. Li, and Y. Gong. An analysis of convolutional neural networks for speech recognition. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4989–4993, April 2015.

[36] Y. Huang, R. Wu, Y. Sun, W. Wang, and X. Ding. Vehicle logo recognition system based on convolutional neural networks with a pretraining strategy. *IEEE Transactions on Intelligent Transportation Systems*, 16(4):1951–1960, Aug 2015.

[37] S. Ji, W. Xu, M. Yang, and K. Yu. 3d convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):221–231, Jan 2013.

[38] J. Jin, K. Fu, and C. Zhang. Traffic sign recognition with hinge loss trained convolutional neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 15(5):1991–2000, Oct 2014.

[39] K. Kang, H. Li, J. Yan, X. Zeng, B. Yang, T. Xiao, C. Zhang, Z. Wang, R. Wang, X. Wang, and W. Ouyang. T-cnn: Tubelets with convolutional neural networks for object detection from videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(10):2896–2907, Oct 2018.

[40] L. Kang, P. Ye, Y. Li, and D. Doermann. Convolutional neural networks for no-reference image quality assessment. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1733–1740, June 2014.

[41] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos. Video super-resolution with convolutional neural networks. *IEEE Transactions on Computational Imaging*, 2(2):109–122, June 2016.

[42] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1725–1732, June 2014.

[43] M. Khalil-Hani and L. S. Sung. A convolutional neural network approach for face verification. In *2014 International Conference on High Performance Computing Simulation (HPCS)*, pages 707–714, July 2014.

[44] S. Kiranyaz, T. Ince, and M. Gabbouj. Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675, March 2016.

[45] S. Lawrence, C. L. Giles, Ah Chung Tsoi, and A. D. Back. Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8(1):98–113, Jan 1997.

[46] P. Le Callet, C. Viard-Gaudin, and D. Barba. A convolutional neural network approach for objective video quality assessment. *IEEE Transactions on Neural Networks*, 17(5):1316–1327, Sep. 2006.

[47] Y. Le Cun and Y. Bengio. *Word-level training of a handwritten word recognizer based on convolutional neural networks*, volume 2. Oct 1994.

[48] S. Lee, H. Zhang, and D. J. Crandall. Predicting geo-informative attributes in large-scale image collections using convolutional neural networks. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 550–557, Jan 2015.

[49] C. Li, N. H. C. Yung, and E. Y. Lam. Human arm pose modeling with learned features using joint convolutional neural network. In *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, pages 398–401, May 2015.

[50] R. Li, Q. Liu, J. Gui, D. Gu, and H. Hu. Indoor relocalization in challenging environments with dual-stream convolutional neural networks. *IEEE Transactions on Automation Science and Engineering*, 15(2):651–662, April 2018.

[51] X. Li, S. Qian, F. Peng, J. Yang, X. Hu, and R. Xia. Deep convolutional neural network and multi-view stacking ensemble in ali mobile recommendation algorithm competition: The solution to the winning of ali mobile recommendation algorithm. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1055–1062, Nov 2015.

[52] Y. Li, D. Liu, H. Li, L. Li, F. Wu, H. Zhang, and H. Yang. Convolutional neural network-based block up-sampling for intra frame coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(9):2316–2330, Sep. 2018.

[53] F. Liu, Y. Huang, W. Yang, and C. Sun. High-level spatial modeling in convolutional neural network with application to pedestrian detection. pages 778–783, May 2015.

[54] G. Lopez-Risueno, J. Grajal, and R. Diaz-Oliver. Target detection in sea clutter using convolutional neural networks. In *Proceedings of the 2003 IEEE Radar Conference (Cat. No. 03CH37474)*, pages 321–328, May 2003.

[55] T. X. Luong, B. Kim, and S. Lee. Color image processing based on nonnegative matrix factorization with convolutional neural network. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 2130–2135, July 2014.

[56] G. Lv. Recognition of multi-fontstyle characters based on convolutional neural network. In *2011 Fourth International Symposium on Computational Intelligence and Design*, volume 2, pages 223–225, Oct 2011.

[57] D. Maturana and S. Scherer. 3d convolutional neural networks for landing zone detection from lidar. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3471–3478, May 2015.

[58] I. Mrázová, J. Pihera, and J. Velemínská. Can n-dimensional convolutional neural networks distinguish men and women better than humans do? In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, Aug 2013.

[59] C. Nebauer. Evaluation of convolutional neural networks for visual recognition. *IEEE Transactions on Neural Networks*, 9(4):685–696, July 1998.

[60] M. Oquab, L. Bottou, I. Laptev, and J. Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1717–1724, June 2014.

[61] R. Oullette, M. Browne, and K. Hirasawa. Genetic algorithm optimization of a convolutional neural network for autonomous crack detection. In *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*, volume 1, pages 516–521 Vol.1, June 2004.

[62] M. Peemen, A. A. A. Setio, B. Mesman, and H. Corporaal. Memory-centric accelerator design for convolutional neural networks. In *2013 IEEE 31st International Conference on Computer Design (ICCD)*, pages 13–19, Oct 2013.

[63] K. Peng and T. Chen. A framework of extracting multi-scale features using multiple convolutional neural networks. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, June 2015.

[64] S. Pereira, A. Pinto, V. Alves, and C. A. Silva. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Transactions on Medical Imaging*, 35(5):1240–1251, May 2016.

[65] E. Poisson, C. Viard Gaudin, and P. . Lallican. Multi-modular architecture based on convolutional neural networks for online handwritten character recognition. In *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02.*, volume 5, pages 2444–2448 vol.5, Nov 2002.

[66] N. P. Ramaiah, E. P. Ijjina, and C. K. Mohan. Illumination invariant face recognition using convolutional neural networks. In *2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, pages 1–4, Feb 2015.

[67] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: An astounding baseline for recognition. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 512–519, June 2014.

[68] T. N. Sainath, B. Kingsbury, A. Mohamed, G. E. Dahl, G. Saon, H. Soltau, T. Beran, A. Y. Aravkin, and B. Ramabhadran. Improvements to deep convolutional neural networks for lvcsr. In *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 315–320, Dec 2013.

[69] T. N. Sainath, A. Mohamed, B. Kingsbury, and B. Ramabhadran. Deep convolutional neural networks for lvcsr. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8614–8618, May 2013.

[70] M. Sankaradas, V. Jakkula, S. Cadambi, S. Chakradhar, I. Durdanovic, E. Cosatto, and H. P. Graf. A massively parallel coprocessor for convolutional neural networks. In *2009 20th IEEE International Conference on Application-specific Systems, Architectures and Processors*, pages 53–60, July 2009.

[71] J. Schlüter and S. Böck. Improved musical onset detection with convolutional neural networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6979–6983, May 2014.

[72] M. Schwarz, H. Schulz, and S. Behnke. Rgb-d object recognition and pose estimation based on pre-trained convolutional neural network features. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1329–1335, May 2015.

[73] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, May 2016.

[74] H. Soltau, G. Saon, and T. N. Sainath. Joint training of convolutional and non-convolutional neural networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5572–5576, May 2014.

[75] S. T. Soman, A. Nandigam, and V. S. Chakravarthy. An efficient multiclassifier system based on convolutional neural network for offline handwritten telugu character recognition. In *2013 National Conference on Communications (NCC)*, pages 1–5, Feb 2013.

[76] D. Strigl, K. Kofler, and S. Podlipnig. Performance and scalability of gpu-based convolutional neural networks. In *2010 18th Euromicro Conference on Parallel, Distributed and Network-based Processing*, pages 317–324, Feb 2010.

[77] M. Sun, D. Zhang, J. Ren, Z. Wang, and J. S. Jin. Brushstroke based sparse hybrid convolutional neural networks for author classification of chinese ink-wash paintings. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 626–630, Sep. 2015.

[78] P. Swietojanski, A. Ghoshal, and S. Renals. Convolutional neural networks for distant speech recognition. *IEEE Signal Processing Letters*, 21(9):1120–1124, Sep. 2014.

[79] M. Szarvas, U. Sakai, and Jun Ogata. Real-time pedestrian detection using lidar and convolutional neural networks. In *2006 IEEE Intelligent Vehicles Symposium*, pages 213–218, June 2006.

[80] M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogata. Pedestrian detection with convolutional neural networks. In *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, pages 224–229, June 2005.

[81] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, May 2016.

[82] W. Tao, F. Jiang, S. Zhang, J. Ren, W. Shi, W. Zuo, X. Guo, and D. Zhao. An end-to-end compression framework based on convolutional neural networks. In *2017 Data Compression Conference (DCC)*, pages 463–463, April 2017.

[83] S. Thomas, S. Ganapathy, G. Saon, and H. Soltau. Analyzing convolutional neural networks for speech activity detection in mismatched acoustic conditions. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2519–2523, May 2014.

[84] W. Thubsaeng, A. Kawewong, and K. Patanukhom. Vehicle logo detection using convolutional neural network and pyramid of histogram of oriented gradients. In *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 34–39, May 2014.

[85] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez. Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE Transactions on Medical Imaging*, 35(5):1273–1284, May 2016.

[86] R. Wagner, M. Thom, R. Schweiger, G. Palm, and A. Rothermel. Learning convolutional neural networks from few samples. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, Aug 2013.

[87] Q. Wang, J. Gao, and Y. Yuan. A joint convolutional neural networks and context transfer for street scenes labeling. *IEEE Transactions on Intelligent Transportation Systems*, 19(5):1457–1470, May 2018.

[88] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, and Y. Yin. Multi-scale rotation-invariant convolutional neural networks for lung texture classification. *IEEE Journal of Biomedical and Health Informatics*, 22(1):184–195, Jan 2018.

[89] Y. Wei, W. Xia, M. Lin, J. Huang, B. Ni, J. Dong, Y. Zhao, and S. Yan. Hcp: A flexible cnn framework for multi-label image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9):1901–1907, Sep. 2016.

[90] C. Wu, W. Fan, Y. He, J. Sun, and S. Naoi. Cascaded heterogeneous convolutional neural networks for handwritten digit recognition. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 657–660, Nov 2012.

[91] Q. . Wu, Y. L. Cun, L. D. Jackel, and B. . Jeng. *On-line recognition of limited-vocabulary Chinese character using multiple convolutional neural networks*. May 1993.

[92] Y. Wu, Y. Liu, J. Li, H. Liu, and X. Hu. Traffic sign detection based on convolutional neural networks. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, Aug 2013.

[93] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva. Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks. *IEEE/ASME Transactions on Mechatronics*, 23(1):101–110, Feb 2018.

[94] P. Xu and R. Sarikaya. Convolutional neural network based triangular crf for joint intent detection and slot filling. In *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 78–83, Dec 2013.

[95] Q. Xu and L. Zhang. Convolutional neural network with corrupted input. In *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, volume 2, pages 77–80, Aug 2015.

[96] Ying-Nong Chen, Chin-Chuan Han, Cheng-Tzu Wang, Bor-Shenn Jeng, and Kuo-Chin Fan. The application of a convolution neural network on face and license plate detection. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 3, pages 552–555, Aug 2006.

[97] H. Yu, R. Hong, X. Huang, and Z. Wang. Obstacle detection with deep convolutional neural network. In *2013 Sixth International Symposium on Computational Intelligence and Design*, volume 1, pages 265–268, Oct 2013.

[98] A. Yuan, G. Bai, L. Jiao, and Y. Liu. Offline handwritten english character recognition based on convolutional neural network. In *2012 10th IAPR International Workshop on Document Analysis Systems*, pages 125–129, March 2012.

[99] Yunlong Bian, Yuan Dong, Hongliang Bai, Bo Liu, Kai Wang, and Yinan Liu. Reducing structure of deep convolutional neural networks for huawei accurate and fast mobile video annotation challenge. In *2014*

IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pages 1–6, July 2014.

[100] Y. Zhao, Q. Dong, S. Zhang, W. Zhang, H. Chen, X. Jiang, L. Guo, X. Hu, J. Han, and T. Liu. Automatic recognition of fmri-derived functional networks using 3-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 65(9):1975–1984, Sep. 2018.