Face Recognition Based on Convolutional Neural Networks

Ruishen Liu*

Faculty of Science, Beijing University of Technology, Beijing, China

*Corresponding author: liuxijie@cscec.com

Abstract. Since science and technology have been progressing steadily in recent years, deep learning's potential applications have expanded greatly. From unlocking the screen of a phone with a human face to driverless technology, which has emerged in recent years. Facial recognition is proving to be a boon to life. Among various deep learning algorithms, the appearance of convolutional neural network (CNN)has made unprecedented progress in image recognition. In this paper, the basic principles of convolutional neural networks are explained, and the most important concepts are introduced. The convolutional neural network is used for experiments. The input layer, convolution layer, pooling layer, fully connected layer, and output layer are the nine layers that make up the traditional and complete convolutional neural network model, which is used as the experimental foundation. LFW dataset is used for training, and the experimental results are given. At the end of the paper, the accuracy and loss functions are analyzed and the accurate results of facial recognition are achieved.

Keywords: Face recognition; Convolutional neural network; Deep learning; LFW dataset.

1. Introduction

1.1. Background and significance of the study

In the 1950s, the first experiments on artificial neural networks were done.Because of the level of development of computer science and the size of the data, it did not have the soil for development for a long time. Since the 1980s, the level of processing information and liberating human labor brought about by deep learning frameworks has been constantly improving. Moreover, deep learning frameworks continue to be successfully and effectively applied to more and more complex practical problems.The development of deep learning is also heavily dependent on the development of hardware and software[1].

Over the past decades, deep learning frameworks have borrowed a great deal of knowledge about mathematics, statistics, and human brain neuroscience. With the rapid advancement of computer theory and hardware in recent years, making the current computing resources can run larger models, so that computer vision has obtained many important research results. Face recognition as an important direction of computer vision research has also obtained great progress.

For the study of image recognition and other fields, scholars have made a lot of research results, and put forward many effective image recognition algorithms. However, the development of facial recognition technology still had a lot of issues and technological roadblocks.For example, in the application field of face recognition, the model was not only prone to be interfered from the external environment, but also difficult to recognize complex facial expressions. It must give consideration to real-time recognition and accuracy. When it was applied to real life, the resultwas not ideal, such as incorrect identification or long response time. On the road of the progress of image recognition technology, it was urgent to study a technology that was not easy to be interfered with by the outside world but couldalso process tasks in real time[2].The academics hope to find a technology that matches real-life scenarios.

With the introduction of convolutional neural network in recent years, the application of deep learning in face recognition can express the original face information more essentially, and the accuracy rate of face recognition has been rapidly improved through the feature extraction and autonomous learning of network deep learning[3].

1.2. The research status

Convolutional neural networks have been studied since the 1980s. In his published paper, Japanese scholar Kunihiko Fukushima proposed a new neural network with deep structure for the first time, which was regarded as the inspiration of the pioneering research on convolutional neural networks. Subsequently, in the 1990s, Professor LeCun completed the development of a check number recognition system in his laboratory. Since then, convolutional neural network has been applied and gradually entered the field of vision of people.

Since 2000, in order to recognize faces, convolutional neural networks are frequently utilized. Because of the important research significance and bright application prospect of face recognition technology, more and more organizations have begun to study it, including Massachusetts University of Technology, Stanford University, Peking University, Harbin Institute of Technology and others.Today's face recognition field has made a lot of progress, making face recognition technology in both detection speed and recognition accuracy have been greatly improved.

2. The Convolutional Neural Network Model

A particular type of neural network known as a convolutional neural network excels at processing data that can be evolved into images. Input, convolution, pooling, fully connected, and output layers are typically present in convolutional neural networks.Convolutional neural networks have found widespread applications in all facets of life, from autonomous automobiles to the ability to unlock computers and mobile phones using a person's face.

2.1. Representation of an Image in aComputer

The image is made up of individual sample points called pixels. The image generated by capturing the real scene through the device is stored as a bitmap after being digitized inside the computer. An image is composed of M rows and N columns of sample points.

For black and white images, each pixel is only black and white, so only one binary bit is needed.For example, 0 for black and 1 for white. For gray images, each pixel needs to be able to display a total of 256 gradients from black to white, typically represented by eight binary bits. For color images, it is necessary to divide them into three channels, including red, green and blue channels, and the pixels in each channel are used to measure brightness. Therefore, it can be seen that the representation method of pixels in the image is as follows: the black and white image is represented by a matrix consisting of binary bits 0 or 1; each pixel of the gray image is represented by a matrix consisting of 0-255; each pixel of the color image is represented by 0-255 and consists of three matrices[4].

2.2. Convolution

In the general form,$f(x)$and $g(x)$can be regarded as two integrable functions over the domain, and the integral can be done as Formula 1:

$$s(\tau) = \int_{-\infty}^{+\infty} f(\tau)g(x-\tau)\,\mathrm{d}\tau \tag{1}$$

Convolution is the name of this operation. Convolution is usually denoted by $*$as shown in Formula 2:

$$s(\tau) = (f * g)(\tau) \tag{2}$$

The function f in the example above is the variable called the input in the convolutional neural network, and the function g in the example above is the kernel.

When the computer is processing data, the processed data will be discretized according to time, so the discretized convolution definition is also needed as Formula 3:

$$s(\tau) = \sum_{\tau=-\infty}^{\infty} f(\tau)g(x-\tau) \tag{3}$$

Most of the time in machine learning, multi-dimensional arrays are used as input data. However, the convolution kernel is different and needs to use self-learning algorithm to automatically optimize into each convolution layer through back propagation.

Convolution operations often need to be carried out in multiple dimensions. For example, a two-dimensional image I is taken as the input of the convolution layer, and a convolution kernel J is used to carry out convolution operations on this image, as described in Formula 4:

$$S(i,j) = (I * J)(i,j) = \sum_m \sum_n I(m,n)J(i-m, j-n) \tag{4}$$

Therefore, we can get the convolution formula for two dimensions[5].

2.3. Convolution Properties

Convolution operation has three important properties: parameter sharing, translation invariance and sparse connectivity.

Parameter sharing: It use the same parameter in different functions of the same model. Each weight matrix element in the fully connected network can only have one interaction with each neuron in a certain input layer while calculating the output of a particular layer.While in the convolutional neural network, each element in the convolution kernel matrix can act on every position of the input. This means that the weight of an input represents the weight of the same convolutional neural network layer. Therefore, in the process of back propagation, only one matrix needs to be changed instead of a separate weight matrix for each layer, which significantly reduces the storage capacity of the model.

Translation invariance: The physical property of parameter sharing combined with the pooling process can make the convolution operation have translation invariant property. For example, I represents the function of the image on integer coordinates, and g represents the transformation function of the image function such that $I' = g(I)$, where the image function I' satisfies $I'(x,y) = I(x-1,y)$. This function moves each pixel in I one unit to the right, that is, the result of the transformation of I and then convolution operation. It's the same thing as convolving I and then applying the translation function$g$ to the output.

Sparse connectivity: In traditional neural networks, the relation between the input and output of two adjacent network layers can be represented by a matrix composed of weight values. Each element value in the matrix represents the connection between two neurons in two adjacent network layers. Any neuron in a fully connected network is connected to every other neuron in the layers above it, creating an extremely dense network structure. However, in the convolutional neural network, the scale of the convolution kernel is much smaller than that of the input, so the neurons in each layer are only related to a few neurons in the previous layer. The reason for doing this is that data such as images and voice have a local structure, so there is no need to connect with all the neurons in the previous layer, and the computation can be greatly reduced[6].

2.4. Pooling

In the image, there is a large probability relationship between two adjacent elements, and the large probability reflects the same thing, which implies that a lot of data can be summed up into a small number of points.Max pooling and average pooling are the two most common types of pooling.

Taking max pooling as an example, an image can be divided into many regions. Max pooling means that in a region, the pixel with the maximum value is used to represent the whole region, and the maximum pixels selected from each block of the image can be spliced together to form a more simplified image compared with the original image.

There are two main purposes to do this: The first purpose is to reduce the features carried by the original image, which can greatly reduce the amount of computation and complexity in the calculation of the convolutional neural network. At the same time, some redundant information can be removed and only the key information can be retained, so that the efficiency of the convolutional neural

network can be greatly improved. The second purpose is to make the image translation invariant. When a pixel changes position around it, pooling can ignore this movement[7].

Pooling also has its limitations. When screening the pixels to remove redundant information, it can also remove some information which might be useful.At the same time,the image cannot recover through the pooling operation.

3. Experiment

3.1. Dataset

This experiment uses LFW as the training dataset of face recognition. The LFW dataset is not collected from a lab. The images were collected from the Internet. The LFW dataset contains more than 13,000 images of faces taken in natural conditions. As can be seen from the following picture, human faces in the LFW data set have various poses and expressions, as well as different lighting conditions. Some faces are even blocked, which greatly increases the difficulty of face recognition.Figure 1 presents the samples of LFW dataset.

Figure 1. The LFW dataset

To add diversity of the dataset, this study also captured more real facial pictures in daily life for experimental recognition of facial information. Face information is captured by laptop camera. Dlib library is used to extract facial features from the images captured by the camera. The experiment took 20,000 images.Figure 2 presents this sample.

Figure 2. Face dataset with captured new images

3.2. Model

This experiment adopts nine neural network layers as experimental model. It includes input layer, convolution layer, pooling layer, fully connected layer and output layer. The connection between each layer is shown in Figure 3.

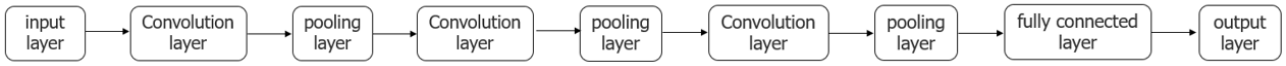

Figure 3. Experimental model of convolutional neural network

The input layer can detect the location information of the face, and then the coordinate and size information of the face is saved in the form of bitmap inside the computer, and it is sent to the neural network for training[8].

The size of the convolution kernel in the convolution layer is (3,3), and the step size is 1. Meanwhile, zero is added in the outer circle to prevent the elements in the outer circle of the matrix from being seldom used while the elements in the center of the matrix are overused. The convolutional layer performs feature extraction on the output results of the input layer.

The pooling layer adopts maximum sampling, and the sampling size is 2×2, which means that the feature maps can be cut into a completely non-overlapping 2×2 matrix, the maximum value of each cut matrix is taken, and then they are reassembled. The pooled matrix is much smaller than the original matrix, which can reduce the amount of computation in the model. At the same time, because the required features have been extracted, so the experimental results will not have a great impact.

Then the dropout method is used, which randomly drops a number of neurons at a certain probability. The purpose of this is not only to get a higher training speed, but also to prevent overfitting.

The fully connection layer can enhance the nonlinear capability of the network, generally using ReLU activation function as Formula 5[9]:

$$f(x) = max(0, x) \tag{5}$$

It can also limit the size of the network. After feature extraction of the network is completed, a fully connection layer is connected, and each neuron of the fully connection layer is connected with all neurons of the previous layer.The results of the previous layer are compressed into a one-dimensional vector[10].

The output layer can apply the results of the full connection layer to the Softmax function for normalization as Formula 6:

$$P(y|x) = \frac{e^{h(x,y_i)}}{\sum_{j=1}^{n} e^{h(x,y_j)}} \tag{6}$$

Normalization makes the probability of predicted results between 0 and 1, and then the maximum probability is selected for prediction. In this experiment, the results are divided into two categories to determine whether the face is correct.

In addition, in the course of this training, the learning rate was set as 0.1, the loss function was the cross entropy, and the optimizer was Adam. In this experiment, more than 30,000 images were taken as data sets, and the ratio of training set and test set was 20:1.

3.3. Analysis

The efficiency of model training has different performance at different learning rates. Figure 4 to 6 show the learning rate curves under different conditions.

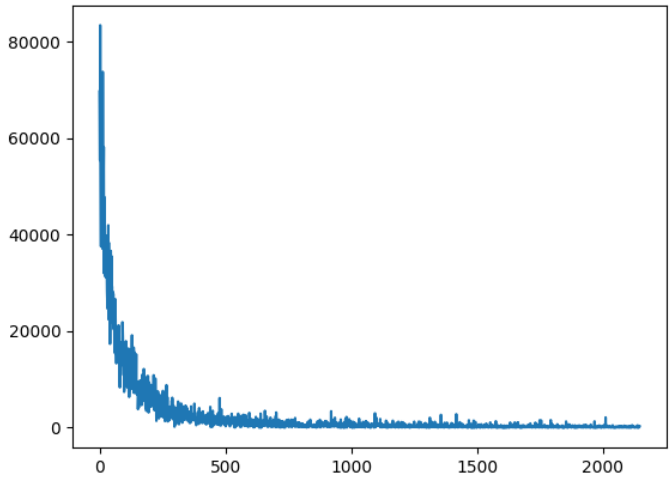

Figure 4. Loss function curve with learning rate of 0.01

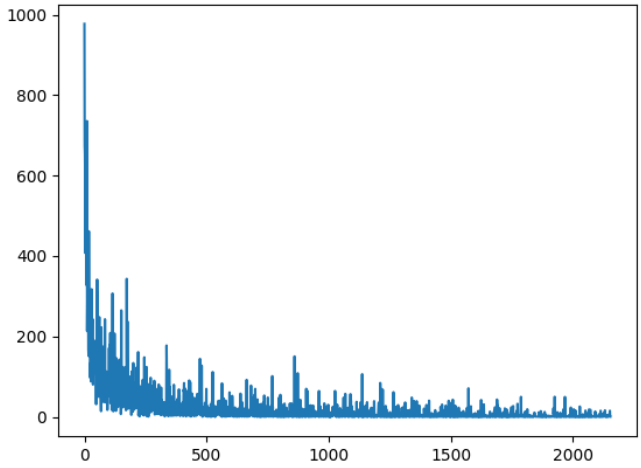

Figure 5. Loss function curve with learning rate of 0.05

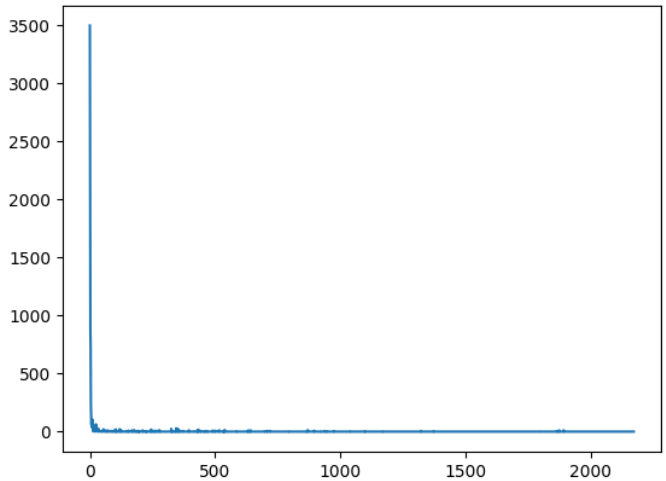

Figure 6. Loss function curve with learning rate of 0.1

As can be seen from the above three figures, the higher the learning rate is, the faster the convergence of training results can be. When the learning rate is 0.1, even few steps are required to converge to a satisfactory result.

In order to see the changing trend of the accuracy curve, the experiment outputs the accuracy rate once in one hundred steps, and adopts the learning rate of 0.01, because the small learning rate can make the accuracy curve change gradually. The accuracy curve is shown in Figure7.

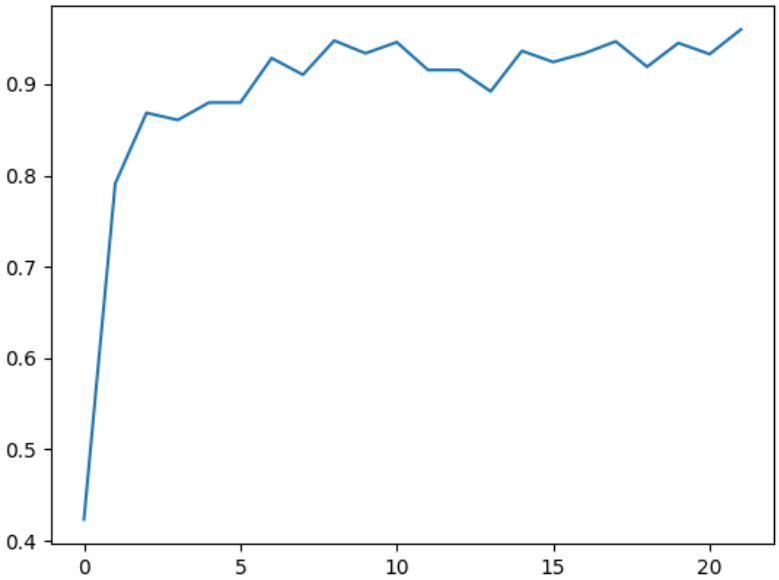

Figure 7. Accuracy curve with learning rate of 0.01

As can be seen from the line graph above, after 300 steps of learning, the accuracy rate can reach more than 80%. Although the subsequent accuracy will fluctuate, the accuracy rate can be stable at more than 90% with gradual and continuous learning.

3.4. Results

The output layer contains two types of output, with yes and no to distinguish the correct face. The results are presented in Figure 8.

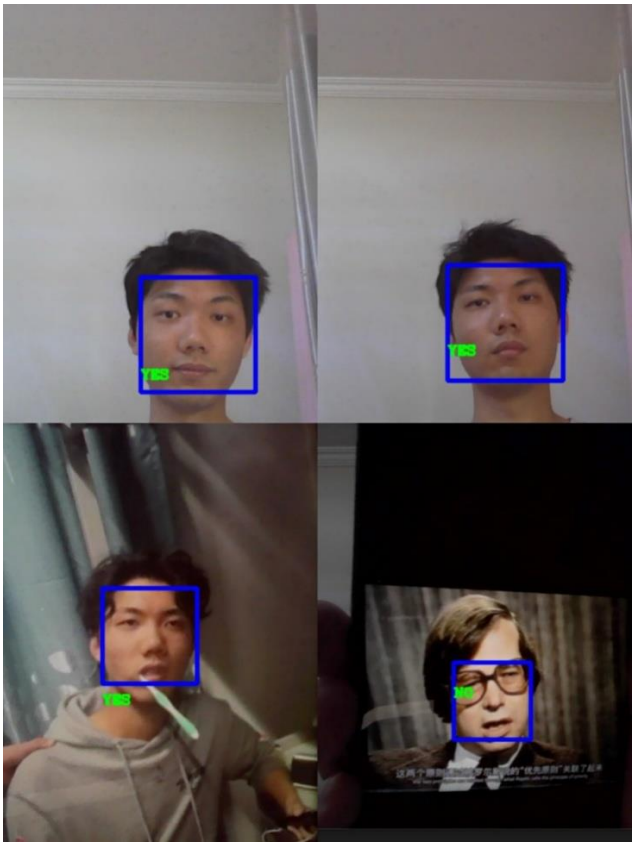

Figure 8. Facial recognition results

The model can successfully recognize facial information, including photographs from mobile phones, as shown in the figures above, demonstrating its viability.

4. Conclusions

Nowadays, living in a dazzling and colorful world, more and more people can feel the convenience of life brought by scientific and technological progress. Computer vision as a major field of computer science, has made great progress. Therefore, it is very necessary to explore the neural network of deep learning.

Convolutional neural network is described in this paper. It can process image information and extract image features from images efficiently through operations such as convolution, pooling and full connection. On this basis, target detection and semantic segmentation can be completed.

On the basis of discussing the principle of convolutional neural network, this paper also performs the experiments. The model is constructed completely and accomplishes the face recognition, which can correctly distinguish the face information under different conditions. It can be seen that convolutional neural network has an excellent performance in face recognition processing. Beyond this paper, it still has many features that can be studied.

References

[1] Zhang Rong, Li Weiping, Mo Tong, A review of deep learning research [J], Information and Control, 2018,47(04):385-397.

[2] Yang li, Wu Yuxi, Wang Junli, Liu Yuli, A review of research on recurrent neural networks [J], Computer application, 2018,38(S2):1-6.

[3] Li Guanghao, Research and application of face recognition system based on convolutional neural network [D], Lanzhou University of Technology, 2021.

[4] Chen Yaodan, Wang Lianming, Face recognition method based on convolutional neural network [J], Journal of Northeast Normal University (natural science edition),2016,48(02):70-76.

[5] Xu Guoan, Face recognition based on convolutional neural network [D], Harbin University of Science and Technology, 2021.

[6] Xu Wenhua, Design and implementation of long text classification algorithm based on deep neural network [D], Nanjing University of Posts and Telecommunications, 2020.

[7] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[C].

[8] Xia Yulu, A review of the development of recurrent neural networks [J], Computer Knowledge and Technology, 2019,15(21):182-184.

[9] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, Yoshua Bengio, Generative Adversarial Nets [M], 2014,2672-2680.

[10] LeCun Y, Bengio Y, Hinton G, Deep Learning [J], Nature, 2015,521(7553):436.