EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

A Review of Image Classification Algorithms in IoT

Xiaopeng Zheng[1,*], Rayan S Cloutier[2]

[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, P R China
[2]Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada
Email: XZ (zxp@home.hpu.edu.cn) RSC (royancloutier@ieee.org)

Abstract

With the advent of big data era and the enhancement of computing power, Deep Learning has swept the world. Based on Convolutional Neural Network (CNN) image classification technique broke the restriction of classical image classification methods, becoming the dominant algorithm of image classification. How to use CNN for image classification has turned into a hot spot. After systematically studying convolutional neural network and in-depth research of the application of CNN in computer vision, this research briefly introduces the mainstream structural models, strengths and shortcomings, time/space complexity, challenges that may be suffered during model training and associated solutions for image classification. This research also compares and analyzes the differences between different methods and their performance on commonly used data sets. Finally, the shortcomings of Deep Learning methods in image classification and possible future research directions are discussed.

Keywords: IOT, Convolutional Neural Network, Image Classification, Deep Learning.

Received on 12 March 2022, accepted on 21 April 2022, published on 21 April 2022

doi: 10.4108/eetiot.v7i28.562

*Corresponding author. Email: zxp@home.hpu.edu.cn

1. Introduction

1.1 Traditional Algorithms

Image classification means that a target picture is given first, and then an algorithm is used to identify the category of the picture. There are many ways to sort images. In term of various semantics of image, they are sorted into object classification, event classification, emotion classification, and scene classification. The major course of image classification involves image preprocessing [1], image characteristics description and extraction [2] and the structuring of classifier [3]. Preprocessing includes operations such as image filter and size normalization, the mission of which is to treat the image more handily; Image features are the description of obvious property, and suitable characters are picked and effectively extracted through a specific image classification algorithm; Classifier is an algorithm that categorize the target image based on the selected characteristic.

The above process describes what is a traditional image classification method. Its properties are mainly determined by feature extraction and classifier devised. The feature extraction that applied in classical image classification algorithms are manually chosen by humans. Usual image characteristics cover low-level visual features like shape, texture, and color, as well as Scale-invariant feature transform [4], Local Binary Patterns [5], Histogram of oriented gradient [6] and so on. Although these features have universality, they are not pertinent to specific images and division ways, and images of some complex scenes, it is indispensable that can exactly describe target images. Common traditional classifiers consist k-Nearest Neighbor [7], Support Vector Machine [8] and so on. These classifiers are straightforward to carrying out and are more effective for easy image

classification missions, but for certain categories with delicate variance, images severe disturbance and other problems, their accuracy is significantly reduced, that is, classical classifiers are inappropriate for intricate image classification.

1.2. Application of CNN in The Internet of Things

Due to the development of computers and the enhance of computing power, Deep Learning [9-11] has gradually entered our field of vision. Compared with traditional image classification algorithms, it no longer needs to manually extract features from the primeval picture, but autonomously learns characteristics from samples through training. Those features are intimately associated with the classifier. And it addresses the hard puzzle of manual characteristics extraction and classifier picking

This technological breakthrough makes AI more widely used in the field of Internet of Things, such as product quality inspection in manufacturing, remote sensing for environmental management or high-resolution cameras for gathering information on the battlefield. Some of sensors are immobile, and others are applied into moving targets, like satellites, drones, and cars.

In the old days, lots of applications about computer vision were restricted to some enclosed platform. However, when associated with IP connectivity technologies, they invent a fresh group of applications which were previously impossible. Computer vision, plus with IP connectivity, data analytics and AI, will be catalysts for mutually, leading to a significate leap in Internet of Things innovation and application.

2. Convolutional Neural Networks

2.1. Neural Networks

Neural network [12] is a significant machine learning [13] technique and the basis of deep learning. As shown in **Figure 1.** Classical Neural Network, in the input and output layers, the count of nodes is fixed, but the quantity of nodes can be randomly specified in the middle layer [14]; The more important is the connection line (connection between neurons), and the corresponding weight of each connection line needs to be obtained through training [15, 16]. The essence of the neural network is made up of numerous neurons [17]. The specific data flow process in the neural network is shown in Figure 2. Neural network processing. In the figure, x1, x2, and x3 are adopted to represent input 1, input 2, input 3 respectively, and w1, w2, w3 represent weight 1, weight 2, weight 3. The bias is b. The nonlinear function is represented by g (·), and the output is represented by y [18-20]. The course can be represented by the below expressions:

$$y = g\left(w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + b\right) \quad (1)$$

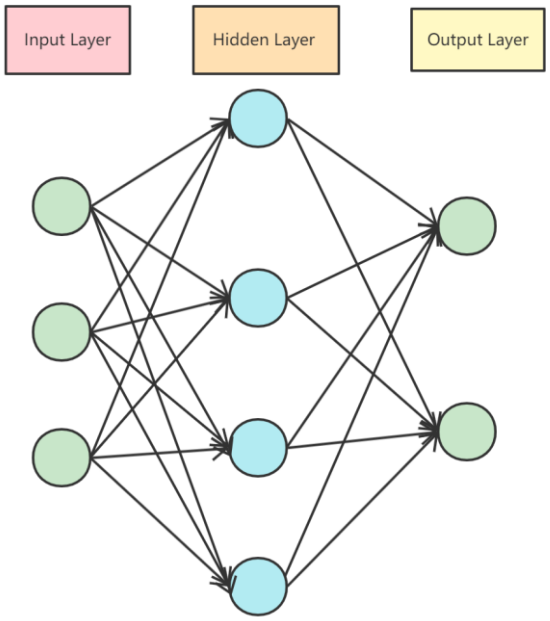

Figure 1. Classical Neural Network

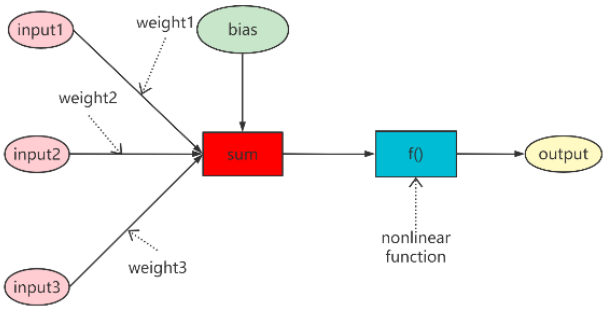

Figure 2. Neural network processing

2.2. Convolutional Neural Networks

Convolutional neural network [21-23] is to add a convolution layer (conv layer) and a pooling layer (pooling layer) based on neural network, and rest hierarchical structures are remains uniform with ordinary neural networks [24]. The flow of its data in the convolutional layer is still illustrated in Figure 1: given an RGB color image (5×5) input into the convolutional layer, the values in parentheses represent the resolution. Then the corresponding input is no longer three values, but three 5 × 5 matrices corresponding to the three-color

channels of the color image [25-27], which are represented by x1, x2 and x3 respectively, and the weights of the CNN are no longer value, but a matrix smaller than the size of the input pixel matrix, so it is called a convolution kernel. Let the size of the convolution kernel be w1, w2 and w3, respectively, and w represents the (2×2) weight matrix [28, 29]. The nonlinear function is represented by G (·), the bias matrix is b, and its output is set to the pixel matrix y [30, 31]. The flow of data in the convolutional layer can be expressed as follows:

$$y = G(w1 \cdot x1 + w2 \cdot x2 + w3 \cdot x3 + b) \quad (2)$$

The parameter sharing mechanism is the biggest feature of the convolutional layer. The weights of the convolution kernels are gained from training, and the weights of the convolution kernels do not alter in the convolution process [32-34]. It means that we can extract the same characteristics in different positions of the initial image through the operation of a convolution kernel.

Pooling layer: Also known as undersampling or downs ampling. It is mainly used for feature dimension reduction , compressing the number of data and parameters [35-37], reducing overfitting, and improving the fault tolerance of t he model. The pooling layer is mainly divided into max p ooling [38] and average pooling. The max pooling is to co rrespond to the region of the filter dimensions on the pictu re, and take the largest value of the pixel in the region to g ain the information related to the characteristics [39-41]. The feature information gained by this technique can pres erves the image better. The average pooling is to take the average value of all the pixels that are not 0 in the above a rea, to obtain the feature data, and this method better retai ns the extraction of the background information of the ima ge. As shown in Figure 3. Pooling operations, a 4×4 imag e matrix is input, and a 2×2 sliding filter is constructed to slide on this image matrix with a step size of 2 [42, 43]. T he role of the sliding filter is to calculate the maximum va lue and the average value of the pixels within its filtering r ange, and finally downsampling the original pixel matrix t o a 2×2 matrix.

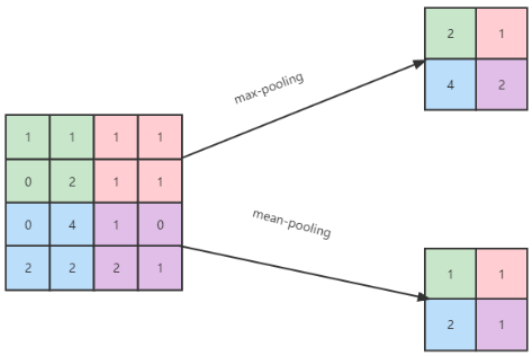

Figure 3. Pooling operations

3. Image Classification Based on Convolutional Neural Networks

The biggest advantage of image classification based on CNN is that extracting certain manual characteristics is not essential for specified image classification methods or classification methods, but to simulate the vision processing system of the brain to abstract the image hierarchy and automatically screen the characteristics [44], for achieving the sorting mission of images. This section first introduces the datasets frequently used in image classification [45], briefly introduces several classic classification models, and finally describes the challenges encountered in the training process and solutions.

3.1. Common Data Sets

The following shows the datasets that are commonly used for classification:

MNIST [46]: The MNIST dataset is one of the most studied datasets in the computer vision and machine learning literature. The goal of this dataset is to correctly classify handwritten digits 0-9, and it itself contains 60,000 training images and 10,000 test images.

CIFA R -10[47]: CIFAR-10 contains 60,000 32*32*3 (RGB) images with a feature vector dimension of 3072, divided into 10 categories. It should be noted that these 10 categories are independent of each other, and each category has 6000 images. Among them, 50,000 images are used for training and 10000 images are used for testing.

CIFA R -100: There are also 60000 color images with an image resolution of 32 × 32, divided into 100 categories, and each category has 600 pictures, covering 500 images for training and 100 images for testing.

ImageNet[48]: The ImageNet dataset is a computer vision dataset founded by Professor Feifei Li of Stanford University, which contains 14,197,122 images and 21,841 Synset indices. A Synset is a node in the WordNet hierarchy, which in turn is a set of synonyms. ImageNet has long been a reference for assessing the capability of image classification algorithms.

3.2. Classical Convolutional Neural Network Models

Typical CNN network structure models commonly used for image classification include LeNet [49], AlexNet [21], GoogLeNet [50], VGGNet[51] and so on. The following is only a brief analysis of CNN's original model and the image classification model that has won the first and second prizes in previous ILSVRC competitions and is more innovative than before, as well as its advantages and disadvantages.

LeNet: LeNet5 was proposed in 1994 and is the fundamental network of the basic network. Through clever design, LeNet5 uses convolution, parameter sharing, pooling and other operations to extract features, avoiding a lot of computational costs, and its network involves a total of 60k parameters. The basic framework of the LeNet is displayed in Figure 4. Convolution process of LeNet. Among them, the convolutional layer extracts spatial features, the pooling layer performs mean downsampling [52-54], and the fully connected layer converts the output of the previous convolutional layer into an overall convolution with a convolution kernel of h ×w, where h and w are the previous The height and width of layer convolution results; Finally, the output layer adopts a soft-max [55] classifier. The model was first applied to digit identification and the result is great. Nevertheless, for the low computational efficiency at that time, the depth of the designed model is relatively shallow, the parameters are not many [56-58], and the structure is relatively simple, so LeNet is not appropriate for more complex image classification missions [59, 60].

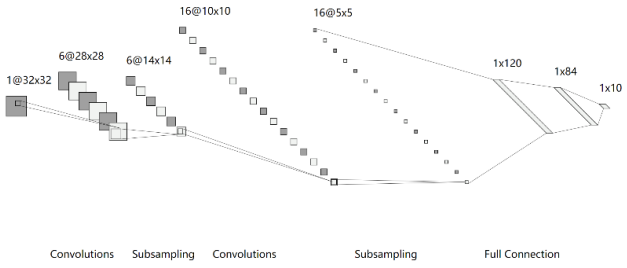

Figure 4. Convolution process of LeNet

AlexNet: This network involves a total of about 60M parameters, and is the ILSVRC2012 champion network. The network structure of AlexNet and LeNet is very similar, but the network has deeper layers and further parameters. The basic framework of AlexNet is shown in

Figure 5. Convolution process of AlexNet. Compared to LeNet, the model applies the ReLU [61] activation function, and the gradient descent is faster than before, so there are less iterations in training. The model also decrease overfitting[62] through dropout[63] and Data Augmentation[64]. Nevertheless, its capacity for describing and extracting image characteristics remains more than limited.

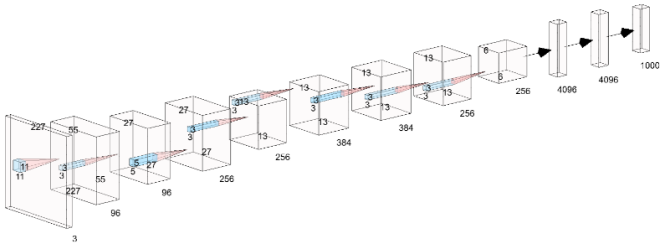

Figure 5. Convolution process of AlexNet

GoogLeNet: GoogLeNet is a deep neural network model based on the Inception module launched by google, and it is also the ILSVRC2014 champion network. The Inception module is shown in

Figure 6. Inception module. There are 4 sections in total. The first part performs 1×1 convolution on the input. It can organize message across channels and enhance the expressive capacity of the network; The second part also applies 1×1 convolution, and then applies 3×3 convolution, which is equivalent to two characteistics conversion; The next part is like the second; The last part is to execute 3 × 3 maxi pooling and then applies 1 × 1 convolution [32]. Before the advent of GoogleNet, our layer often used only one operation, such as convolution or pooling, and the size of the convolution kernel of the convolution operation was also fixed. However, in practical situations, in images of different scales, different sizes of convolution kernels are required to achieve the best performance. For the same image, convolution kernels of different sizes perform differently because their receptive fields are different [65]. Therefore, we hope that the network can choose by itself, and Inception can meet such needs. An Inception module provides a variety of convolution kernel operations in parallel, and the network chooses and uses it by adjusting the parameters during the training process. Compared with the previous network model, the depth of GoogLeNet has been greatly increased, reaching an unprecedented 22 layers. For it has just 1/12 the quantity of parameters of AlexNet, the computational amount of the model is obviously diminished, and the precision of image classification has attained a new level. Although the layers of the GoogLeNet have attained 22, it is highly tough to further deepen the layer, for with the deepening of the model level, the difficulties about gradient dispersion grows more and more severe, which makes the network tough to train.

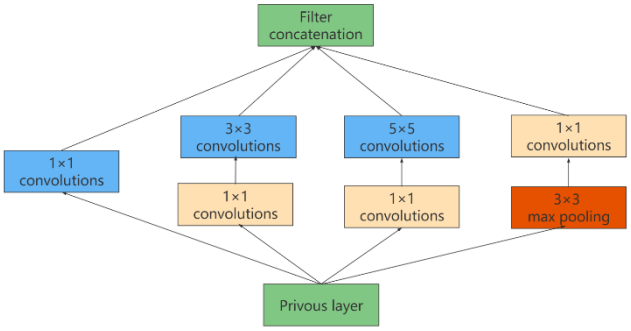

Figure 6. Inception module

VGGNet: This model is the runner-up network of ILSVRC2014. VGG is improved based on AlexNet, which can be seen as a deepened version of AlexNet. And the entire network consists of convolutional layers and fully connected layers [66], VGG uses small convolutional cores (3×3), as shown in Figure 7. Convolution process of VGGNet. In the convolution structure of VGGNet, a 1*1 convolution kernel is imported, and a nonlinear transformation is introduced without affecting the input and output dimensions, which increases the expressiveness of the network and reduces a certain amount of calculation. The model uses a higher number of channels and a wider feature degree. Each channel represents a Feature Map, and more channels represent more abundant image features. The number of channels in the first layer of the VGG network is 64, and each subsequent layer has been doubled to a maximum of 512 channels. The increase in the number of channels allows more information to be extracted. Although the model did not win the championship in ILSVRC2014, its performance is almost the same as the championship.

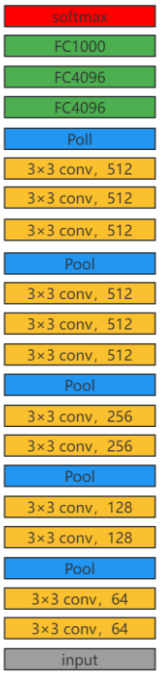

Figure 7. Convolution process of VGGNet

ResNet: This model is the winner network of ILSVRC 2015, and Its appearance is a milestone in the history of CNN images. The model is designed to solve the "degenerate" problem, that is, when the network depth increases, the network accuracy saturates or even declines[67]. The reason why the model cannot get better learning effect is that as the model grows more intricate, the optimization of stochastic gradient descent[68] grows more trouble. Consequently, Dr. He put forward residual learning to solve the degradation problem. As shown in **Figure 8.** Residual module, when the input is represented by x, the learned characteristics are noted as H(x), and currently we expect that it can learn the residual $F(x) = H(x) - x$. In this way, the initial learning characteristic is represented as F(x)+x. The cause is that the residual learning is easier than direct learning from raw characteristic[69]. If the residual is 0, the stacking layer does the identity mapping at this time, at least the network capability will not be degraded, and the residual will not be 0, which will make the stacking layer learn new characteristics depended on the input characteristics, resulting in greater properties.

Why residual learning is relatively easier, intuitively look at the residual learning needs to learn less content, because the residual will generally be relatively small, learning difficulty is smaller. However, we can analyze the questions from a mathematical thinking of view, the residual model can be represented as:

$$y_l = h(x_l) + F(x_l, W_l)$$
$$x_{l+1} = f(y_l) \quad (3)$$

where x_l and x_{l+1} represent the input and output of the l th residual unit, respectively. Note that every residual unit typically involves a multilayer framework. F is the function that means the learned residual. $h(x_l) = x_l$ means the identity map, and f means a ReLU activation function[70]. The learning traits we find from shallow l to deep L can be represented as:

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_i, W_i) \quad (4)$$

adopting chain rules, we can obtain the gradient of the reverse course:

$$\frac{\partial loss}{\partial x_l} = \frac{\partial loss}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_l} = \frac{\partial loss}{\partial x_L} \cdot \left(1 + \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} F(x_i, W_i)\right) \quad (5)$$

The first factor of the equation, $\frac{\partial loss}{\partial x_L}$, means that the loss function attains a gradient of L, the 1 means that the short-circuit mechanism can propagate the gradient without loss, while another residual gradient requires to go through a layer with weights, and the gradient do not pass immediately. The residual gradient will not be all -1, and even though it is relatively little, the presence of 1 will not result in the gradient to lost. Therefore, residual learning is more relaxing than before.

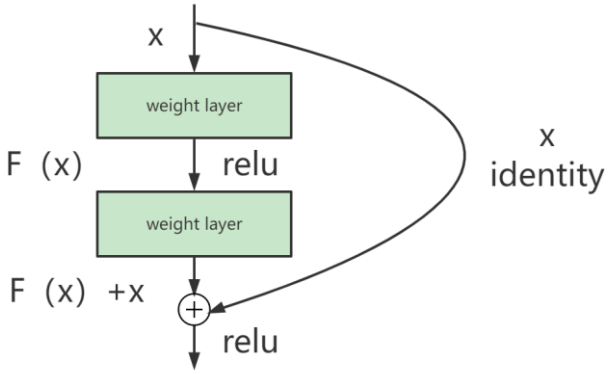

Figure 8. Residual module

SENet [71]:This model is the champion network of ILSVRC 2017, and like the emergence of ResNet. It reduces the error rate of the previous model to a large extent. The full name of SENet is Squeeze-and-Excitation Networks, which is mainly composed of two parts: Squeeze and Excitation. The model is shown in Figure 9. Squeeze-and-Excitation Networks. Among them, what Squeeze does is to compress the size of the original image H*W*C to 1*1*C, which is equivalent to compressing H*W into one dimension. In practice, it is generally realized by global average pooling. After H*W is compressed into one dimension, it is equivalent to obtaining the global view of the previous H*W with this dimension parameter, and the perception area is wider. Excitation will get the 1*1*C representation of Squeeze, add a fully connected layer to predict the importance of each channel, get the importance of different channels, and then apply it to the corresponding channel of the previous feature map, then do the follow-up operation. One of the great advantages of SENet is that it can be easily integrated into existing networks to improve network performance, and the cost is very small.

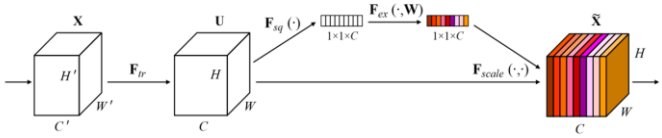

Figure 9. Squeeze-and-Excitation Networks

Mobile Net[72]: Mobile Net came out in 2017 and was created by Google. The original intention of Mobile net is to shine after AlexNet won the ImageNet championship in 2012. Since then, the structure of the network has become deeper and more complex. For mobile and embedded devices, this is clearly inappropriate. Through this paper, the authors detail their design of a lightweight, low-latency network that can be easily applied to mobile and embedded devices.

The main innovation of Mobile Net is to replace ordinary convolution with Depthwise separable convolution, and the core idea of Depthwise separable convolution is to split ordinary convolution into two parts: Depthwise + Pointwise. The Depthwise convolution is shown Figure **10.** Depthwise convolution, which does not do channel fusion after convolution operations. That is, its convolutional kernel only does convolution operations with each layer of the input texture map. The Pointwise convolution is shown in **Figure 11.** Pointwise convolution, which uses a 1*1 convolution kernel to perform channel fusion on the feature map after Depthwise convolution. Among them, Depthwise corresponds to grouped convolution, and Pointwise corresponds to concatenated information.

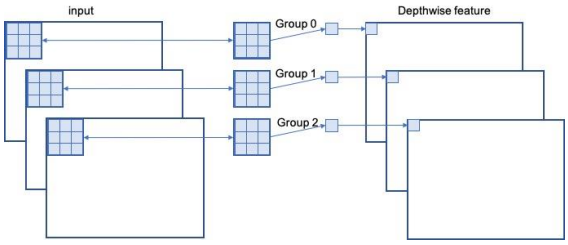

Figure 10. Depthwise convolution

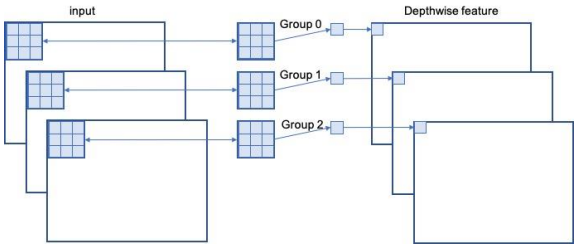

Figure 11. Pointwise convolution

3.3 Training Precautions and Techniques

The following is a brief analysis of the common problems encountered in the training of Convolutional Neural Network models and their solutions.

Overfitting: The model can only fit the data well within the training set, and perform poorly outside the training set[73]. The main reasons for this phenomenon are as follows: (a) the complexity of the model is too high, including too many parameters or over-training; (b) the noise data in the sample interferes too much, causing the model to over-remember the noise features, Instead, it ignores the real input and output features[74]; (c) The amount of data is limited. Therefore, the problem can be solved in three ways. For the problem of too many model parameters, the model depth can be reduced; for the problem of over-training, methods such as pair loss function [75], regularization, random deactivation, etc. can be used. If there is too much noise in the data image, the data can be preprocessed to achieve noise reduction. For the problem of too little data, on the one hand, the data set can be expanded, and on the other hand, the images in the original training set can be flipped, enlarged, and translated and added to the training set again.

Underfitting[76]: The model cannot fit the data well within the training set mainly due to the insufficient depth of the network model. Therefore, this problem can be solved by deepening the layer of the network.

Gradient vanishing [77],Gradient exploding [78]: Since the convolutional neural network adopts the back-propagation method, this method uses chain derivation, and the gradient on the previous layer is the product of the gradient from the latter layer. When the number of network layers is relatively deep, most of the continuous multiplication factor is less than 1, and the final product tends to 0, which is the so-called gradient disappears; Similarly, if most of the continuous multiplication factor is greater than 1, then the final product tends to infinity, that is so-called gradient explosion. For the vanishing gradient problem, the commonly used solutions include: introducing the Residual module used in the ResNet model or improving the activation function. For the exploding gradient problem, in addition to improving the activation function, optimization techniques such as batch normalization (BN) [79-81]and weight regularization can also be performed.

4. Algorithm Comparison

This section compares and analyzes representative CNN models in the references from the perspective of temporal/spatial complexity of CNN models and accuracy on ImageNet datasets.

4.1 Classification Accuracy

This summary is mainly to analyze the error rate of the model described above. As shown in Table 1, the Top-1 error rate indicates that the label learned by the model takes the category with the highest predicted probability as the classification result, so the Top-1 error rate represents the classification result. Among the learned labels, the class with the highest predicted probability is not the ratio of the correct class, and so on. The Top-5 error rate represents the ratio of the five classes with the highest predicted probability of the learned labels that do not contain the correct class. Among them, the second and third columns are the error rates on the validation set, and the fourth column is the error rate on the test set. The test set used is the designated test set for the ILSVRC competition that year, so it is of great reference significance, especially in 2017 The ILSVRC champion network SENet achieved a Top-5 error rate of 2.25% on the test set, a qualitative leap over the previous network performance. From AlexNet in 2012 to SENet in 2017, the error rate dropped from 15.3% to 2.25%, which displays that CNN has developed very fast in recent years, but it is far from reaching the bottleneck.

Table 1. Error rate comparison

Model	Top-1 Error(val,%)	Top-5 Error(val,%)	Top-5 Error(test,%)
AlexNet	36.7	15.4	15.3
GoogleNet	-	7.89	6.66
VGGNet	-	8.43	7.32
ResNet-50	20.74	5.25	-
ResNet-101	19.87	4.60	-
ResNet-152	19.38	4.49	3.57
SENet-154	18.68	4.47	2.25

4.2 FLOPS and Number of Parameters

The time complexity and space complexity of the model are important indicators for measuring the quality of the model. The time complexity is usually measured by the number of floating-point operations per second (FLOPs), and the space complexity is measured by the number of model parameters. The following focuses on the analysis of space complexity. Excessive space complexity means that there are many parameters to be trained, and the amount of data is greatly increased. This undoubtedly puts forward higher requirements for the sample size of the data set, and the ability to characterize the data distribution is also enhanced. Although it is undoubtedly a good thing for classification tasks on some complex datasets, it is also fatal for some simple datasets and may cause serious overfitting problems. Table 2 lists the FLOPs and number of parameters of the above classical models.

Table 2. Time complexity and space complexity

CNN Network Model	FLOPs	Number of parameters
LeNet	4×10^{5}	6×10^{4}
AlexNet	7×10^{9}	6×10^{7}
GoogleNet	1.5×10^{10}	5×10^{6}
VGGNet-16	1.5×10^{11}	1.38×10^{8}
ResNet-50	3.86×10^{10}	2.5×10^{7}
SENet	3.87×10^{10}	2.75×10^{7}

5. Conclusion

Humans have hundreds of millions of neurons, which together constitute the human nervous system. A deep convolutional neural network is also an information processing system, is like the human nervous system. Convolutional neural network is an important method of deep learning, which is widely used in computer vision-related tasks, especially in image classification tasks. This paper focuses on image classification based on deep convolutional neural networks, firstly introduces the basic models from LeNet to ResNet, then introduces the commonly used datasets for image classification experiments, and finally compares and analyzes the performance of classic network models.

Currently, CNN has caused the focus of lot of researchers because of its prominent properties like weight sharing, pooling operations, and multi-layer structures. CNN enters the initial information into the network straightly, and performs learning from the training data, which avoids the weaknesses of manual feature extraction, and its whole classification course is auto. Though these traits have allowed it widely applied, it does not imply that the existing network is free of flaws. It remains an open question about how to train a deep network model with the deeper layers, and there are still many difficulties and challenges that we need to overcome one by one in the future.

References

[1] S. Bhattacharyya, "A Brief Survey of Color Image Preprocessing and Segmentation Techniques," *Journal of Pattern Recognition Research*, vol. 1, no. 1, pp. 120-129, 2011.

[2] M. A. Vega-Rodriguez, *Review: Feature Extraction and Image Processing*. Feature extraction and image processing, 2002.

[3] D. Zhang, B. Liu, C. Sun, and X. Wang, "Learning the Classifier Combination for Image Classification," *Journal of Computers*, vol. 6, no. 8, pp. 1756-1763, 2011.

[4] M. Grabner, H. Grabner, and H. Bischof, "Fast approximated SIFT," in *Asian Conference on Computer Vision*, 2006.

[5] L. He, C. Zou, L. Zhao, and D. Hu, "An Enhanced LBP Feature Based on Facial Expression Recognition," in *IEEE Engineering in Medicine & Biology Conference*, 2005, pp. 3300-3303.

[6] T. T. Do and E. Kijak, "FACE RECOGNITION USING CO-OCCURRENCE HISTOGRAMS OF ORIENTED GRADIENTS," in *IEEE International Conference on Acoustics*, 2012.

[7] G. Amato and F. Falchi, "Local Feature based Image Similarity Functions for kNN Classification," in *ICAART 2011 - Proceedings of the 3rd International Conference on Agents and Artificial Intelligence, Volume 1 - Artificial Intelligence, Rome, Italy, January 28-30, 2011*, 2011.

[8] T. Joachims, "Making Large-Scale SVM Learning

Practical," *Technical Reports*, vol. 8, no. 3, pp. 499-526, 1998.

[9] L. Deng and D. Yu, "Deep Learning: Methods and Applications," *Foundations & Trends in Signal Processing*, vol. 7, no. 3, pp. 197-387, 2014.

[10] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, and A. Y. Ng, "On Optimization Methods for Deep Learning," in *International Conference on Machine Learning*, 2011.

[11] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[12] S. Bhardwaj, S. Tewari, and S. Jain, "Study on Future of Artificial Intelligence in Neural."

[13] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine Learning on Big Data: Opportunities and Challenges," *Neurocomputing*, vol. 237, no. MAY10, pp. 350-361, 2017.

[14] S. C. Satapathy and D. Wu, "Improving ductal carcinoma in situ classification by convolutional neural network with exponential linear unit and rank-based weighted pooling," *Complex Intell. Syst.*, vol. 7, pp. 1295-1310, 2020/11/22 2021.

[15] I. Aliyu, K. J. Gana, A. A. Musa, M. A. Adegboye, and C. G. Lim, "Incorporating Recognition in Catfish Counting Algorithm Using Artificial Neural Network and Geometry," *Ksii Transactions on Internet and Information Systems*, vol. 14, no. 12, pp. 4866-4888, Dec 2020.

[16] T. Vaidya and K. Chatterjee, "Enhancing the stability of active harmonic filter using artificial neural network-based current control scheme," *IET Power Electronics*, vol. 13, no. 19, pp. 4601-4609, Dec 2020.

[17] Y.-D. Zhang, "A five-layer deep convolutional neural network with stochastic pooling for chest CT-based COVID-19 diagnosis," *Machine Vision and Applications*, vol. 32, 2021, Art no. 14.

[18] D. R. Nayak, "Detection of unilateral hearing loss by Stationary Wavelet Entropy," *CNS & Neurological Disorders - Drug Targets*, vol. 16, no. 2, pp. 15-24, 2017.

[19] Y. Jiang, "Exploring a smart pathological brain detection method on pseudo Zernike moment," *Multimedia Tools and Applications*, vol. 77, no. 17, pp. 22589-22604, 2018.

[20] J. Li, "Detection of Left-Sided and Right-Sided Hearing Loss via Fractional Fourier Transform," *Entropy*, vol. 18, no. 5, 2016, Art no. 194.

[21] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in neural information processing systems*, vol. 25, no. 2, 2012.

[22] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "learning and transferring mid-level image representions using convolutional neural networks," 2017.

[23] S. Zagoruyko and N. Komodakis, "Learning to Compare Image Patches via Convolutional Neural Networks," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[24] P. Prakash, A. Banerjee, and P. K. Perepu, "Determination of the relative inclination and the viewing angle of an interacting pair of galaxies using Convolutional Neural Networks," *Monthly Notices of the Royal Astronomical Society*, vol. 497, no. 3, pp. 3323-3334, Sep 2020.

[25] Y. D. Zhang, "A seven-layer convolutional neural network for chest CT based COVID-19 diagnosis using stochastic pooling," *IEEE Sens. J.*, pp. 1-1. doi: 10.1109/JSEN.2020.3025855

[26] D. S. Guttery, "Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network," *Information Processing and Management*, vol. 58, 2, 2021, Art no. 102439.

[27] W. Zhu, "ANC: Attention Network for COVID-19 Explainable Diagnosis Based on Convolutional Block Attention Module," *Computer Modeling in Engineering & Sciences*, vol. 127, 3, pp. 1037-1058, 2021.

[28] H. Uzen and K. Hanbay, "LM Filter-Based Deep Convolutional Neural Network for Pedestrian Attribute Recognition," *Journal of Polytechnic-Politeknik Dergisi*, vol. 23, no. 3, pp. 605-613, Sep 2020.

[29] S. Vagios *et al.*, "THE USE OF DEEP CONVOLUTIONAL NEURAL NETWORKS (CNN) TO OBJECTIVELY ASSESS THE ROLE OF ABSTINENCE ON IVF DEVELOPMENTAL OUTCOMES," *Fertility and Sterility*, vol. 114, no. 3, pp. E141-E141, Sep 2020.

[30] E. Villain *et al.*, "Convolutional neural network for discriminating between Multiple System Atrophy and Healthy Control, comparing MRI modalities and highlighting the disease signature," *Movement Disorders*, vol. 35, pp. S121-S122, Sep 2020.

[31] Q. Abbas, "Object Recognition using Template Matching and Pre-trained convolutional neural network," *International Journal of Computer Science and Network Security*, vol. 20, no. 8, pp. 69-79, Aug 2020.

[32] Q. Zhou, "ADVIAN: Alzheimer's Disease VGG-Inspired Attention Network Based on Convolutional Block Attention Module and Multiple Way Data Augmentation," *Front. Aging Neurosci.*, vol. 13, 2021, Art no. 687456.

[33] Z. Zhang and X. Zhang, "MIDCAN: A multiple input deep convolutional attention network for Covid-19 diagnosis based on chest CT and chest X-ray," *Pattern Recognition Letters*, vol. 150, pp. 8-16, 2021.

[34] Y.-D. Zhang, D. R. Nayak, X. Zhang, and S.-H. Wang, "Diagnosis of secondary pulmonary tuberculosis by an eight-layer improved convolutional neural network with stochastic pooling and hyperparameter optimization," *Journal of Ambient Intelligence and Humanized Computing*, Accessed on: 2020/10/23. doi: 10.1007/s12652-020-02612-9 [Online]. Available: https://doi.org/10.1007/s12652-020-02612-9

[35] J. Jo, H. I. Koo, J. W. Soh, and N. I. Cho, "Handwritten Text Segmentation via End-to-End Learning of Convolutional Neural Networks," *Multimedia Tools and Applications*, vol. 79, no. 43-44, pp. 32137-32150, Nov 2020.

[36] S. K. Ghosh, B. Biswas, and A. Ghosh, "Development of intuitionistic fuzzy special embedded convolutional neural network for mammography enhancement," *Comput. Intell.*, vol. 37, no. 1, pp. 47-69, Feb 2021.

[37] M. C. Kim and J. H. Lee, "Color reproduction in virtual lip makeup using a convolutional neural network," *Color Research and Application*, vol. 45, no. 6, pp. 1190-1201, Dec 2020.

[38] A. Giusti, D. C. Cirean, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Fast Image Scanning with Deep Max-Pooling Convolutional Neural Networks," *IEEE*, 2013.

[39] K. Wu, "SOSPCNN: Structurally Optimized Stochastic Pooling Convolutional Neural Network for Tetralogy of Fallot Recognition," *Wireless Communications and Mobile Computing*, vol. 2021, p. 5792975, 2021/07/02 2021, Art no. 5792975.

[40] V. Govindaraj, "Deep Rank-Based Average Pooling Network for Covid-19 Recognition," *Computers, Materials & Continua*, vol. 70, 2, pp. 2797-2813, 2022.

[41] X. Cheng, "PSSPNN: PatchShuffle Stochastic Pooling Neural Network for an Explainable Diagnosis of COVID-19 with Multiple-Way Data Augmentation," *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021, Art no. 6633755.

[42] H. Shirai *et al.*, "Estimation of the Number of Convallaria Keiskei's Colonies UsingUAVImages Based on a Convolutional Neural Network," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 15, no. 10, pp. 1552-1554, Oct 2020.

[43] R. Abed, S. Bahroun, and E. Zagrouba, "KeyFrame extraction based on face quality measurement and convolutional neural network for efficient face recognition in videos," *Multimedia Tools and Applications*, vol. 80, no. 15, pp. 23157-23179, Jun 2021.

[44] T. Ameen, L. Chen, Z. X. Xu, D. D. Lyu, and H. Y. Shi, "A Convolutional Neural Network and Matrix Factorization-Based Travel Location Recommendation Method Using Community-Contributed Geotagged Photos," *Isprs International Journal of Geo-Information*, vol. 9, no. 8, Aug 2020, Art no. 464.

[45] D. Anderson, "Deep fractional max pooling neural network for COVID-19 recognition," *Frontiers in Public Health*, vol. 9, 2021, Art no. 726144.

[46] M. León, A. Moreno-Báez, R. Magallanes-Quintanar, and R. D. Valdez-Cepeda, "Assessment in subsets of MNIST Handwritten Digits and their Effect in the Recognition Rate," *Journal of Pattern Recognition Research*, vol. 2, no. 2, 2011.

[47] H. Li, H. Liu, X. Ji, G. Li, and L. Shi, "CIFAR10-DVS: An Event-Stream Dataset for Object Classification," *Frontiers in Neuroscience*, vol. 11, 2017.

[48] D. Jia, D. Wei, R. Socher, L. J. Li, L. Kai, and F. F. Li, "ImageNet: A large-scale hierarchical image database," 2009, pp. 248-255.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *IEEE*, 2016.

[50] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, and A. Rabinovich, "Going Deeper with Convolutions," *IEEE Computer Society*, 2014.

[51] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv*, 2014.

[52] M. Jaderberg, A. Vedaldi, and A. Zisserman, "Speeding up Convolutional Neural Networks with Low Rank Expansions," *Computer ence*, vol. 4, no. 4, p. XIII, 2014.

[53] X. Liu, J. Pool, S. Han, and W. J. Dally, "Efficient Sparse-Winograd Convolutional Neural Networks," 2018.

[54] S. Srinivas and R. V. Babu, "Data-free parameter pruning for Deep Neural Networks," *Computer Science*, pp. 2830-2838, 2015.

[55] X. Qi, T. Wang, and J. Liu, "Comparison of Support Vector Machine and Softmax Classifiers in Computer Vision," in *International Conference on Mechanical*, 2018, pp. 151-155.

[56] S. C. Satapathy, "Fruit category classification by fractional Fourier entropy with rotation angle vector grid and stacked sparse autoencoder," *Expert Syst.*, vol. 39, no. 3, 2022, Art no. e12701.

[57] Z. Zhu, "PSCNN: PatchShuffle Convolutional Neural Network for COVID-19 Explainable Diagnosis," *Frontiers in Public Health*, vol. 9, 2021, Art no. 768278.

[58] C. Tang, "Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform," *Multimedia Tools and Applications*, vol. 77, no. 17, pp. 22821-22839, 2018.

[59] S. Purwar, R. Tripathi, A. W. Barwad, and A. K. Dinda, "Detection of Mesangial hypercellularity of MEST-C score in immunoglobulin A-nephropathy using deep convolutional neural network," *Multimedia Tools and Applications*, vol. 79, no. 37-38, pp. 27683-27703, Oct 2020.

[60] M. L. George, T. Govindarajan, K. A. Rajasekaran, and S. R. Bandi, "A robust similarity based deep siamese convolutional neural network for gait recognition across views," *Comput. Intell.*, vol. 36, no. 3, pp. 1290-1319, Aug 2020.

[61] A. B. Jiang and W. W. Wang, "Research on optimization of ReLU activation function," *Transducer and Microsystem Technologies*, 2018.

[62] Z. W. Yuan and J. Zhang, "Feature extraction and image retrieval based on AlexNet," in *Eighth International Conference on Digital Image Processing (ICDIP 2016)*, 2016.

[63] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," *JMLR.org*, 2015.

[64] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," 2017.

[65] R. K. Pandey and R. A. Ganesan, "DeepInterpolation: fusion of multiple interpolations and CNN to obtain super-resolution," *IET Image Processing*, 2021.

[66] L. Wang, Y. Xiong, Z. Wang, and Y. Qiao, "Towards Good Practices for Very Deep Two-Stream ConvNets," *Computer Science*, 2015.

[67] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway Networks," *Computer Science*, 2015.

[68] A. Bordes, L. Bottou, and P. Gallinari, "SGD-QN: Careful Quasi-Newton Stochastic Gradient Descent," *Journal of Machine Learning Research*, vol. 10, no. 3, pp. 1737-1754, 2009.

[69] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," 2016.

[70] J. Schmidt-Hieber, "Nonparametric regression using deep neural networks with ReLU activation function," 2017.

[71] H. Jie, S. Li, S. Gang, and S. Albanie, "Squeeze-and-Excitation Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, 2017.

[72] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.

[73] M. Cogswell, F. Ahmed, R. Girshick, L. Zitnick, and D. Batra, "Reducing Overfitting in Deep Networks by Decorrelating Representations," *Computer Science*, 2015.

[74] S. Yeom, I. Giacomelli, M. Fredrikson, and S. Jha, "Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting," in *IEEE Computer Security Foundations Symposium*, 2017.

[75] L. Lin, N. J. Higham, and J. Pan, "Covariance structure regularization via entropy loss function," *Computational Statistics & Data Analysis*, vol. 72, no. 3, pp. 315-327, 2014.

[76] S. Narayan and G. Tagliarini, "An analysis of underfitting in MLP networks," in *IEEE International Joint Conference on Neural Networks*, 2005.

[77] S. Squartini, S. Paolinelli, and F. Piazza, "Comparing Different Recurrent Neural Architectures on a Specific Task from Vanishing Gradient Effect Perspective," in *IEEE International Conference on Networking*, 2006.

[78] R. Pascanu, T. Mikolov, and Y. Bengio, "Understanding the exploding gradient problem," *Arxiv Preprint Arxiv*, 2012.

[79] N. Gajhede, O. Beck, and H. Purwins, "Convolutional Neural Networks with Batch Normalization for Classifying Hi-hat, Snare, and Bass Percussion Sound Samples," *ACM*, pp. 111-115, 2016.

[80] M. Simon, E. Rodner, and J. Denzler, "ImageNet pre-trained models with batch normalization," 2016.

[81] J. Kohler, H. Daneshmand, A. Lucchi, M. Zhou, K. Neymeyr, and T. Hofmann, "Towards a Theoretical Understanding of Batch Normalization," 2018.