Journal of Physics: Conference Series

Research on Image Recognition Method Based on Deep Learning

To cite this article: Luo Jing 2019 *J. Phys.: Conf. Ser.* **1395** 012008

Research on Image Recognition Method Based on Deep Learning

Luo Jing

Guangzhou Nanyang Polytechnic College, Guangzhou, 510925, China

yangguanglj@126.com

Abstract. With the rise of artificial intelligence technology, computer vision, image processing, image recognition, voice recognition and other technologies have gradually entered the public's vision. This paper focuses on image recognition. This paper discusses the original image acquisition, image information preprocessing, image feature extraction, pattern classification and recognition results. Analyzes VGG-19 network, VGG improved has realized the image quick recognition, studies input and output points, the number of neurons by convolution neural network, and finally concludes that the deep learning method can shorten the training time, improve the identification results, also the requirement for images in the image library is greatly reduced, the demand for hardware is not too high.

1. Introduction

Deep learning is a new field in machine learning research, which is based on neural network simulating human brain for analysis and learning. Neural network overcomes the shortcomings of traditional artificial intelligence based on logical symbols in dealing with intuitive and unstructured information, and has the features of self-adaptation, self-organization and real-time learning. Deep learning mainly includes convolutional network, loop and recursive network, etc. CNN mainly deal with data with similar grid structure. For example, time series data and image data. With the increasing development of computer hardware, deep learning neural network has made remarkable achievements in computer image recognition [1].

2. Original image acquisition based on deep learning

Nand Flash memory chip, high-speed interface and memory are mainly used in image acquisition system. Compared with the traditional data acquisition system, the current high-speed image acquisition system mainly includes data input acquisition module, display module, storage module, data output module and core control chip FPGA.

Among them, the FPGA is the control core of the system, all modules are connected by the FPGA, and the flow direction of image data in each step is controlled by the FPGA.

The function of the data input acquisition module is to receive the image and video signals transmitted by the digital camera and transmit them to the FPGA for subsequent operation. There is only one kind of acquisition interface, i.e. Camera Link, which is a very popular high-speed camera interface nowadays. The Camera Link interface receives the LVDS signal transmitted by the camera and converts it into TTL signal used in the system and sends it to the FPGA [2].

Display module is an optional auxiliary function, which is used to observe the collected image during the acquisition process or to reproduce the image information stored in the memory chip after acquisition. The interface of the display module is DVI interface. When used, it can work by connecting the DVI display side.

The storage module is also the core of the high-speed image acquisition system. The image data collected by the FPGA is cached and stored in the system. DDR2 SDRAM is used for data caching and finally stored in NAND Flash memory.

3. Image preprocessing

3.1. Image smoothing processing

Choose the jeep in the picture gallery as the object of processing: filter the picture, first define a long window with a length of odd L, N is a positive integer.

$$L = 2N + 1 \tag{1}$$

When loading the picture, the signal samples in the window are $X(i-N),\cdots X(i),\cdots X(i+N)$, where $X(i)$ is the sample value of the signal located in the center of the window. After the sample values of L signal are arranged in order from small to large, the sample values at I are defined as the output value of median filter [3].

Median filtering is a kind of non-linear signal processing technology that can effectively suppress noise. The basic principle of median filtering is to replace the value of a point in a digital image with the median value of each point in a neighborhood of the point, so that the surrounding pixel values are close to the real value, thus eliminating isolated noise points.

When the image is processed by median filtering, the output is set to $g(x, y)$ and the original object is $f(x, y)$. Two-dimensional median filtering is used to complete the formula as follows:

$$g(x, y) = med\{f(x-k, y-1), (k, 1 \in w)\} \tag{2}$$

W refers to a two-dimensional template. The size is generally set to 3*3, 5*5 areas. According to the requirements of the original image, different shapes can be used. Commonly used graphics are linear, circular, cross and circular. Using formula (1) to count the number of signal samples, formula (2) can suppress noise.

3.2. Image sharpening processing

Laplace operator filtering is a linear filtering operation for image sharpening. Laplace operator is a second-order differential operator in n-dimensional Euclidean space, which is defined as divergence $(\vartriangle . f)$ of gradient $(\vartriangle f)$. Therefore, if f is a second-order differentiable real function, then the Laplace operator of f is defined as:

$$\nabla f = \nabla 2 f = \nabla\nabla \bullet f \tag{3}$$

The Laplace operator of f is also all non-mixed second-order partial derivatives in Cartesian coordinates xi:

$$\nabla f = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2} \tag{4}$$

In order to be better applied in image sharpening processing. As a second-order differential operator, Laplace operator maps C function to C function, which is established when $k \geq 2$.

$$\text{Operator } \Delta : C(R) \to C(R) \tag{5}$$

Then the X direction can be expanded as follows:

$$\frac{\partial^2 f}{\partial y^2} = f(x+1, y) + f(x-1, y) - 2f(x, y) \tag{6}$$

The Y direction can be expanded as follows:

$$\frac{\partial^2 f}{\partial y^2} = f(x, y+1) + f(x, y-1) - 2f(x, y) \tag{7}$$

X direction and Y direction are operated and calculated simultaneously.In this expression, we introduce a set of two-dimensional arrays $\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. In the second-order differential operator transformation, the pixel values of the image diagonal are loaded and finally expressed as:

$$\begin{aligned} \nabla 2 f = [& f(x+1, y) + f(x-1, y) + f(x+1, y) + f(x-1, y) + f(x-1, y-1) + \\ & f(x+1, y-1) + f(x-1, y+1) + f(x+1, y+1)] - 8 f(x, y) \end{aligned} \tag{8}$$

Sharpen the image and apply formula (3) to formula (8) to make the blurred image clear .After completing the operation, the corresponding two-dimensional array parameters of the filter are expressed as $\begin{pmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. The function $g(x, y)$ is obtained from the image after multiple processing.

$$g(x, y) = f(x, y) - \nabla 2 f(x, y) \tag{9}$$

The color of digital image ranges from 0 to 255. After filter processing by Laplace operator, the color outside the range may appear. This new color value needs to be re-calibrated. The following functions are adopted to complete.

$$I(x, y) = round\left(\frac{g(x, y) - \min[g(x, y)]}{\max[g(x, y)] - \min[g(x, y)]} \times 255\right) \tag{10}$$

The picture is the contrast effect before and after sharpening treatment. After processing, the outline of the image is clearer and the details are more prominent. After sharpening the image before and after the contrast, using formula (9) and formula (10) to process the image, the outline of the image is clearer and the details are more prominent.

4. HOG image feature extraction

After image processing, image features are analyzed. Image features are extracted mainly from the aspects of color, texture, shape and spatial relationship. The process is divided into the following aspects.

- Consider the image in the image database as a 3D image of x, y, Z (gray level) [4].

- Use Gamma correction method to standardize (Normalize) the color space of the input image; adjust the contrast of the image, and reduce the impact of local shadows and illumination changes of the image, while suppressing noise interference.
- Calculate the gradient of each pixel of the image (including size and direction), capture outline information, and further weaken the interference of light. Divide the image into small pixel units, the commonly used standard is: 6*6 pixels.
- The HOG features of each block can be obtained by concatenating the features of all the pixel units in a block. By concatenating the HOG features of all blocks in the image, the HOG of the detected image can be obtained.

5. Deep learning recognition method

Firstly, the layer name, size, number, step size, extension and trainable parameters of VGGNet (19 layers) are collected. The convolution layer, full connection layer and softmax classification layer in VGGNet (19 layers) are analyzed. In the process of image recognition, we study the performance of full connection layer, and reduce the number of full connection layer by 2.

5.1. Experimental sample data

The image sample database contains some pictures of jeeps, cars and buses. Car pictures include name and model, the same type of car in the sample database, from shooting angle, distance, light, attitude and so on. In some pictures, noise, blur and other interference factors are added [5].

5.2 Establishment of network model

VGGNet (19 layers) removes two layers, determines the number of nodes in each layer, the activated function and its weights among layers, etc. [6].

5.3. Determine the number of nodes in and out of the network.

In the experimental data, in order to improve the recognition rate, a multi-dimensional array is formed based on the feature data of the extracted image, and the number of nodes in the input layer is determined according to the dimension of the multi-dimensional array. In the vehicle picture gallery, the recognition array is composed of five fixed values. So the number of network input points in the experiment is 5. The total category of vehicles is divided into three categories, and the output nodes of the neural network are 3 .

5.4. Experiment

After the experiment of sample parameters, a three-layer network structure is defined by the types of vehicles, which allows the recognition error to be precise to the last two decimal points. The information output of three types of vehicles is (0,0,1), (0,1,0), (1,0,0). [7]. The experimental results are as follows:

Table 1. Recognition of Vehicle Information

No.	Number of Hidden Layer Neurons	Training time (s)	Car recognition rate (%)
1	4	7.22	92

2	9	8.93	93.4
3	10	10.32	98.2
4	11	10.52	96.7
5	30	19.53	89

By comparing the experimental results, it is found that the increase of the number of neurons in the hidden layer leads to the prolongation of training time and the extension of dynamic recognition time, but the recognition rate of the vehicle increases and then decreases gradually after reaching the peak. Increasing the number of neurons and recognition layers blindly will not improve the recognition rate.

Table 2. The test results of the car in

No.	Array invariants	Actual output	Recognition results
1	0.9047 0.3628 0.6921 2.6555 4.9337	0 0 0.9989	√
2	0.8427 0.1054 -0.6493 2.6293 4.9881	1 0 0	×
3	0.9821 0.6146 0.8124 2.6274 4.8144	0 0 0.9999	√
4	0.8973 1.0011 0.8965 2.4946 4.6282	0 0.0026 0.9945	√

6. Conclusion

Deep learning allows computers to construct complex concepts through simpler concepts. By mapping a set of data inputs to outputs of a worthwhile mathematical function, it can improve the recognition rate of image and the process of image recognition is analyzed. This paper analyses VGG-19 network, realizes fast image recognition through improved VGG, studies the input and output points and the number of neurons through CNN, and finally draws a conclusion that compared with the original image recognition method, the deep learning method can shorten the training time, improve the recognition results, and greatly reduce the requirements for the pictures in the image database and the hardware requirements is not too high.

7. Reference

[1] Wu Yubo. *Development and Application of Image Recognition Technology* [J]. Electronic Technology and Software Engineering, 2017 (4): 86

[2] Zhu Renbing. *Research on Face Recognition Method Based on Deep Learning* [D]. Xi'an University of Technology, 2015.

[3] Xie Xiaohua, Lai Jianhuang, Zheng Weishi. *Face Illumination Normalization Based on Quadratic Multinomial Model* [J]. Journal of Electronics, 2010 (8): 1791-1797

[4] Wei Zheng. Research and Implementation of Face Recognition Based on Deep Learning on Caffe Platform [D]. Xi'an University of Electronic Science and Technology, 2015.

[5] Jin Lianwen, Zhong Zhuoyao, Yang Zhao, et al. *A Review of the Application of Deep Learning*

Journal of Physics: Conference Series **1395** (2019) 012008 doi:10.1088/1742-6596/1395/1/012008

in Handwritten Chinese Character Recognition [J]. Journal of Automation, 2016, 42 (8): 1125-1141.

[6] Liu Wanjun, Liang Xuejian, Qu Haicheng. *Image Recognition Algorithm Based on Convolutional Neural Network with Double Optimization* [J]. Pattern recognition and artificial intelligence, 2016, 29 (9): 856-864.

[7] Qinfeng, Liu Dongxia, Sun Bingda, et al. *Image Recognition of Four Alfalfa Leaf Diseases Based on Deep Learning and Support Vector Machine* [J]. Journal of China Agricultural University, 2017, (7): 123-133.

Acknowledgments

Research on the Retrospective Platform of Agricultural Product Quality and Safety in 2019 at the School Level (NY-2019KYYB-27); 2019 Guangdong Higher Vocational College Cloud Computing and Big Data Professional Committee Research Planning Project (GDYJSKT19-13).