

The eye of artificial intelligence - Convolutional Neural Networks

Jiamin Jiang

School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou
Jiangsu, 215028, China

Jiamin.Jiang21@student.xjtlu.edu.cn

Abstract. Inspired by the biological visual system, the convolutional neural network has been widely studied and invented in the field of artificial intelligence. As one of the important algorithms in artificial neural networks, convolutional neural networks have shown outstanding application potential in fields such as image recognition, computer vision, and natural language processing. This article will focus on exploring the powerful capabilities of convolutional neural networks in image processing. By delving into the implementation process of a convolutional neural network, readers will gain a deeper understanding of its working principles. In addition, this article will briefly introduce three classic models of convolutional neural networks, providing readers with more background knowledge. Next, this paper will analyze in detail two typical application cases of convolutional neural networks in the field of image processing: intelligent transportation systems and dental imaging technology. These cases demonstrate the successful application of convolutional neural networks in practical scenarios, pointing the way for their future development. In the future, convolutional neural networks will be more widely used in fields such as image and video processing as data scale increases and computing power improves. By using techniques such as model compression and hardware optimization, it is made more suitable for low-power and high-efficiency environments, and its interpretability and applicability are enhanced through data augmentation and model interpretation techniques.

Keywords: Convolutional Neural Network, Artificial Neural Network, Artificial Intelligence, Image Recognition.

1. Introduction

With the continuous evolution of the field of artificial intelligence, Artificial Neural Network (ANN) has laid the foundation for the rise of deep learning. Meanwhile, during the development of artificial neural network, Convolutional Neural Network (CNN) have rapidly developed and become an important algorithm model. In the mid-19th century, scientists gradually began to attempt to use software or hardware to simulate the working principles of the human brain, in order to create systems that can autonomously perform intelligent tasks, namely artificial intelligence. And artificial neural network is one of the most important models in artificial intelligence. Inspired by neurons in neurobiology, artificial neural network simulates the learning and feedback processes of the human brain. The artificial neural network had limited progress until the emergence of deep learning, a multi-level approach that extracts features from data through nonlinear transformations. The paper discusses the convolutional neural

network, a type of deep learning that extracts grid-structured data features, such as images or videos, through convolutional computation. The convolutional neural network is capable of effectively capturing and extracting local features from data through convolutional computation and pooling operations. They then extract the final abstract features from the raw data by overlaying and splitting the convolutional and pooling layers. It is important to note that the paper primarily focuses on the analysis of images using convolutional neural networks, rather than other types of grid-structured data. Therefore, this paper gives an introduction to convolutional neural network and roughly describes its operation. Its characteristics are discussed around two applications of convolutional neural networks. In addition, the current problems of convolutional neural networks and possible future directions are discussed.

2. Overview of the Convolutional Neural Network

Convolutional neural network is a feedforward neural network used for image processing and is one of the representative algorithms of deep learning. Convolutional neural network is composed of variants of the Multi-Layer Perceptron (MLP), designed under the inspiration of biologists Hubel and Wiesel's early research on the visual cortex of cats [1]. At the origin of neural network, Hubel and Wiesel found through experiments that the visual cortex of cats also exhibits a hierarchical structure in information processing, and the difficulty of extracting information gradually increases with the increase of levels [2]. The model they obtained is shown in the following figure 1:

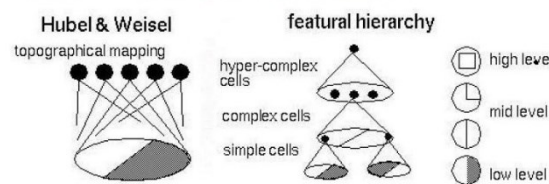


Figure 1. The hierarchical cortical model of cat's visual cortex

Based on the influence of Hubel and Wiesel experiments, Yann and LeCun et al. proposed convolutional neural network and designed and trained the first truly meaningful convolutional neural network model (LeNet model) using backpropagation algorithm and random gradient descent method [3]. This lays the foundation for convolutional neural network in the field of image recognition. There is a very similar concept between biological neural network and convolutional neural network, which is the receptive field. The receptive field refers to the area on the “neuron” that is only related to the input data in each convolutional layer [4]. It is one of the most important concepts in convolutional neural network, which emulates the response range of biological neurons to visual inputs. This concept will be mentioned in the subsequent detailed introduction of the convolutional neural network operation process. A convolutional neural network mainly consists of the following five layers: data input layer, convolutional computation layer, activation function layer, pooling layer and fully connected layer [5]. Afterwards, this paper will provide a specific introduction to these five layers. In the data input layer, the original image data will be preprocessed, such as normalization and PCA dimensionality reduction. This step facilitates subsequent operations on image data [5]. The convolutional computation layer is the most important layer for feature extraction in convolutional neural network. This layer has two key operations: local cross-correlation and receptive field. The size of the receptive field depends on the size of the convolutional kernel (filter) and the structure of the network [6]. In addition, different layers will have different sizes of receptive fields. Its size determines the local information range of input data that neurons can observe, and directly affects the performance and learning ability of the network. After selecting the filter size and stride, and fill values, perform convolution calculations to extract different input features and reduce the impact of noise in the convolutional computation layer. The activation function layer uses Rectified Linear Unit and other activation functions to nonlinearly map the output results of the convolutional layer in most cases [7]. The pooling layer performs feature dimensionality

reduction and extracts features to reduce the possibility of overfitting while ensuring that the features remain unchanged. This paper mainly explores the processing of images by convolutional neural network, and the main role of pooling layers can be simply understood as compressing images during image processing. The fully connected layer transforms the two-dimensional feature map output by convolution into a one-dimensional vector by adjusting parameters and changing weights. This layer implements an end-to-end learning process, which involves inputting an image to ultimately obtain a vector or information.

3. Models of the Convolutional Neural Network

There are many models of convolutional neural network that can be used to process images, such as LeNet, AlexNet, GoogLeNet, etc. This chapter provides a brief overview of several typical models and their respective characteristics.

As mentioned earlier, LeNet is one of the earliest convolutional neural network models that can be used to recognize handwritten digits [3]. It basically has a relatively complete structure, which is the main five-layer structure mentioned earlier. However, the structure of LeNet is relatively simple compared to other models, as shown in Figure 2. In addition, due to the relatively small size of the convolutional kernel and the use of a sigmoid function with an output range of only 0 to 1 as the activation function, the achievable functions are relatively limited.

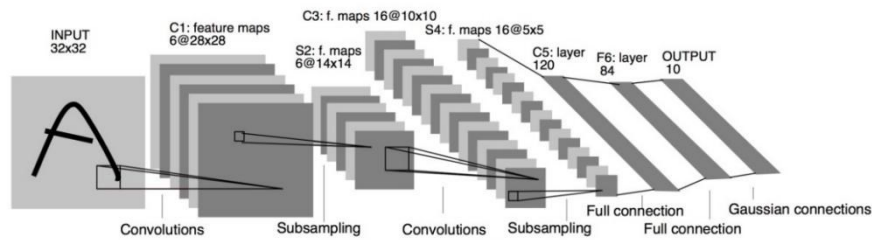


Figure 2. Flowchart of LeNet

AlexNet is a new model proposed by Alex Krizhevsky et al. in 2012 [8]. It was the first model to use a deep neural network to solve image problems. Compared to LeNet, AlexNet has a deeper network structure, such as 5 convolutional layers and 3 fully connected layers. Therefore, AlexNet has better feature expression and learning ability, and a deep network structure can improve generalization ability and performance. Figure 3 shows the network structure of the AlexNet model, which can visually display the difference in the number of layers compared to the LeNet model. Unlike LeNet, AlexNet uses a more efficient ReLU activation function. In addition, AlexNet introduces a local response normalization layer and utilizes Dropout technique to reduce overfitting. For features of different sizes, AlexNet uses convolutional kernels of different sizes in the last two convolutional layers to learn and extract features.

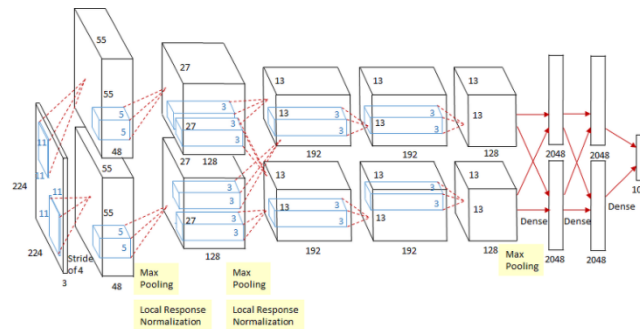


Figure 3. The network structure of an AlexNet model

GoogLeNet was proposed by the Google Brain team in 2014 [9]. Compared to the first two, GoogLeNet is unique in that it not only has depth, but also width in the horizontal direction. It uses the Inception module to extract features of different scales by parallelizing multiple convolution kernels of different sizes and pooling operations. The use of Inception in this model can be seen in Figure 4 below [10]. Unlike the previous two, GoogLeNet uses Dense Connectivity. By using small convolution kernels and adopting global average pooling to reduce the number of parameters in the network, computational and memory consumption can be reduced, which can improve the efficiency of the model. GoogLeNet improves model stability by introducing batch normalization technology to accelerate the convergence process of the model and reduce dependence on initial weights.

This paragraph mainly introduces the characteristics and differences of the three models. Convolutional neural network has other models that may differ in network structure, connection methods, and activation function selection. According to different requirements, different network models can be suitable for different application ranges.

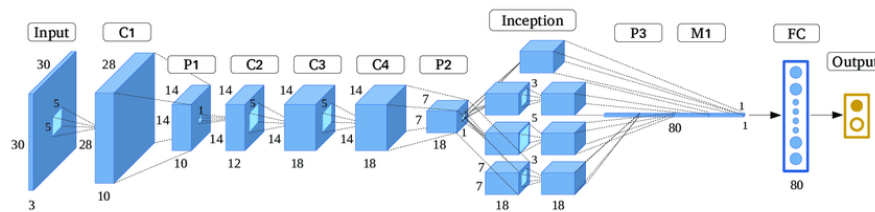


Figure 4. The network structure of an GoogLeNet model

4. Applications of the Convolutional Neural Network

In this chapter, two applications of image recognition using Convolutional Neural Networks are discussed, one in Intelligent Transportation Systems and the other in the field of dental imaging.

In the field of intelligent transportation systems, the research results of convolutional neural network can be applied in many scenarios. For example, license plate recognition technology can enable traffic management personnel to monitor road traffic more efficiently and achieve real-time monitoring and tracking of illegal vehicles. Traffic sign recognition technology can identify various traffic signs on the road in real time, providing accurate and timely traffic information for drivers to improve their safety while driving. License plate recognition technology is achieved through license plate positioning, character segmentation, and letter recognition. During the image acquisition stage, the surrounding environment may have an impact on the recognition results. To address this issue, Xiang et al. [11] proposed a lightweight fully convolutional network for license plate detection. This network reduces the proportion of input information, greatly accelerating computation speed and reducing computational costs. At the same time, it combines multi-level visual features and adopts dense linking and inflation methods to more fully utilize input information and capture the correlation between features. Besides, Asif et al. [12] proposed a new method for light invariance. In order to detect license plates, this technology utilizes heuristic energy maps to preprocess the extracted vertical edges, preserving high edge density areas. Finally, the AlexNet mentioned earlier is used to extract features and classify them to detect the information needed by the transportation department. With this method, the recognition accuracy of convolutional neural network has been greatly improved, with a numerical increase to 91.5%, and the average recognition time is only 0.16 seconds. Traffic sign recognition belongs to the external perception information of intelligent transportation systems and is implemented through deep learning. The basic process of implementation includes image preprocessing, feature extraction, and classification recognition. Compared to license plates, traffic signs contain more information and have fewer inter class differences. For greater convenience in analysis and understanding, Wang et al. [13] proposed a solution. Firstly, extract the maximum stable extremum region from the grayscale image of the HSV color space to obtain candidate traffic sign regions. Next, convolutional neural network is introduced to learn the shape, color, pattern content and other attributes of traffic signs, and to constrain their

recognition as traffic sign areas. Finally, if it is a traffic sign area, it is classified to achieve traffic sign detection. Convolutional neural network greatly improves the efficiency of urban transportation.

Convolutional neural network also has a wide range of applications in dental imaging, specifically in traditional disease detection and diagnosis. Diagnosis is completed through image preprocessing, boundary segmentation, feature extraction, and diagnostic classification. Lee et al. [14] evaluated the accuracy of deep convolutional neural network algorithm in diagnosing dental caries on apical slices. Through preprocessing and transfer learning, the diagnostic accuracy of premolars, molars, and two combined tooth models was 89.0%, 88.0%, and 82.0%, respectively. Lee et al. [15] achieved results in the diagnosis of periodontal disease on periapical X-rays by combining deep learning algorithms with pre trained deep convolutional neural network structures and self-trained networks. Their research results show that the diagnostic accuracy of premolars is 81.0%, and the diagnostic accuracy of molars is 76.7%. In addition, they also used 64 premolars and 64 molars clinically diagnosed with severe periodontal disease for testing, and the accuracy of predicting tooth extraction was 82.8% and 73.4%, respectively [15]. Orhan et al. [16] used an artificial intelligence system with deep convolutional neural networks to diagnose periapical periodontitis in cone beam CT (CBCT), with an accuracy rate of up to 92.8%. From this, it can be seen that convolutional neural networks have very high accuracy in processing dental medical images, and sometimes may even exceed human accuracy.

5. Discussion

At present, convolutional neural network still has many problems. The biggest problem people encounter when developing artificial intelligence is the difficulty in achieving low power consumption and high efficiency, which is currently the main goal pursued by humans. The number of parameters in convolutional neural network is usually large, which can lead to excessive computational overhead. This limits real-time performance in practical applications, and most people do not have the ability to afford to use it. Secondly, convolutional neural network is prone to overfitting when dealing with imbalanced datasets, which can lead to unstable prediction results, resource waste, and limited generalization ability. Besides, many of the decision-making processes of convolutional neural network cannot be explained in natural language as a black box model. This may limit its application scenarios. In the future, with the increase in data scale and the improvement of computing power, the application of convolutional neural network in fields such as image and video processing will become increasingly widespread. With low consumption and high efficiency as the primary goals, convolutional neural network will also continue to be updated and iterated. By adopting model compression technology and lightweight model structure design, convolutional neural network can reduce the complexity and number of parameters of the model, thereby achieving lightweight. This makes it more suitable for resource constrained environments, such as mobile devices and embedded systems. Meanwhile, parallel optimization and hardware optimization can improve the efficiency of the network, making it suitable for applications with high real-time requirements, such as video analysis and real-time object detection. To solve the problem of overfitting, algorithms can reduce data imbalance through techniques such as data augmentation, data balancing, and regularization. Enhance the interpretability of algorithms and expand their application scenarios through visualization, interpretive models, and transfer learning techniques [17].

6. Conclusion

In conclusion, the convolutional neural network is a relatively new artificial neural network algorithm. It is inspired by biological vision systems and has become one of the most important computational algorithms. The network is widely used in image recognition, computer vision, and natural language processing. This paper focuses on the image processing capabilities of convolutional neural networks. This paper provides an in-depth understanding of the implementation process of convolutional neural networks. Additionally, it briefly introduces three models of convolutional neural networks: LeNet, AlexNet, and GoogLeNet. Furthermore, it analyzes two applications of convolutional neural networks in image processing, namely, intelligent transportation system and dental imaging applications. The

paper concludes by discussing the challenges and future directions of convolutional neural networks. However, it suffers from several shortcomings. As previously mentioned, convolutional neural networks originated in the 1980s and 1990s, and there are relatively few references available to fully guarantee the accuracy of all available data. One limitation of this paper is that it only presents a convolutional neural network model without comparing it to other artificial neural network models. This makes it difficult to visually evaluate the strengths and weaknesses of convolutional neural networks.

References

- [1] Y. LeCun, Y. Bengio and G. Hinton, Deep learning, in *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.
- [2] B. Li, Y. Todo and Z. Tang, Artificial Visual System for Orientation Detection Based on Hubel–Wiesel Model, in *Brain Sci.*, vol. 12, no. 4, p470, Apr. 1, 2022. Doi: 10.3390/brainsci12040470.
- [3] C. Nandini and S. S. Reddy, Detection of communicable and non-communicable disease using Lenet-Bi-Lstm model in pathology images, in *International Journal of System Assurance Engineering and Management*, vol. 15, no. 1, 2024. Doi: 10.1007/s13198-022-01702-5
- [4] L. Fu, M. Cai, X. Hou, Z. Xu, and M. Tao, Receptive Field: A Survey, in *IEEE Access*, vol. 9, pp. 3553-3565, 2021, doi: 10.1109/ACCESS.2020.3040702.
- [5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998. Doi: 10.1109/5.726791.
- [6] S. S. Nandha, V. P. Athish and D. Rajeswari, Train Track Crack Prediction Using CNN with LeNet - 5 Architecture, in *INCOFT*, Nov. 1-5, 2022. Doi: 10.1109/INCOFT55651.2022.10094528
- [7] V. Nair and G. E. Hinton, rectified linear units improve restricted boltzmann machines, in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, Haifa, Israel, 2010, pp. 807–814.
- [8] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.
- [9] C. Szegedy, W. Jia and et al., Going Deeper with Convolutions, in *IEEE Conference on Computer Vision and Pattern Recognition*, New York, pp.1-9, 2015
- [10] Z. Huang, L. Ni, N. Chen and et al., Improved GoogleNet Model for Tired Driving Detection, in *ICCWAMTIP*, Dec. 1-7, 2023. Doi: 10.1109/ICCWAMTIP60502.2023.10387133
- [11] H. Xiang, Y. Zhao, Y. Yuan and et al., Lightweight fully Convolutional Network for License Plate Detection, in *Optik*, vol:178, pp.1185-1194, 2019.
- [12] M. R. Asif, C. Qi, T. Wang and et al. License Plate Detection for Multi-national Vehicles: An Illumination Invariance Approach in Multi- Lane Environment, in *Computers and Electrical Engineering*, vol. 78, pp. 132-147, 2019.
- [13] F. Wang, J. Wang, B. Li and et al., Deep Attribute Learning Based Traffic Sign Detection, in *Journal of Jilin University (Engineering and Technology Edition)*, vol:48, no.1, pp.319-329, 2018. (in Chinese)
- [14] JH. Lee, DH. Kim, SN. Jeong and et al. Detection and Diagnosis of Dental Caries Using a Deep Learning-based Convolutional Neural Network Algorithm, in *J Dent*, vol:77, pp. 106- 111, 2018.
- [15] JH. Lee, DH. Kim, SN. Jeong and et al. Diagnosis and Prediction of Periodontally Compromised Teeth Using a Deep Learning-based Convolutional Neural Network Algorithm, in *J Periodontal Implant*, vol:48, no.2, pp.114- 123, 2018
- [16] K. Orhan, IS. Bayrakdar, M. Ezhov and et al. Evaluation of Artificial Intelligence for Detecting Periapical Pathosis on Cone Beam Computed Tomography Scans, in *Int Endod J*, vol:53, no.5, pp. 680- 689, 2020.

- [17] M. Sara, A. Hala, A. A. Elsayed and S. Hassan, Iterative magnitude pruning-based light-version of AlexNet for skin cancer classification, in *Neural Computing and Applications*, vol. 36, no. 3, pp.1413-1428, 2024. Doi: 10.1007/s00521-023-09111-w