Research on Image Recognition Based on Neural Network

Jinhai Zhang*
Shandong Jiaotong University, Weihai, Shandong, China

*Corresponding author: zhang_yi63@163.com

Abstract. With the rapid development of Internet and mobile Internet, the image data in the network shows explosive growth. Image data is simple and intuitive, and contains rich information, which is widely used as a carrier of information exchange. Content-based image recognition can start from the image itself, first extract significant features from the image, and then recognize the image according to the differences of features, with good recognition effect. Neural network plays a very important role in artificial intelligence. Neural network is the key to realize artificial intelligence by learning data through algorithms and then making predictions and decisions. The image recognition of neural network is based on the image content. Neural network starts from the image data itself, extracts the image features of the bottom layer from the data, and then establishes the relationship between the image semantics of the bottom layer and the top layer through the algorithm of neural network, thus realizing the image recognition.

Keywords: Image recognition, Image features, Support vector machine, Neural network

1. Introduction
With the development of Internet and mobile Internet, the storage, transmission and acquisition of information have opened a new chapter. The rapid development of the network cannot be separated from the rapid rise of the army of netizens. When netizens surf the Internet, they also generate a large amount of information while acquiring it. Through Tencent QQ and WeChat developed by Tencent's communication department, real-time communication is carried out through text, voice and video, which greatly promotes information exchange. Through microblog to understand social trends, understand social customs, pay attention to people you like and share your life. Through blog to record their feelings, record their work, study and life, record the problems encountered and solutions through blog posts, posted to the network, to help people solve problems. Through discussion in the forum, focus on a certain field, become relevant experts, cultivate a group of like-minded friends, and promote the development of the industry.

There has been a massive increase in data in the network. Through these devices, you can share your life through various online platforms, opening the connection between individuals and the whole world, so that the scholar does not go out and knows everything in the world. Picture data can show a variety of rich content, and pictures are intuitive and easy to understand. They are widely used to transmit
information and become an important method for people to obtain and share information. However, the amount of image data in the network is very large, so it is very important to identify the meaning of the required pictures from the countless pictures. In the face of massive image data, through machine learning and image recognition technology, images can be accurately and quickly recognized from the network to obtain the desired information.

Machine learning and artificial intelligence have entered people's field of vision and have promoted the development of image recognition. Machine learning is to simulate the human brain, so that machines can have the ability of autonomous learning like human beings. People's learning process is a rather complicated process. People need to know what they want to do, and then carry out practice. In the process of practice, they find out the skills, and then continuously adjust and practice, finally they can learn something. Machine learning is to simulate the learning process of human brain and finally realize the autonomous learning of machine. Human beings mainly acquire the surrounding things through vision, then present images through eyes, then learn the characteristics of images, through certain abstraction and reasoning, finally form relevant concepts in the brain, and then show what they have learned through certain ways. Combining machine learning with image recognition, mining valuable information from images and deriving good products will surely bring great convenience to life.

2. Target Detection Based on Neural Network

Region CNN (R-CNN for short) was proposed by Ross Girshick, which is the foundation work of a series of algorithms for region convolution neural network that uses depth learning for target detection. After DPM algorithm was in a bottleneck period for several years, the accuracy of ILSVRC 2013 data set was significantly improved.

R-CNN did not adopt the previous sliding window scheme, one of the reasons is that the calculation cost is relatively high, which will produce many windows to be classified; Secondly, the width-to-height ratio of different types of rectangular frames previously used is inconsistent, which leads to the inability to use windows of the same size when scanning images. Using linear support vector machine to classify region generation, calculating feature vectors generated by all regions of each input picture, and inputting into SVM for classification; And the position coordinate regression of the object is carried out in the full connection layer.

![Coordinate regression process framework](image)

Fig. 1. Coordinate regression process framework

The input of R-CNN detection method can only be of fixed size. In order to meet the input of this fixed size, one is to intercept the image area that meets the size requirement, which may result in the intercepted image not covering the entire target. Another method is to scale the image, which will produce distortion. However, in CNN, a fixed size is not required for the input convolution layer, only the first layer of the full connection layer needs to input a size of a specific size, because the weight
matrix between the full connection layers needs a size of a specific size, and other full connection layers also do not need a fixed size input. To solve this problem, only one step is needed between the last convolution layer and the first full connection layer to convert images with different sizes into fixed sizes and input them into the full connection layer.

YOLO detection framework is based on the prediction from the feature map of the whole network, so there will be no two stage situation where ROI needs to be intercepted, so the probability of misclassifying the background will be much smaller. What we have learned is a generalized representation, which can be trained on natural images and tested on portraits. Therefore, performance degradation is not easy to occur for input that has not been seen. In contrast, YOLO did not do a good job in positioning accurately, especially for small targets. The whole picture is divided into S x S grids. Assuming that the center of the object to be predicted is in a grid, the object will be predicted from this grid.

3. Image Recognition Based on Neural Network

Neural networks are machine learning algorithms that simulate biological neural networks. The traditional machine learning algorithm is rich in image recognition theory, but the traditional machine learning algorithm requires human design of image features, which requires designers to have rich experience. At the same time, the algorithm has a single structure and poor ability to express complex functions, resulting in low accuracy of image recognition. Artificial neural network has undergone the development of shallow learning and deep learning. Artificial neural network can learn the characteristics of images autonomously, and has advantages in learning massive image data and can improve the accuracy of image recognition. In view of the advantages of artificial neural network, which can extract distributed features autonomously and express complex functions, it plays an important role in image recognition.

The artificial neural network graph has a data input layer, through which n-dimensional data can be input, and then through a series of hidden layers, more reasonable m-dimensional data can be output. Each layer and each layer are completely connected with each other, and the nodes in each layer are independent of each other to form a network structure. Through the relevant algorithm of machine learning, the given image data are preprocessed, the relevant parameter values are initialized first, then the data parameters are updated through the image data used for training, a better artificial neural network model is learned, and then the image recognition accuracy rate of the artificial neural network is counted through the image data used for testing.

There are many training methods for neural network models, and the back propagation algorithm has a very prominent significance in the training of artificial neural networks. The principle of the back propagation algorithm is very simple, that is, the relevant training is carried out through samples, then
the difference between the actual output value and the expected value is calculated, and the relevant loss function can be calculated, and then the relevant weight and deviation are updated through the relevant gradient descent algorithm to obtain the minimum loss function. In order to make the objective function have a good generalization ability, there can be no under-fitting and over-fitting, and to ensure the optimal empirical risk and the optimal structural risk of the artificial neural network model, the artificial neural network model can be trained in the following way, which is shown in Fig 2.

The number of nodes in each layer of the neural network has a great influence on the network. The number of neural network nodes represents the number of learning features. In order for the network to learn as many useful features as possible, it is hoped that the number of nodes can be as large as possible, so that the information of a given picture can be learned more comprehensively and effectively. If there are too many nodes, the network can learn more useful and effective features, but if there are too many nodes, the number of connections between layers will become very large, which is very unfavorable to machine learning. The network structure becomes extremely complex and the learning difficulty becomes very large, so the learning time will become very long, which is not conducive to the training of neural networks. Therefore, it is more important to select a suitable number of nodes.

The number of layers of neural networks will also have a great impact on the learning ability of neural networks. For the simulation of complex functions, the deep network structure can well pass through layer-to-layer transfer, and more complex functions can be simulated through continuous iteration of simpler functions. The deeper the neural network in depth learning is, the more abstract features can be better learned and the semantic meaning of images can be better expressed through layer-to-layer transfer and abstraction, which is very important for image recognition. Shallow complex neural networks can be converted into multi-layer simple neural networks, but simple multi-layer neural network layers and layers are connected in a full connection mode. The more layers, the more complex the network structure, the explosive growth of parameters to be learned. Therefore, it is necessary to select a neural network with appropriate depth to meet the requirements.

The training process of neural network needs a series of steps to generate the required model. With the help of the back propagation algorithm, firstly, the related data need to be normalized to ensure the unification of the data. Then, the weight and deviation need to be initialized by randomization method. Then, the actual value needs to be calculated by the forward propagation algorithm. Then, the related weight needs to be updated by the expected value by the backward propagation algorithm. Finally, the training of the model needs to be completed through continuous iteration.

In the training process, randomization is needed to initialize the weights. A good initialization of weights is very important. If the weights are given unreasonably, it is easy to enter the local optimal solution, causing the weights to saturate and the update speed to slow down, which is very unfavorable to the learning of functions. Therefore, selecting reasonable initial values can ensure that the machine can fully learn to reach the optimal solution.

Convolution layer and pooling layer in convolution neural network are simple cells and complex cells simulating biological neural system. Convolution layer is used for feature extraction. Different convolution check should have different feature planes. Convolution layer reduces network parameters through local connection of receptive fields and weight sharing of convolution kernel, which can simplify the model and further reduce training difficulty. Pooling layer converts multiple pixels of convolution layer into individual pixels, which can be regarded as the secondary extraction of features, can enhance the characteristics of features and avoid over-fitting of network.

The image processing part includes median filter module, Gaussian filter module and Sobel edge extraction module. Median filtering and Gaussian filtering are image preprocessing operations, and Sobel edge extraction module is the feature extraction part of image processing. The three modules are the same as the data receiving module in system design and all have the transparent transmission function to enable the original data to be output, so that the image processing effect can be observed independently during debugging, the coupling of each module can be reduced, and the system-level debugging can be facilitated. The key signal acquisition module is an auxiliary module for Sobel edge
extraction, which is used to assist in adjusting the threshold value for Sobel edge extraction. Although the threshold value for Sobel edge extraction is automatically calculated according to the image characteristics, the process of adding the key adjustment threshold value is favorable for Sobel edge debugging and comparison of processing effects.

The image display part completes the control of VGA display and realizes real-time display of gray image of Y component or processed image. This involves VGA timing standard and analog-to-digital conversion. VGA timing standard is a standard defined in the display industry, so to control VGA displays must be followed.

The internal module of FPGA includes PLL phase locked loop, which is almost the content that FPGA function must involve. The PLL phase-locked loop outputs a 40 MHz system clock processed in FPGA by multiplying the clock input from the crystal oscillator. PLL uses altera's IP core, which has low clock jitter and offset and high clock performance.

The value of each pixel after Gaussian filtering is obtained by weighted average of other pixel values in its domain. However, this system takes 3x3 template as its filtering neighborhood, because taking too large is easy to reduce image edge details and is not conducive to FPGA processing. Therefore, the Gaussian filter module, like the median filter module, first instantiates a 3x3 neighborhood image processing template. Therefore, the image buffer processing method of median filtering 3x3 template is adopted here, and the processing method is the same as that in fig.1. FPGA is easy to implement the template relative to its processing capability and has good adaptability to filter out image noise.

The calculation process of Gaussian filter is divided into two steps: weighted sum, multiplying the data in the template by the corresponding coefficients, and then finding their sum; Take the average output, that is, divide by 16. Here, the operation of dividing by 16 can directly remove the high 8-bit data of the sum, i.e. the low 4-bit data, which is convenient and does not consume the resource consumption brought by the division operation.

Edge detection module is the core of image processing. Due to the pipelined processing mode of FPGA, many algorithms based on PC software are difficult to implement and work, such as QSTU algorithm based on entropy selection algorithm and variance between classes. At present, the edge gradient threshold of FPGA for edge extraction basically adopts a fixed empirical threshold. This processing method has poor generality and needs to adjust the threshold in a specific environment.

This system makes full use of the advantages of FPGA pipelining and parallel computing, and realizes Sobel edge extraction method with non-maximum suppression and threshold adaptive selection. After non-maximum suppression processing, the edge thinning effect can be achieved. At the same time, threshold selection can better remove false edges and capture details.

The module mainly includes gradient operation part, threshold selection part, image buffer part and binary display edge part. Sobel gradient operation part realizes the multi-direction gradient operation of 5x5 pixel window center 9 pixels, and determines whether it is the edge center by comparing the gradients of two adjacent pixels in the maximum gradient direction of the center point, thus achieving the effect of edge thinning. The threshold selection module calculates the maximum value of the second derivative based on the gradient histogram, which is used as the threshold for edge extraction. The image gradient buffer module buffers the gradient value of the central pixel of 5x5 windowing buffer. The binary edge display section compares the gradient with the threshold, and if the gradient is greater than the threshold, displays the central pixel of the 5x5 windowing buffer as an edge.

The gradient calculation of the center point passes through a 5x5 windowing buffer section, a four-direction gradient calculation section of the center 3x3 pixel, a four-direction gradient comparison section and a non-maximum suppression section, and finally obtains the gradient value of the center vertex of the center 5x5 pixel. The shift register uses an IP core (alt shift _ tapes) with a depth of 640 and a width of 8 bits. the 5x5 register groups buffer 25 image data. when the image flows through the module, it traverses the whole frame of image, and the gradient values of all pixels can be calculated. In order to calculate the large data cache without accumulated delay, the statistics module is implemented purely by LE and register logic, and does not use internal DRAM resources. The advantage of this is
that there is no calculation delay, but it will consume more LE resources. The comparison module calculates the second derivative of the current input gradient value. When a frame of pixel gradient value flows through the comparison module, all second derivatives are traversed once, so the maximum second derivative can be obtained.

From the processing process of the threshold selection module, it can be seen that the threshold of the current frame of the video is obtained through statistics of the image of the previous frame. Since the images of two adjacent frames differ unintentionally, there is no influence on the convenience of extraction. This module calculates a gradient threshold according to the specific scene of each frame of image, so it has the function of threshold adaptive selection.

4. Conclusion
The recognition model of artificial neural network is a machine learning model simulating biological neural system. Artificial neural network solves the shortage of artificial feature extraction. Artificial feature extraction requires experts with rich experience to ensure better feature extraction. Artificial neural network can extract features autonomously and then use them for image recognition. This paper focuses on the design of the number of nodes. In order to ensure effective feature extraction, a large number of nodes is required. However, too many nodes will lead to too many network parameters, increased learning difficulty, decreased learning speed and prone to over-fitting. Therefore, the number of nodes is half of the dimension of input data, which can better identify images. This paper also studies the depth of the network. With more depth of the network, the network can learn more abstract and better structured features through layer-by-layer transmission, which can improve the recognition rate of pictures. However, the deeper the depth of the network, the slower the learning speed of the network and the over-learning phenomenon will also occur. It is necessary to reasonably select the depth of the network. This paper demonstrates that the network with two hidden layers has better recognition effect than the network with one layer.

References
[1] Han S, Mao H, Dally W J, et al. Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding[J]. international conference on learning representations, 2016.
[2] Courbariaux M, Hubara I, Soudry D, et al. Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1[J]. 2016.
[3] Hinton G, Vinyals O, Dean J. Distilling the Knowledge in a Neural Network[J]. Computer Science, 2017, 14(7):38-39.
[4] Gu Yan. based on data warehouse technology in hospital information systems (HIS) Scheme [j]. computer systems applications. 2018
[5] Yi Jing. and implementation of data mining technology in the hospital information [j]. Chongqing University of medical sciences 2017.
[6] Kantardzic M. Data Mining Concept、 Models、 Methods and Algorithms[M]. IEEE Press.2016
[7] Bryan Ford, Pyda Srisuresh,and Dan Kegel.Peer-to-Peer Communication Across Network Address Translators.In USENIX Annual Technical Conference, Anaheim, CA, April,2015.
[8] Hill R,Wang J, Nahrstedt K. Quantifying Non-functional Requirements: A Process Oriented Approach[C]. IEEE International Requirements Engineering Conference,Kyoto,Japan,2018:352-353