Asia Pacific Science Press

Journal of Advances in Engineering and Technology
ISSN(O): 3005-7507
ISSN(P): 3005-7493
https://doi.org/10.62177/jaet.v1i3.25

Application of CNN Classic Model in Modern Image Processing

Zhouyi Wu
Central South University, Changsha Hunan, 410083, China
Corresponding author: Zhouyi Wu
Published: Nov 13, 2024

Abstract: As a deep learning model, convolutional neural network (CNN) has been widely used in the field of modern image processing and has shown excellent performance. From LeNet to AlexNet, to classic models such as VGGNet and ResNet, CNN has achieved remarkable success in tasks such as image classification, object detection and segmentation through multi-level feature extraction and automatic learning capabilities. This paper first explores the basic structure and working principle of CNN, analyzes the advantages and limitations of classic models, and reviews its specific applications in image processing. By introducing the optimization strategies of various models, it further explores the improvement path of CNN in the field of image processing, including model compression, lightweight design and the introduction of new network structures. In short, the continuous optimization of CNN has enabled it to show powerful performance in multiple complex tasks, promoted the rapid development of image processing technology, and provided strong support for applications in more fields.
Keywords: Convolutional Neural Network; Image Classification; Object Detection; Model Optimization; Deep Learning

1. Introduction

1.1 Research Background and Importance

With the rapid development of deep learning, convolutional neural network (CNN) has become one of the core tools in the field of image processing. From image classification to object detection, CNNs outperform traditional methods and have been widely adopted in many practical applications[1]. Since LeCun et al. first proposed CNNs, this technology has achieved remarkable progress through large amounts of data training and powerful computing power[2]. In particular, with the support of large-scale data sets and high-performance hardware, CNNs have demonstrated excellent performance in tasks such as image recognition, object detection, and semantic segmentation, and have been widely used in fields such as medical image processing and autonomous driving[3].

The reason why CNNs have achieved such great success in the field of image processing is mainly due to their ability to automatically extract image features, especially when processing high-dimensional data[4]. Through operations such as convolution and pooling, CNNs can effectively extract multi-level feature representations from raw image data, thereby greatly improving the recognition ability of the model[5]. In recent years, with the further optimization of network structure and the introduction of new architectures, CNNs have been greatly improved in both accuracy and speed. This provides strong technical support for solving practical image processing problems and promotes the rapid development of fields such as image recognition[6].

1.2 Research Objectives

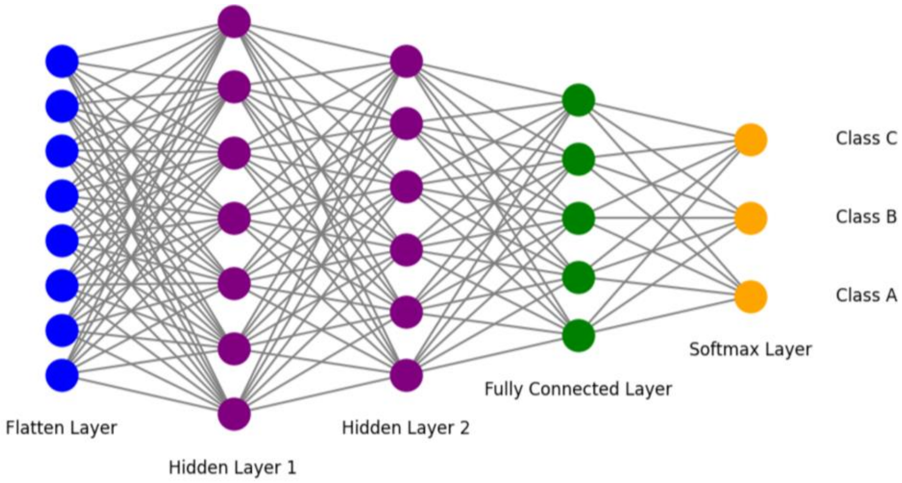
This paper aims to systematically review the application of convolutional neural networks (CNNs) in modern image processing, focusing on classic CNN models such as LeNet, AlexNet, and VGGNet, as well as their applications in image classification, object detection, and segmentation. By reviewing the structure and principles of these classic models, we analyze their advantages and limitations in practical applications and look forward to the future development of CNN. In addition, this paper also explores the latest research progress in the optimization of CNN model structure in recent years, in order to provide a reference for future researchers.

2. Basics of Convolutional Neural Networks

2.1 Basic Structure and Principles of Convolutional Neural Networks

Convolutional Neural Networks (CNN) is a deep learning model specially designed to process grid-like input data (such as images). The core of CNN is to extract the features of input data layer by layer through multiple construction modules such as convolutional layers, pooling layers, and fully connected layers (Figure 1), thereby realizing the modeling and classification of complex data [7]. The basic architecture of CNN usually contains multiple convolutional layers, which extract features through filters (also called convolution kernels). Each convolution kernel can extract local features of the input image, such as edges and corners, and as the number of network layers increases, the extracted features gradually transition from local information to global abstract information[8].

Figure1: Convolutional Neural Network Structure

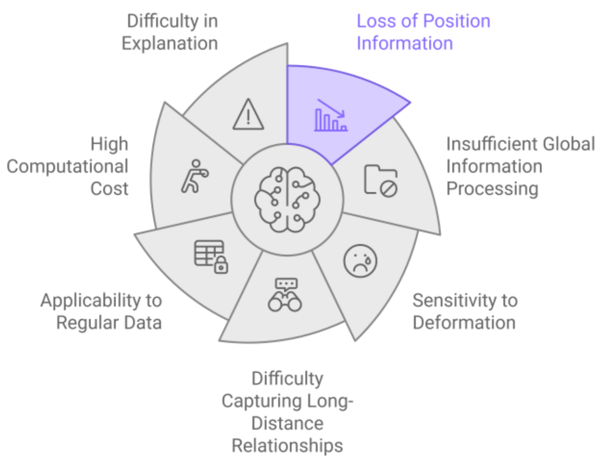


The convolution operation is the core of CNN. It processes the input data locally by moving the convolution kernel, thereby maintaining the spatial structure and greatly reducing the number of parameters. The convolution layer is usually followed by a pooling layer (such as maximum pooling or average pooling), which reduces the size of the feature map by downsampling, thereby reducing the computational complexity and suppressing model overfitting [9]. In addition, the fully connected layer is responsible for classifying the features extracted by the convolution layer and the pooling layer, and finally outputs the prediction results. The back propagation algorithm optimizes the performance of CNN by adjusting the model parameters and improves the accuracy of the model [10].

2.2 Advantages and limitations of classic CNN

CNN has shown significant advantages in processing high-dimensional data such as images, especially in image classification and object detection. It has been widely used in fields such as image classification. First, CNN can significantly reduce the number of network parameters by sharing weights, making model training more efficient and reducing the risk of overfitting [11]. In addition, the local receptive field of the convolution layer can capture the local pattern of the input data, which makes CNN very effective in processing data with spatial structure (such as images) [12]. Another significant advantage is the hierarchical structure of CNN, which can gradually abstract low-level features into high-level concepts, thereby improving the model's ability to handle complex tasks.

Figure2: Limitations of Classical Neural Networks



Nevertheless, CNN also has certain limitations (Figure 2). First, CNN relies on a large amount of labeled data to achieve good learning results, but in many practical scenarios, it is challenging to obtain sufficient labeled data [13]. Second, the network structure of CNN is complex and requires a lot of computing resources. Especially when processing large-scale data sets, the computing time and storage space requirements may be extremely high [14]. In addition, CNN has limited ability to handle rotation, invariance, and scale invariance, and often requires data enhancement and other techniques to improve the generalization performance of the model [15].

3. Overview of Classic CNN Models

3.1 LeNet and AlexNet

LeNet-5 is a convolutional neural network (CNN) proposed by Yann LeCun et al. in the 1990s. It is one of the earliest deep learning models for image recognition. LeNet-5 mainly consists of two convolutional layers and two pooling layers, followed by a fully connected layer and an output layer. It has been successfully applied to tasks such as handwritten digit recognition (MNIST dataset) and has aroused widespread interest in CNN models in the field of computer vision [16]. LeNet has a relatively simple architecture and a small number of parameters, which makes it perform well in the early days with limited computing resources. However, the limitations of LeNet are also obvious, such as its weak ability to process complex and large-scale data, shallow network depth, and difficulty in capturing complex high-dimensional features.

AlexNet was proposed by Krizhevsky et al. in 2012. It greatly improved the performance of image classification by introducing a deeper network structure and ReLU activation function [17]. The success of AlexNet largely depended on the improvement of computing power at the time (especially the widespread use of GPU computing). It processed image classification tasks through 5 convolutional layers and 3 fully connected layers. In the ImageNet Challenge, AlexNet made significant breakthroughs and promoted the rapid development of deep learning in the field of image processing [18]. Compared with LeNet, AlexNet has a deeper network depth and can better capture complex image features. However, AlexNet has a complex structure, a huge number of parameters, and requires powerful computing resources for training, which is also a major challenge for its application.

3.2 VGGNet and Subsequent Models

After AlexNet, VGGNet was proposed by Simonyan and Zisserman, and the number of network layers was further deepened. The most prominent feature of VGGNet is that it uses a smaller 3×3 convolution kernel, but by stacking more convolutional layers (such as VGG16 and VGG19), a deeper network structure is obtained, which can capture more detailed image features[19]. VGGNet performs well in tasks such as image classification, especially achieving high accuracy on the ImageNet dataset. However, the depth of VGGNet also brings a large computational cost, long training time and high memory requirements[20]. Although VGGNet improves the accuracy of the model, its limitations in parameter quantity and computing resource consumption also pose challenges to the optimization of subsequent models.

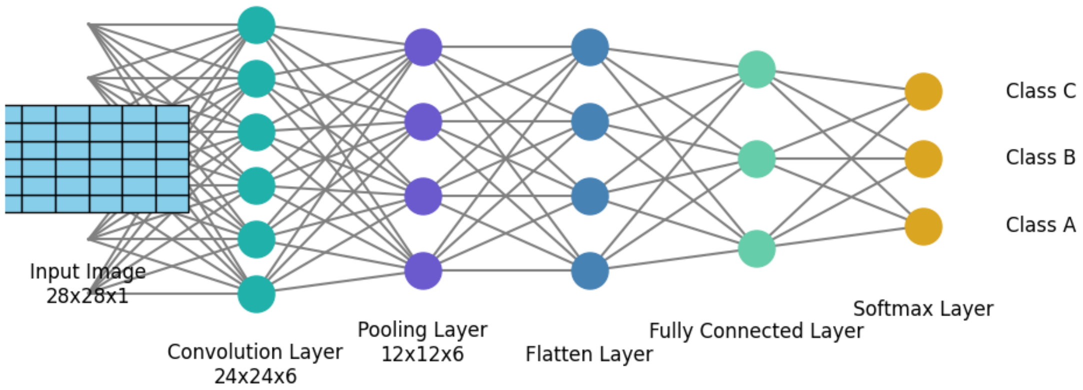
After VGGNet, the emergence of ResNet marked an important innovation in the CNN model. ResNet effectively solves the problem of gradient vanishing and gradient exploding in deep networks by introducing residual connections (skip connections)[21]. This innovation allows the network to be very deep (such as ResNet-50 and ResNet-101) without significantly increasing the difficulty of training. In addition, the structure of ResNet greatly improves the trainability and accuracy of the model, making it dominant in multiple computer vision tasks[22]. With the emergence of models such as ResNet, the application of deep CNN in tasks such as image recognition, target detection, and semantic segmentation has been further expanded, promoting the widespread application of deep learning technology.

4. Application of CNN in Modern Image Processing

4.1 Image Classification and Recognition

Convolutional neural networks (CNNs) are widely used in the field of image classification and recognition and have achieved remarkable success. By stacking multiple convolutional layers and pooling layers, CNNs can extract multi-level feature representations from the original image (Figure 3), thereby achieving efficient classification of complex images. Today, CNNs are widely used in tasks such as face recognition, handwritten digit recognition, and object classification. For example, deep CNNs trained on large image databases (such as ImageNet) can accurately distinguish thousands of different categories of objects and reach or even exceed the recognition level of human experts. Such applications have not only promoted the development of intelligent security, autonomous driving, and other fields, but have also been widely used in medical image analysis, satellite remote sensing image processing, and other fields, greatly improving the accuracy and efficiency of classification and recognition. CNN's automatic feature extraction capability enables it to have excellent generalization and robustness when faced with complex and diverse image data, making it the core method of modern image classification technology.

Figure3: Enhanced Image Processing Neural Network Structure



4.2 Object Detection and Segmentation

In addition to image classification, CNN also performs well in the field of object detection and segmentation. Object detection refers to accurately locating and identifying multiple objects in an image, while segmentation refers to accurately dividing the pixels in an image into different semantic regions. CNN has significantly improved the speed and accuracy of object detection by introducing methods such as region proposal network (RPN) and Faster R-CNN, making it widely used in fields such as autonomous driving and video surveillance. The segmentation task achieves accurate pixel-level segmentation through architectures such as full convolutional network (FCN) and U-Net, which are widely used in application scenarios such as medical image processing and road scene analysis. These technologies can help computers not only identify targets, but also accurately determine the shape and boundaries of targets, thereby achieving more sophisticated image processing and analysis. In modern image processing tasks, object detection and segmentation have become key links and provide important technical support for intelligent applications in multiple fields.

5. Optimization and Future Development of Classic CNN Models

5.1 Model Improvement Strategy

Although the classic convolutional neural network (CNN) model has achieved remarkable results in the field of image processing, with the continuous increase in application requirements, further optimization of the model has become an inevitable trend. First, reducing the parameters and computational complexity of the network is one of the main improvement directions. Although deep models such as VGGNet and ResNet have powerful performance, their huge number of parameters places high demands on hardware resources. To this end, model compression technologies such as weight pruning, quantization, and knowledge distillation are widely used to reduce the size of the model without losing accuracy. Secondly, innovation in network structure is also an important improvement strategy. For example, lightweight networks such as MobileNet and ShuffleNet have significantly improved the computational efficiency of the network by introducing technologies such as deep separable convolution and group convolution, making them more suitable for resource-constrained devices such as mobile applications. Finally, optimizing training methods, such as using adaptive learning rates, enhancing data processing, and applying more regularization techniques (such as Dropout and Batch Normalization), can improve the generalization ability of the model and reduce overfitting, thereby performing more robustly in practical applications.

5.2 Future Research Directions of CNN

With the continuous development of artificial intelligence and deep learning technology, CNN still has many research directions worth exploring in the future. First, multi-task learning and self-supervised learning are one of the hot topics in future CNN research. These technologies can improve the learning ability of the model without large-scale labeled data, and share features in multiple tasks, thereby improving training efficiency. Secondly, the combination of CNN with other models, such as combining CNN with recurrent neural networks (RNN), transformers and other models, is expected to achieve better performance in tasks such as time series data processing and video analysis. In addition, the application of CNN in the processing of three-dimensional data (such as 3D images and point cloud data) will also become a new research direction. Finally, how to further improve the interpretability and controllability of CNN is also a focus of future research. Most of the current CNNs are regarded as "black box" models, and it is difficult to explain their decision-making process. Future research will focus on improving the transparency of the model so that it can be better applied to high-risk fields such as medical diagnosis and autonomous driving.

6. Conclusion

Since its proposal, convolutional neural networks (CNNs) have made great progress and wide application in the field of image processing. From the earliest LeNet to the later AlexNet that promoted the deep learning craze, to the deeper VGGNet and ResNet, the classic CNN model has been continuously optimized and improved to solve many complex image processing problems. Through automatic feature extraction and multi-level feature representation, CNN has demonstrated extremely high accuracy and efficiency in tasks such as image classification, object detection, and image segmentation. In modern applications, CNN is not only widely used in daily image classification tasks, but also in high-demand scenarios such as medical image analysis, autonomous driving, and security monitoring, which has greatly promoted the development of related technologies.

Nevertheless, CNN still has the potential for further development. With the advancement of hardware and algorithms, the lightweight model, improved computing efficiency, and innovation of network structure will continue to promote the application of CNN on resource-constrained devices. In addition, combined with emerging technologies such as self-supervised learning and multi-task learning, CNN will improve learning ability and generalization performance without a large amount of labeled data. In the future, the application field of CNN will be more extensive, especially in three-dimensional data processing and model interpretability. Further research and innovation will provide strong technical support for more practical application scenarios. Through continuous optimization and innovation, CNN will continue to lead the development of image processing technology and make greater contributions to the field of artificial intelligence.

References

[1] Srinija, S.: Image Classification Using Convolutional Neural Networks. International Journal for Research in Applied Science and Engineering Technology, 2022.

[2] Tummala, M.: Image Classification Using Convolutional Neural Networks. International Journal of Scientific and Research Publications (IJSRP), 2019.

[3] Chimakurthi, V. N. S. S.: Application of Convolution Neural Network for Digital Image Processing. Engineering International, 2020.

[4] Rawat, W., Wang, Z.: Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. Neural Computation, 2017.

[5] Shen, X., Chen, Y., Tao, X., Jia, J.: Convolutional Neural Pyramid for Image Processing. ArXiv, 2017.

[6] Tian, Y.: Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm. IEEE Access, 2020.

[7] Bao, C., Li, Q., Shen, Z., Tai, C., & Xiang, X.: Approximation Analysis of Convolutional Neural Networks. East Asian Journal on Applied Mathematics. (2023).

[8] Zhang, Q., Chang, J., Meng, G., Xu, S., & Xiang, S.: Learning Graph Structure via Graph Convolutional Networks. Pattern Recognit., 95, 308-318 (2019).

[9] Yamashita, R., Nishio, M., Do, R., & Togashi, K.: Convolutional Neural Networks: An Overview and Application in Radiology. Insights into Imaging, 9(6), 611-629 (2018).

[10] Ayyadevara, V.K.: Convolutional Neural Network. In: Pro Deep Learning with TensorFlow, pp. 179-215 (2018).

[11] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J.: A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Transactions on Neural Networks and Learning Systems, 33, 6999-7019 (2020).

[12] Gabrielsson, R. B., & Carlsson, G.: A Look at the Topology of Convolutional Neural Networks. ArXiv, abs/1810.03234 (2018).

[13] Zhou, H., & Sun, Q.: Research on Principle and Application of Convolutional Neural Networks. IOP Conference Series: Earth and Environmental Science, 440 (2020).

[14] Mao, T., Shi, Z., & Zhou, D.: Theory of Deep Convolutional Neural Networks III: Approximating Radial Functions.

Neural Networks, 144, 778-790 (2021).

[15] Lei, F., Liu, X., Dai, Q., & Ling, B.: Shallow Convolutional Neural Network for Image Classification. SN Applied Sciences, 2 (2019).

[16] Tang, H.: Image Classification based on CNN: Models and Modules. 2022 International Conference on Big Data, Information and Computer Network (BDICN), pp. 693-696 (2022).

[17] Lin, Z., Li, B., Yong, L., Xia, Z., Wang, Y., Wu, J.: FPGA Acceleration of CNNs-Based Malware Traffic Classification. Electronics (2020).

[18] Weerasena, H., Mishra, P.: Revealing CNN Architectures via Side-Channel Analysis in Dataflow-based Inference Accelerators. ArXiv, abs/2311.00579 (2023).

[19] Ding, X., Xing, L., Lin, T., Wang, J., Li, Y., Miao, Z.: Evaluating CNNs for Military Target Recognition. Lecture Notes in Computer Science, 628-638 (2019).

[20] Khushi, H. M. T., Masood, T., Jaffar, A., Akram, S., Bhatti, S.: Performance analysis of state-of-the-art CNN architectures for brain tumour detection. International Journal of Imaging Systems and Technology (2023).

[21] Long, M., Yang, J., Xia, S., Wei, X.: Classification of rotor blade number of rotor targets micro-motion signal based on CNN. 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), pp. 182-186 (2021).

[22] Yan, Y.: AlexViT: Novel Diabetic Retinopathy Image Classification. 2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI), pp. 913-918 (2023).