PAPER • **OPEN ACCESS**

Research on Face Recognition Based on CNN

To cite this article: Jie Wang and Zihao Li 2018 *IOP Conf. Ser.: Earth Environ. Sci.* **170** 032110

IOP Conf. Series: Earth and Environmental Science **170** (2018) 032110 doi:10.1088/1755-1315/170/3/032110

Research on Face Recognition Based on CNN

Jie Wang and Zihao Li

School of Electrical Engineering, Zhengzhou University, Zhengzhou, China

Abstract. With the development of deep learning, face recognition technology based on CNN (Convolutional Neural Network) has become the main method adopted in the field of face recognition. In this paper, the basic principles of CNN are studied, and the convolutional and downsampled layers of CNN are constructed by using the convolution function and downsampling function in opencv to process the pictures. At the same time, the basic principle of MLP Grasp the full connection layer and classification layer, and use Python's theano library to achieve. The construction and training of CNN model based on face recognition are studied. To simplify the CNN model, the convolution and sampling layers are combined into a single layer. Based on the already trained network, greatly improve the image recognition rate.

1. Introduction

Intelligent systems appear more and more in people's lives, and often need to be identified when using intelligent systems. Traditional methods of identification mainly identify individuals with some personal characteristics, such as identity documents, such as documents and keys, which have obvious shortcomings. They are easily forgotten, lost or faked. If you use some of the personal characteristics to identify the effect will be quite good, such as: face recognition, fingerprinting and so on.

In terms of algorithms, there are sharing parameters between the convolution layer and the convolution layer of CNN. The advantage of this is that the memory requirements are reduced, and the number of parameters to be trained is correspondingly reduced. The performance of the algorithm is therefore improved. At the same time, in other machine learning algorithms, the pictures need us to perform preprocessing or feature extraction. However, we rarely need to do these operations when using CNN for image processing. This is something other machine learning algorithms cannot do. There are also some shortcomings in depth learning. One of them is that it requires a lot of samples to construct a depth model, which limits the application of this algorithm. Today, very good results have been achieved in the field of face recognition and license plate character recognition, so this topic will do some simple research on CNN-based face recognition technology.

2. Convolution neural network

2.1. Convolutional neural network introduction

With the development of convolutional neural networks, the achievements made in various competitions are getting better and better, making it the focus of research. In order to improve the training performance of the forward BP algorithm, an effective method is to reduce the number of learning parameters. This can be done by convolution of the spatial relationship of the neural network. Convolutional neural network, the network structure is proposed, it minimizes the input data pretreatment. In the structure of convolution neural network, the input data is input from the initial input layer, through each layer processing, and then into the other hierarchy, each layer has a convolution kernel to obtain the most significant data characteristics. The previously mentioned obvious features such as translation, rotation and the like can be obtained by this method.

2.2. Convolution neural network basic structure

Neural network can be divided into two kinds, biological neural network is one of them, and artificial neural network is another kind. Here mainly introduces artificial neural network. An artificial neural network is a data model that processes information and is similar in structure to the synaptic connections in the brain. Neural network is composed of many neurons; the output of the previous neuron can be used as the input of the latter neuron. The corresponding formula is as follows:

$$h_{W,b}(x) = f\left(W^T x\right) = f\left(\sum_{i-1}^{3} W_i x_i + b\right) \tag{1}$$

This unit is also called Logistic regression model. When many neurons are linked together, and when they were layered, the structure can now be called a neural network model. Figure 1 shows a neural network with hidden layers.

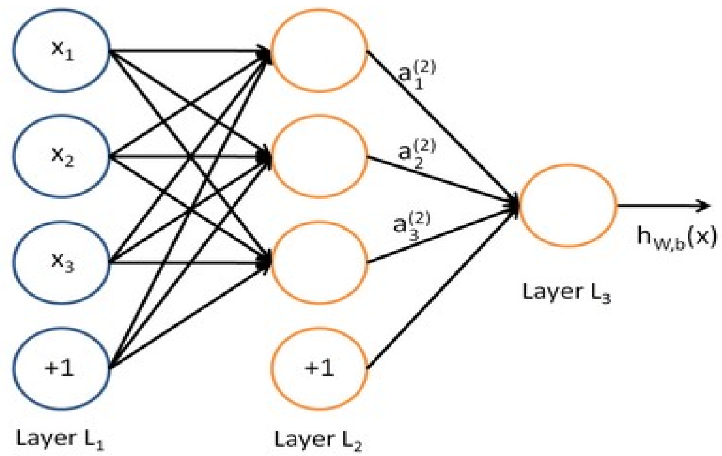

Figure 1. Neural Networks.

In this neural network, X_1, X_2, X_3 are the input of the neural network. +1 is the offset node, also known as the intercept term. The leftmost column of this neural network model is the input layer of the neural network, the rightmost column of which is the output layer of the neural network. The middle layer of the network model is a hidden layer, which is fully connected between the input layer and the output layer. The values of all the nodes in the network model cannot be seen in the training sample set. By observing this neural network model, we can see that the model contains a total of 3 input units, 3 hidden units and 1 output unit.

Now, use n_l to represent the number of layers in the neural network, and the number of layers in this neural network is 3. Now mark each layer, the first layer can be expressed by L_l, then the output layer of the neural network L_1, its output layer is L_{nl}, in this neural network, the following parameters exist:

$$(\mathrm{W}, b) = \left(W^1, b^1, W^2, b^2\right) \tag{2}$$

$W_{ij}^{(l)}$ is the connection parameter between the jth cell of layer 1 and the i th cell of layer l+1, and b_i^l is the offset of the i th cell of layer 1+1. In this neural network model, set $a_i^{(l)}$ to represent the output value of the first few cells in this layer. Let l denote this layer and i the first few cells in this layer.

Given that the set of parameters W and b have been given, we can use the formula $h_{w,b}(x)$ to calculate the output of this neural network. The following formulas are calculation steps:

$$
\begin{aligned}
a_1^2 &= f\left(W_{11}^1 x_1 + W_{12}^1 x_1 + W_{13}^1 x_1 + b_1^{(1)}\right) \\
a_2^2 &= f\left(W_{21}^1 x_1 + W_{22}^1 x_1 + W_{23}^1 x_1 + b_2^{(1)}\right) \\
a_3^2 &= f\left(W_{31}^1 x_1 + W_{32}^1 x_1 + W_{33}^1 x_1 + b_3^{(1)}\right) \\
h_{W,b}(x) &= a_1^3 = f\left(W_{11}^2 a_1^2 + W_{12}^2 a_2^2 + W_{13}^2 a_3^2 + b_1^{(2)}\right)
\end{aligned}
\tag{3}
$$

The calculation of forward propagation is as shown in equation (3). Neural network training methods and Logistic regression model is similar, but due to the multi-layered neural network, but also the need for gradient descent + chain derivation rule.

3. CNN Model Construction and Training

3.1. CNN model

At present, the typical architecture of neural network is divided into the following categories: LeNet5, AlexNet, ZF Net, GooLeNet, and VGGNet, the following will LeNet5 architecture for a detailed analysis. LeNet5 is a CNN classic structure that existed long ago, and it is mainly used in the recognition of handwritten fonts. It contains a total of seven layers of structure, except for the input layer, each of the other has training parameters, and each layer contains a plurality of Feature Maps, we can extract the input features through a convolution kernel. And each feature contains multiple neurons. The picture below shows the architecture of LeNet5:

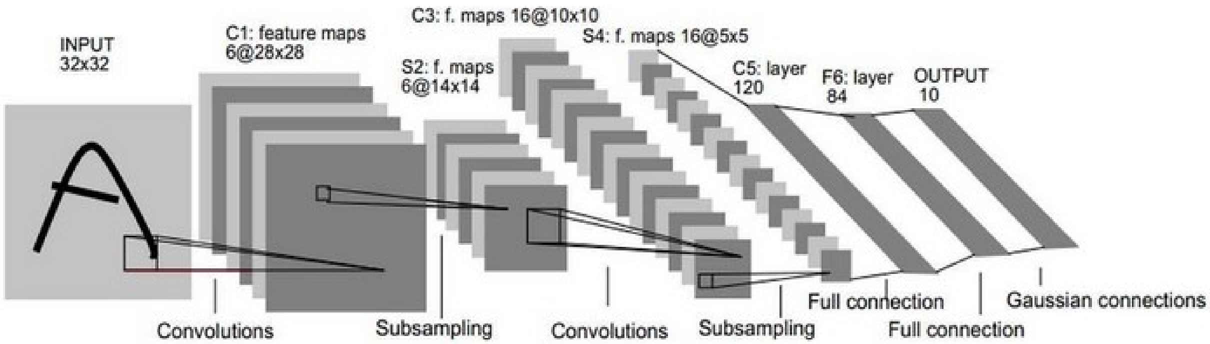

Figure 2. LeNet5 structure diagram.

As shown in Figure 2, a size of 32*32 images through the input layer into the network structure. The layer in the input layer is a convolution layer, which is represented by C1. The number of convolution kernels is 6 and the size is 5*5. After this layer processing, the number of neurons is 28*28*6, trainable parameters are (5*5+1)*6. The next layer of the C1 layer is a downsampled layer, shown in the figure, whose input is the output of the layer convolutional layer, 28*28 in size, 2*2 in the spatial neighborhood of the sample, and the way it is sampled Is to add 4 numbers, multiply them by a trainable parameter, and then add a trainable offset to output the result through the sigmoid function. The number of neurons in layer S2 is 14*14*6. After passing through the S2-layer sampling tube, the size of each feature plot it gets is a quarter of the output from its previous convolution layer. The layer after layer S2 is still a convolutional layer, with a total of 16 convolution kernels, and the size of each convolution kernel is the same as that of C1. This layer is called the C3 layer in the above figure. The size of the output feature layer in this layer is 10*10. The 6 features in the S2 layer are connected with all the features in the C3 layer. The features obtained in this layer the figure is a different combination of the output features of the previous layer. The S4 layer is the same as the S2 layer, and its sampling type is 16. So far, the network structure has reduced the number of neurons to 400. The next layer of C5 is still a convolutional layer, which is fully connected with the previous layer, the size of its convolution kernel is still 5*5, this time C5 layer image processing, the image size becomes 5-5+1=1, which means that only one neuron output, in this layer contains a total of 120 convolution kernel, so the final output of neurons is 120. The last layer of F6, this layer is a fully connected layer, by calculating the input vector and the weight vector between the dot product, plus a bias, and finally through the sigmoid function to deal with the results.

3.2. Face image collection and processing
Image processing based on convolutional neural networks needs to collect a large number of pictures for the computer to learn. This topic will take a lot of people a lot of images, after collecting a lot of images cropped irrelevant parts of the face. This article uses the face detection and cut saved in the created folder.

At this time, the collected images have been trimmed and resized. Then all the images are stitched and stitched in the olivettifaces face dataset, each line represents the category of two people, after all the face images stitching together, and then get the small face database gray degree treatment. The figure below is the subject of the face data set to be trained:

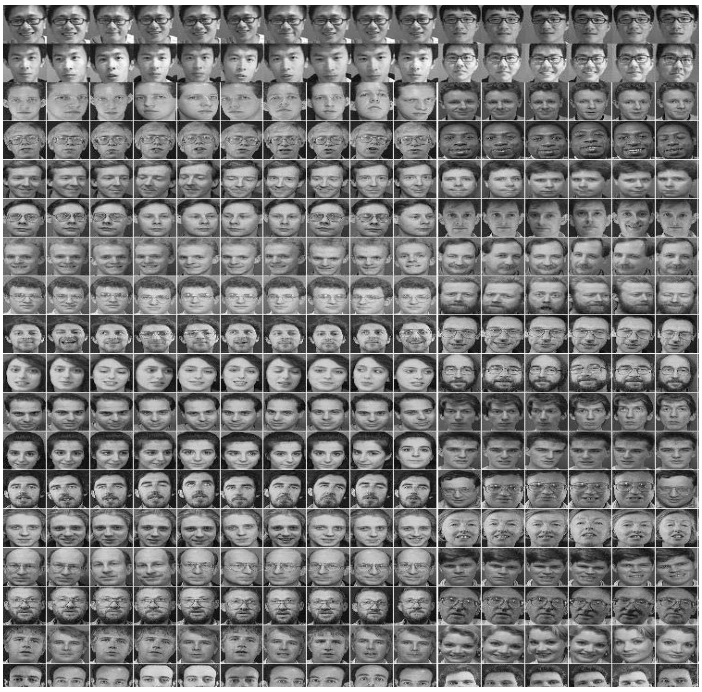

Figure 3. Face data set.

3.3. Convolution neural network model construction
CNN designed in this paper contains the following layers of structure, which are the input layer, conv, pooling, all connected layer, output layer, convolutional layer and the downsampled layer there will be many. In this paper, the reference to LeNet5 model to achieve this article CNN model set up. The design of the model will be a convolution layer and a downsampled layer merged into a layer, named "LeNetConvPoolLayer", a total of two layers "LeNetConvPoolLayer" in the third layer convolution plus sampling layer connected a full connection Layer, named "HiddenLayer", this fully connected layer is similar to the hidden layer in a multi-layer perceptron. The last layer is the output layer, because it is a multi-faceted face classification, so Softmax regression model is used, named "LogisticRegression." Figure 4 for the design of convolution neural network structure:

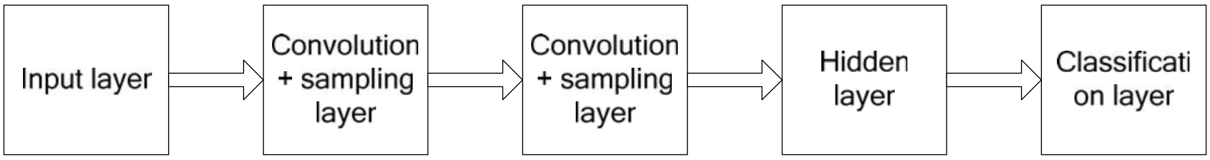

Figure 4. CNN structure diagram.

The first is the input layer of the image input, in this design collected a total of 44 people face, each person's face number is 10, and a total of four 440 sample data, the size of each face image is 57*47=2679. And each image is a grayscale image. The face data set after the face image is collected and processed is the input of the convolution neural network.

The first layer after the input layer is the first convolutional and downsampling layer, the image data input in this layer is 57*47, the size of the convolution kernel is 5 * 5, so the resulting image size after convolution is (57-5+1)*(47-5+1)=(53, 43). After the convolution operation, the image is downsampled to the maximum, resulting in an image size of 26*21.

The input to the second convolution plus sample layer is the output of the first convolution plus sample layer, so the size of the input image in this layer is 26*21. Similar to the operation of convolution plus sampling layer in the first layer, the image is convolution processed first, and the size of the convolutioned image is 22*17. Subsequent image under the maximum downsampling operation, the resulting image size is 11*8.

4. Summary

This paper studies the basic structure of CNN, as well as the basic principles of CNN. Convolutional and downsampled layers of CNN are constructed using the opencv convolution function and the downsampling function. At the same time, the basic principle of multi-layer perceptron MLP is studied to grasp the full connection layer and classification layer, and the use of the Python Python library to achieve. This article simplifies the CNN model by layering the convolutional and sampling layers together. The model consists of two convolution plus sampling layers, a fully connected layer, and a Softmax classification layer. This model is used to train the face data set to optimize the model parameters.

References

[1] Wang Jue, Shi Chunyi. Machine Learning [J]. Journal of Guangxi Normal University (Natural Science Edition), 2013.
[2] Zhang Cuiping, su guangda.Review of face recognition technology [J] .Journal of Image and Graphics, 2015.
[3] Guo Wei, Cai Ning. Convolutional Network Coding [J]. Journal of China Academy of Electronic Science and Technology, 2016.