Image recognition algorithms based on deep learning

Hongli Ma\textsuperscript{1,a}, Fang Xie\textsuperscript{2,b*}, Tao Chen\textsuperscript{3,c}, Lei Liang\textsuperscript{4,d} and Jie Lu\textsuperscript{5,e}

\textsuperscript{1}College of Mechanical Engineering, Xijing University, Xi’an, Shaanxi, China
\textsuperscript{2}College of International Exchange and Education, Hebei University, Baoding, Hebei, China
\textsuperscript{3}email: xj@xijing.com.cn, \textsuperscript{4}email: 1632420948@qq.com, \textsuperscript{5}email: 1140665174@qq.com
\textsuperscript{b}email: 983217250@qq.com
\textsuperscript{c}email: 1468786256@qq.com

Abstract. Convolutional neural network is a very important research direction in deep learning technology. According to the current development of convolutional network, in this paper, convolutional neural networks are induced. Firstly, this paper induces the development process of convolutional neural network; then it introduces the structure of convolutional neural network and some typical convolutional neural networks. Finally, several examples of the application of deep learning is introduced.

1. Introduction

In recent years, with the gradual rise of machine learning methods represented by deep neural networks, due to the increase of training data and the significant improvement of computing power, the network structure of deep neural network and its suitable optimization algorithms become more and more complex, but its performance is also better and better, accompanied by the classical network structure of many different types of data processing tasks, for example, convolutional neural network, recurrent neural network, and residual neural network. Among them, convolutional neural network plays a leading role in the deep learning task, and it has a good application prospect in the field of image recognition.

2. Development of convolutional neural networks

In 1962, Hubel and Wiesel showed that the visual information transmitted from the retina to the brain was stimulated by multi-level receptive fields, and the concept of receptive field was first put forward\textsuperscript{[1]}. Receptive field is the area size mapped by the pixels on the feature map output by each layer of convolutional neural network on the input picture. In 1988, LeNet proposed by Lecun used gradient based back propagation algorithm to supervise the training of the network\textsuperscript{[2]}. LeNet5 network gradually converts the original image into a series of feature maps through alternately connected convolution layer and lower sampling layer, and transmits these features to the fully connected neural network to recognize and classify the image according to the characteristics of the image. In 2012, the AlexNet network of Hinton was proposed to use the relu function as the activation function of convolutional neural network. Its effect exceeded sigmoid in deeper networks, optimized the gradient dispersion problem of sigmoid in deeper networks, and established the position of convolutional neural network in deep learning. With the increase of data volume, the problems encountered become...
more and more complex, followed by more excellent models such as AlexNet[3], VGG[4], GoogleNet[5] and MobileNet[6].

3. Convolutional neural network structure
Convolution neural network is mainly divided into convolution layer, pool layer and full connection layer. Convolution layer mainly undertakes a lot of calculation and feature extraction, pool layer does not undertake calculation and feature extraction again, and full connection layer can be understood as a hidden layer in neural network.

3.1. Convolution
The convolution layer is mainly responsible for feature extraction. It retains the spatial features of two-dimensional images. For example, for a 32 * 32 * 3 image, the input of the convolution layer is a 32 * 32 * 3 matrix without any change. To extract these features, it is very important to set the size, number and step of convolution kernel. Convolution check the input image for processing. The calculation method is inner product, that is, the convolution kernel is multiplied by the position of the convoluted area of the input image and then summed. The process is shown in Figure 1. 3 * 3 convolution is used to check the matrix of 4 * 4 input image for convolution, and the step size is set to 1.

![Figure 1. Convolution process diagram](image1)

3.2. Pool layer
Pool layer is a method of compressing and sampling images. According to the locality principle of the original input image, it can be divided into maximum pooling, average pooling and so on. The pooling operation can also be referred to as down sampling. After obtaining the convolution features of the input image, the feature image output by the convolution layer is pooled with a pooled area with the size of M * N. As shown in figure 2.

![Figure 2. Maximum pooling and average pooling](image2)

3.3. Fully Connected Layer
Fully connected layer can be simply understood as hidden layer, which has weight and nonlinear function. Like the traditional neural network, neurons are fully connected. In the general convolution model, there is at least one layer of fully connected network. Convolution layer and pool layer are stacked to form a simple convolution neural network. In order to better complete the task, the
two-dimensional feature information output by convolution layer and pool layer is transformed into one-dimensional feature information. Usually, a full connection layer is added behind convolution layer and pool layer.

4. Classical convolutional neural network structure

This section will introduce some classical convolution network structures, such as LeNet\(^2\), AlexNet\(^3\), VGG\(^4\), and GoogleNet\(^5\).

4.1. LeNet

LeNet\(^2\) was created by Yann Lecun in 1998. It is the first of convolutional neural network method, which is used for handwritten numeral recognition. LeNet also known as leNet-5\(^7\), has one input layer, two convolution layers, two pool layers, two full connection layers and one output layer (the last full connection layer is the output layer). The first input layer, C1 and S2 are the first convolution layer and the first pool layer respectively; C3 and S4 are the second convolution layer and the second pool layer respectively; C5 is the third convolution layer; F6 is the full link layer; The last layer is the output layer.

4.2. AlexNet

AlexNet\(^3\) was proposed by Geoffrey and Alex to defend the champion in the third ISLVRC competition. Features of AlexNet network structure: (1) AlexNet adopts dual GPU operation; (2) The activation function selected by AlexNet is the relu function, which is much faster than the traditional sigmoid function and tanh function. (3) The pooling method uses the pooling window larger than the step size (overlapping pooling) to prevent over fitting; (4) Enhance learning ability through parameter optimization strategy. AlexNet has an 8-layer structure, with layers 1 to 5 as convolution layers and the last three as full connection layers.

4.3. VGG

VGG\(^4\) model is based on the relationship between network depth and performance. It was proposed by Oxford University in 2014, that is a deeper model with good performance. The classical VGG models mainly include VGG-16 and VGG-19. This paper mainly introduces VGG-16. VGG-16 contains 16 layers of network, including 13 convolution layers and 3 full connection layers, excluding pool layer and nonlinear function layer. Compared with the AlexNet model, its convolution kernel size is fixed as 3 * 3 and the step size is set as 2. However, due to the large number of structural layers, the training will be slow.

4.4. GoogleNet

GoogleNet\(^5\) was proposed by Google team in the fifth ISLVRC competition, and its efficiency is better than VGG. The traditional network optimization method is to increase the depth and width, but the Google team proposed the initial module structure to make its network structure better, that is, to construct basic neurons to establish a high-performance computing network. In the inception module, the step size is fixed to 1, the convolution kernel is 1 * 1, 3 * 3 and 5 * 5, the padding processing values are 0, 1 and 2, and the 1 * 1 convolution kernel is used to reduce the amount of parameters during operation. The three filters 1 * 1, 3 * 3 and 5 * 5 are connected in series to form a "component block", which can not only sparse the neural network, but also make full use of the great advantages of computer for dense matrix calculation. The concept network structure is shown in Figure 3:
5. Application examples of deep learning

5.1. Application of deep learning in medical imaging
In recent years, with the development of computer technology, medical imaging technology based on deep learning has been further developed. Jin Wang[8] and others extracted 74 texture features from X-ray chest films from multiple angles as the input of algorithm model, and introduced convolution sparse coding algorithm to reconstruct JSRT data set, so that the sensitivity and specificity indexes can reach 78.8 and 76.9. Guoxiang Ma[9] and others induced the focus detection methods by using the image enhancement method of CT images, and compared and analyzed the standard Gan, pix2pixgan and Cyclegan models in the medical field.

5.2. Application of deep learning in automatic driving
In recent years, with the rapid development of automobile industry, automatic driving has become a hot spot. By modifying the resolution of the input image, Zhihong Yuan[10] and others accelerated the detection speed, and modified part of Yolov3 network to form Yolov3-bt network, so that the pedestrian detection rate reached 87% and the obstacle detection rate reached 92%. Kangming Zhang[11] and others proposed an improved detection model for SSD algorithm. Through the actual measurement of KITTI data set, the accuracy is improved by 13% and the detection time is shortened by 15%.

5.3. Application of deep learning in face recognition
In recent years, face recognition has been applied to many aspects of life and has become some channels for people to travel and pay. By combining convolution neural network and generative countermeasure network, Yan Li[12] reduces the gap between the generated image and the real face image, and solves the problem of insufficient samples. Yi Jiang[13] and others proposed a method to generate mask wearing face image by combining generative countermeasure network and spatial transformation network. Experiments show that this method successfully wears mask on face image, and the image has high authenticity.

6. Conclusion
This paper first introduces the network structure of CNN, and then introduces some classical network models and their working process. It is found from the above that in the field of image recognition in the future, how to achieve more accurate image recognition is the top priority in the future.

Acknowledgments
The Shaanxi Science and Technology Plan project (Grant No. 2021KRM025)

Reference
[1] Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. The Journal of physiology, 1962, 160(1): 106-154.
[2] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[3] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25: 1097-1105.

[4] Ding X, Zhang X, Ma N, et al. Repvgg: Making vgg-style convnets great again[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13733-13742.

[5] Aldegheishem A, Anwar M, Javaid N, et al. Towards sustainable energy efficiency with intelligent electricity theft detection in smart grids emphasising enhanced neural networks[J]. IEEE Access, 2021, 9: 25036-25061.

[6] Srinivasu P N, SivaSai J G, Ijaz M F, et al. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM[J]. Sensors, 2021, 21(8): 2852.

[7] Rouhsedaghat M, Wang Y, Ge X, et al. Facehop: A light-weight low-resolution face gender classification method[C]//International Conference on Pattern Recognition. Springer, Cham, 2021: 169-183.

[8] Wang Jin, Feng Youhong, Research on medical image detection method based on deep learning technology [J]. Electronic design engineering, 2021,29 (20): 169-173.

[9] Ma Guoxiang, Yan Chuango, Zhang Zhihao, Sen Gan. CT image lesion detection method based on data enhancement [J]. Computer system application, 2021,30 (10): 187-194.

[10] Yuan Zhihong, sun Qiang, Li Guoxiang, Bai shuzhan, Yan Ying, Zhang Zhenhua. Automatic driving target detection based on yolov3 [J]. Journal of Chongqing University of Technology (NATURAL SCIENCE), 2020,34 (09): 56-61.

[11] Zhang Kangming, Cao Xin. Research and application of traffic image recognition based on deep learning [J]. Software engineering, 2021,24 (06): 32-36 + 27.

[12] Li Yan. Generation of face virtual expression samples based on deep learning [J]. Electronic production, 2021 (16): 65-67 + 97.

[13] Jiang Yi, LV Rongzhen, Liu Mingzhu, Han Chuang. Facial mask image synthesis based on generated countermeasure network [J / OL]. Journal of intelligent systems, 2021 (06): 1-8 [2021-10-26].