EXPERIMENTAL REPORT OF FACE RECOGNITION BASED ON MACHINE LEARNING

Hanqing Liu
School of Mathematical Sciences
Nankai University, Tianjin, China
hanqliu@163.com

ABSTRACT

With the deep development of computer science, face recognition, as an important component of biometric information recognition, including specific tasks such as face detection, feature extraction and recognition, and face recognition, has gradually become one of the popular research areas in recent decades. Machine learning methods, including deep learning methods that have been rapidly revitalized within this century, are then providing excellent and reliable solutions to problems in an increasing number of computer science fields. In this paper, we will summarize the main work and methods in the field of machine learning, and introduce the excellent work done in the field of face recognition using machine learning methods. Also, the article will introduce several classical methods for face recognition.

Keywords: face recognition, biometric information recognition, machine learning, deep learning, neural network

1. EXPERIMENTAL OBJECTIVES AND TASKS

Face recognition, is a biometric technology for identity recognition based on the information of human face features. A series of related technologies that use a camera or a camera to capture an image or video stream containing a human face, and automatically detect and track the face in the image, and then perform face recognition on the detected face, are often also called portrait recognition and facial recognition.

The research of face recognition system started in 1960s, and was improved after 1980s with the development of computer technology and optical imaging technology, while it really entered the primary application stage in late 1990s, and was mainly realized by the technology of the United States, Germany and Japan; the key to the success of face recognition system lies in whether it has the cutting-edge core algorithm and makes the recognition result have practical recognition rate and recognition The "face recognition system" integrates various professional technologies such as artificial intelligence, machine recognition, machine learning, model theory, expert system, video image processing, etc., and needs to combine the theory and implementation of intermediate value processing, which is the latest application of biometric recognition. It is the latest application of biometric recognition, and its core technology implementation shows the transformation of weak artificial intelligence to strong artificial intelligence.

The face recognition system mainly includes four components, which are: face image acquisition and detection, face image pre-processing, face image feature extraction, and matching and recognition.

1.1 Face image acquisition and detection

1.1.1 Face image acquisition

Different face images can be captured by the camera lens, such as static images, dynamic images, different positions, different expressions and other aspects can be well captured. When the user is within the capture range of the capture device, the capture device will automatically search and capture the user's face image.

1.1.2 Face detection

Face detection is mainly used in practice for pre-processing of face recognition, i.e., to accurately calibrate the position and size of a face in an image. The face image contains rich pattern features, such as histogram features, color features, template features, structure features and Haar features. Face detection is to pick out the useful information from these and use these features to achieve face detection.

The mainstream face detection method uses Adaboost learning algorithm based on the above features. Adaboost algorithm is a method used for classification, which combines some relatively weak classification methods together to combine new very strong classification methods.

The face detection process uses the Adaboost algorithm to select some rectangular features (weak classifiers) that best represent the face, constructs the weak classifier into a strong classifier according to weighted voting, and then connects a number of trained strong classifiers in series to form a cascaded classifier with a cascade structure, which effectively improves the detection speed of the classifier.

1.2 Face image pre-processing

Image pre-processing for faces is the process of processing images based on face detection results and ultimately serving feature extraction. The original image acquired by the system is often not directly usable due to various conditions and random interference, and it must be subjected to image pre-processing such as grayscale correction and noise filtering at the early stage of image processing. For face images, the pre-processing process mainly includes light compensation, grayscale transformation, histogram equalization, normalization, geometric correction, filtering and sharpening of face images.

1.3 Face image feature extraction

The features that can be used in face recognition systems are usually classified as visual features, pixel statistical features, face image transformation coefficient features, face image algebraic features, etc. Face feature extraction is done for certain features of the face. Face feature extraction, also known as face characterization, is the process of feature modeling of a human face. The methods of face feature extraction are summarized into two main categories: one is knowledge-based characterization methods; the other is characterization methods based on algebraic features or statistical learning.

The knowledge-based characterization method is mainly based on the shape descriptions of face organs and the distance characteristics between them to obtain feature data that can help face classification, and its feature components usually include Euclidean distance, curvature and angle between feature points. The human face consists of parts such as eyes, nose, mouth, and chin, etc. Geometric descriptions of these parts and the structural relationships between them can be used as important features for face recognition, and these features are called geometric features. Knowledge-based face characterization mainly includes geometric feature-based methods and template matching methods.

1.4 Face image matching and recognition

The feature data of the extracted face image is searched and matched with the feature template stored in the database by setting a threshold value, and when the similarity exceeds this threshold value, the result obtained by matching is output. Face recognition is to compare the features of the face to be recognized with the obtained face feature template, and to judge the identity information of the face according to the degree of similarity. This process is further divided into two categories: confirmation, which is a one-to-one image comparison process, and recognition.

1.5 Advantages of Face Recognition

The advantage of face recognition lies in its naturalness and its imperceptibility to the individual being measured.

By naturalness, we mean that the recognition method is the same as the biometric features used by humans (or even other organisms) for individual recognition. For example, face recognition, humans also distinguish and confirm their identity by observing and comparing faces, and other natural recognition includes voice recognition, body shape recognition, etc. Fingerprint recognition, iris recognition, etc. are not natural because humans or other creatures do not distinguish individuals by such biometric features.
Undetectability is also important for an identification method, which makes it less offensive and less likely to be spoofed because it is less likely to attract attention. Face recognition has this characteristic, it uses entirely visible light to acquire face image information, unlike fingerprint recognition or iris recognition, which require the use of electronic pressure sensors to capture fingerprints, or infrared light to capture iris images, these special capture methods are easily detected and thus more likely to be deceived by disguise.

1.6 Difficulties of face recognition

Face recognition is considered to be one of the most difficult research topics in the field of biometric recognition and even in the field of artificial intelligence. The difficulty of face recognition is mainly brought about by the characteristics of human face as a biometric feature.

1.6.1 Similarity

There is little difference between different individuals, and all faces have similar structures, and even the structural appearance of face organs is similar. Such a feature is advantageous for localization using faces, but unfavorable for distinguishing human individuals using faces.

1.6.2 Fickleness

The appearance of human face is very unstable, and people can produce many expressions through face changes, and the visual images of human face vary greatly in different observation angles. In addition, face recognition is also affected by various factors such as lighting conditions (e.g. day and night, indoor and outdoor, etc.), many coverings of human face (e.g. masks, sunglasses, hair, beard, etc.), and age.
In face recognition, the first class of variations is supposed to be enlarged and used as a criterion to distinguish individuals, while the second class of variations should be eliminated because they can represent the same individual. The first class of

variation is usually referred to as inter-class difference, while the second class of variation is referred to as intra-class difference. For faces, intra-class variation is often greater than inter-class variation, making it exceptionally difficult to distinguish individuals using inter-class variation when disturbed by intra-class variation.

2. BACKGROUND KNOWLEDGE - ABOUT DIGITAL IMAGE PROCESSING

2.1 Main purpose

Generally speaking, there are three main purposes for processing (or processing and analysis) of images.

2.1.1 Improve the visual quality of the image, such as performing image brightness and color transformation, enhancing and suppressing certain components, and performing geometric transformation of the image to improve the quality of the image.

2.1.2 Extraction of certain features or special information contained in an image that are extracted often facilitates computer analysis of the image. The process of extracting features or information is a pre-processing for pattern recognition or computer vision. The extracted features can include many aspects, such as frequency domain features, grayscale or color features, boundary features, region features, texture features, shape features, topological features, and relationship structures.

2.1.3 Transformation, encoding and compression of image data for image storage and transmission.

Regardless of the purpose of image processing, an image processing system consisting of a computer and image-specific equipment is required to input, process, and output image data.

2.2 Common methods

The following methods are commonly used for digital image processing.

2.2.1 Image transformation

Due to the large image array, processing directly in the spatial domain involves a large amount of computation. Therefore, various methods of image transformation, such as Fourier transform, Walsh transform, discrete cosine transform and other indirect processing techniques are often used to convert the processing in the spatial domain to transform domain processing, which not only reduces the computational effort but also results in more efficient processing (e.g., Fourier transform can be digitally filtered in the frequency domain). The emerging research wavelet transform has good localization properties in both time and frequency domains, and it also has wide and effective applications in image processing.

2.2.2 Image coding compression

Image coding compression techniques reduce the amount of data (i.e., the number of bits) describing an image in order to save image transmission, processing time and reduce the memory capacity occupied. Compression can be obtained without distortion or with the allowed distortion. Coding is the most important method in compression techniques, and it is the first and more mature technique developed in image processing technology.

2.2.3 Image enhancement and restoration

The purpose of image enhancement and restoration is to improve the quality of the image, such as removing noise and improving the sharpness of the image. Image enhancement does not consider the reason for image degradation and highlights the part of the image of interest. For example, enhancing the high-frequency component of the image can make the outline of objects in the image clear and details obvious; for example, enhancing the low-frequency component can reduce the impact of noise in the image. Image restoration requires a certain understanding of the causes of image degradation, and generally speaking, a "degradation model" should be established according to the degradation process, and then a certain filtering method should be used to restore or reconstruct the original image.

2.2.4 Image segmentation

Image segmentation is one of the key techniques in digital image processing. Image segmentation is the extraction of meaningful feature parts of an image, whose meaningful features are edges and regions in the image, which are the basis for further image recognition, analysis and understanding. Although many methods of edge extraction and region segmentation have been researched, there is not yet an effective method that is universally applicable to various images. Therefore, the research on image segmentation is still in progress, and it is one of the hot spots of research in image processing.

2.2.5 Image description

Image description is a necessary prerequisite for image recognition and understanding. As the simplest binary image can be used to describe the characteristics of the object using its geometric properties, the general image description method uses two-dimensional shape description, which has two types of methods: boundary description and region description. For special texture images can be described by two-dimensional texture features. With the in-depth development of image processing research, the research of three-dimensional object description has been started, and methods such as volume description, surface description, and generalized cylinder description have been proposed.

2.2.6 Image classification (recognition)

Image classification (recognition) belongs to the category of pattern recognition, the main content of which is image segmentation and feature extraction after some pre-processing (enhancement, recovery, compression), so as to carry out judgment classification. Image classification often uses classical pattern recognition methods with statistical pattern classification and syntactic (structural) pattern classification, and in recent years the newly developed fuzzy pattern recognition and artificial neural network pattern classification have also received increasing attention in image recognition.

3. BACKGROUND KNOWLEDGE - MACHINE LEARNING

Machine learning is a multidisciplinary profession, covering knowledge of probability theory, statistics, approximation theory, and complex algorithms, using computers as tools and aiming to simulate human learning in real time, and structuring existing content to effectively improve learning efficiency.

Machine learning is a common research hotspot in the field of artificial intelligence and pattern recognition, and its theories and methods have been widely used to solve complex problems in engineering applications and science. 2010 Turing Award winner is Professor Leslie vlliant of Harvard University for, among other things, establishing Probably Approximate .The 2011 Turing Award winner was Professor Judea Pearll of the University of California, Los Angeles, whose main contribution was the establishment of an artificial intelligence approach based on probabilistic statistics. All these research results have contributed to the development and prosperity of machine learning.

Machine learning is the science of how to use computers to simulate or implement human learning activities, and it is one of the most intelligent and cutting-edge research areas in artificial intelligence. Since 1980s, machine learning has attracted wide interest in the artificial intelligence community as a way to achieve artificial intelligence, especially in the last decade or so, research work in the field of machine learning has developed rapidly and it has become one of the important topics of artificial intelligence. Machine learning has been applied not only in knowledge-based systems, but also in many fields such as natural language understanding, non-monotonic reasoning, machine vision, and pattern recognition. Whether a system has the ability to learn or not has become a sign of "intelligence". Machine learning research is mainly divided into two types of research directions: the first type is the research of traditional machine learning, which mainly studies the learning mechanism and focuses on exploring the learning mechanism of simulating human; the second type is the research of machine learning in big data environment, which mainly studies how to effectively use information and focuses on obtaining hidden, effective and understandable knowledge from huge amount of data.

After 70 years of tortuous development, machine learning has reaped breakthroughs in many aspects, most representative of which is in the field of image recognition, by drawing on the multi-stratified structure of the human brain, the layer-by-layer analysis and processing mechanism of neuron connection interaction information, and the powerful parallel information processing capability of self-adaptation and self-learning.

4. ADABOOST-BASED FACE DETECTION METHOD

Adaboost is an iterative algorithm whose core idea is to train different classifiers (weak classifiers) for the same training set, and then aggregate these weak classifiers to form a stronger final classifier (strong classifier).

Boosting, also known as augmented learning or boosting, is an important integrated learning technique that can augment a weak learner with prediction accuracy only slightly higher than random guesses into a strong learner with high prediction accuracy, which provides an effective new idea and method for the design of learning algorithms when it is very difficult to construct strong learners directly. As a meta-algorithmic framework, Boosting can be applied to almost all currently popular machine learning algorithms to further enhance the prediction accuracy of the original algorithm, and it is widely used and has a great impact. AdaBoost is one of the most successful representatives, and is rated as one of the top ten algorithms for data mining. The success of AdaBoost is not only that it is an effective learning algorithm, but also that it has turned Boosting from a conjecture to a real practical algorithm. Some techniques used by the algorithm, such as breaking the original sample distribution, have also brought important insights to the design of other statistical learning algorithms; the related theoretical research results have greatly promoted the development of integrated learning.

The algorithm is actually a simple process of weak classification algorithm enhancement, and this process can improve the classification ability of data by continuous training. The whole process is shown as follows.

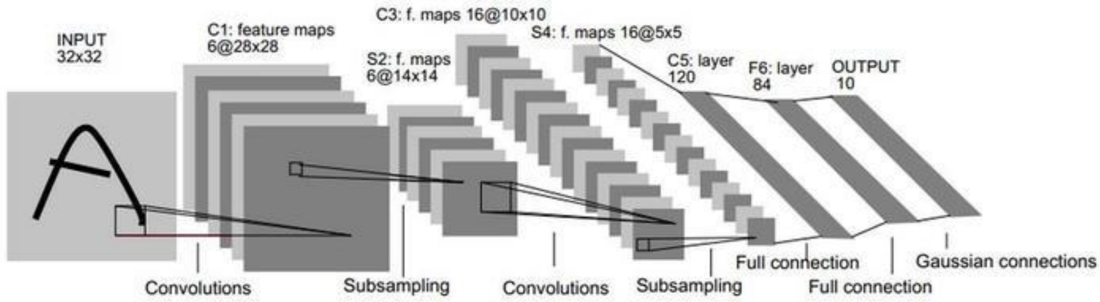
1. first obtain the first weak classifier by learning on N training samples.
2. to form a new training sample of N with the misclassified samples and other new data, and to obtain a second weak classifier by learning from this sample.
3. the third weak classifier is obtained by learning from the samples that were misclassified in both 1 and 2 plus other new samples to form another new training sample of N.
4. the final boosted strong classifier. That is, the class into which a certain data is classified is determined by the respective classifier weights.

From the process described in the Adaboost algorithm, it is clear that the algorithm is implemented to initialize the sample weights according to the size of the training set so that they satisfy a uniform distribution, and in subsequent operations the weights of the samples after the iterations of the algorithm are changed and normalized by the formula. The samples are misclassified resulting in an increase in the weights and a corresponding decrease in the weights conversely, which indicates that the misclassified training sample set includes a higher weight. This causes the training sample set to focus more on the hard-to-identify samples in the next round, and further learning for the misclassified samples is used to obtain the next weak classifier until the samples are correctly classified. The strong classifier construction is completed when the specified number of iterations or the expected error rate is reached.

The Adaboost algorithm system has a high detection rate and is less prone to over-adaptation. However, the implementation of this algorithm requires a large training sample set in order to achieve higher detection accuracy. In each iteration, a weak classifier is trained for each sample in that sample set, and each sample has many features, so the computational effort to train the optimal weak classifier from the large number of features increases. The typical Adaboost algorithm uses a backtracking search mechanism, and although a greedy algorithm is used to obtain the best local weak classifier each time when training a weak classifier, it does not ensure that the overall best one is selected after weighting. After selecting the weak classifier with minimum error, the weights of each sample are updated to increase the weights corresponding to the misclassified samples and reduce the weights of the correctly classified samples relatively. The execution effect depends on the selection of the weak classifier, and the search time increases, so the training process makes the time used by the whole system very large, and therefore limits the wide application of the algorithm. On the other hand, in the implementation of the algorithm, the cascade classifier is constructed by gradually approximating the expected value in terms of both detection rate and false recognition rate of positive samples, and this construction process can be realized only after iterative training to generate a large number of weak classifiers. This leads to the introduction of circular approximation for training classifiers that consume more time.

5. LENET5-BASED FACE RECOGNITION METHOD

LeNet5 is a small network, but it contains the basic modules of deep learning: convolutional layer, pooling layer, and full linkage layer. It is the foundation of other deep learning models, and we will analyze LeNet5 in depth here. At the same time, we will deepen our understanding of the convolutional and pooling layers through example analysis.



LeNet-5 has 7 layers without input and each layer contains trainable parameters; each layer has multiple Feature Maps, each FeatureMap extracts one feature of the input through one type of convolutional filter, and then each FeatureMap has multiple neurons.

5.1 INPUT layer - input layer

First is the data INPUT layer, where the size of the input image is uniformly normalized to 32*32. This layer is not considered a LeNet-5 network structure, and traditionally, the input layer is not considered as one of the network hierarchies.

5.2 C1 layer - convolutional layer

Input image: 32*32

Convolution kernel size: 5*5

Convolution kernel type: 6

Output featuremap size: 28*28 (32-5+1)=28

Number of neurons: 28*28*6

Trainable parameters: (5*5+1) * 6 (5*5 per filter = 25 unit parameters and one bias parameter, 6 filters in total)

Number of connections: (5*5+1)*6*28*28=122304

The first convolution operation is performed on the input image (using 6 convolutional kernels of size 5*5) to obtain 6 C1 feature maps (6 feature maps of size 28*28, 32-5+1=28). For the convolutional layer C1, each pixel in C1 is connected to 5*5 pixels and 1 bias in the input image, so there are 156*28*28=122304 connections in total. connections. There are 122,304 connections, but we only need to learn 156 parameters, which is mainly achieved by weight sharing.

5.3 S2 layer - pooling layer (lower sampling layer)

Input: 28*28

Sampling area: 2*2

Sampling method: 4 inputs are summed, multiplied by a trainable parameter, and a trainable bias is added. Results are obtained by sigmoid

Sampling type: 6

Output featureMap size: 14*14 (28/2)

Number of neurons: 14*14*6

Number of connections: (2*2+1)*6*14*14

The size of each feature map in S2 is 1/4 of the size of the feature map in C1.

The first convolution is followed by a pooling operation using a 2*2 kernel, resulting in S2, six 14*14 feature maps (28/2=14). S2 is a pooling layer that sums the pixels in the 2*2 region of C1 and multiplies them by a weight factor plus a bias, and then maps the result again. Also there are 5x14x14x6=5880 connections.

5.4 C3 layer - convolutional layer

Input: combination of all 6 or several feature maps in S2

Convolution kernel size: 5*5

Convolution kernel type: 16

Output featureMap size: 10*10 (14-5+1)=10

Each feature map in C3 is connected to all 6 or several feature maps in S2, indicating that the feature maps in this layer are different combinations of the feature maps extracted from the previous layer

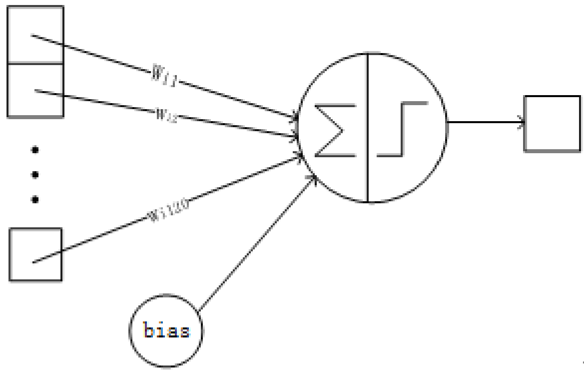
One way of existence is that the first 6 feature maps of C3 take as input a subset of 3 adjacent feature maps in S2. The next 6 feature maps take as input a subset of 4 adjacent feature maps in S2. The next 3 take as input a subset of 4 non-adjacent feature maps. The last one takes all feature maps in S2 as input. Then: trainable parameters: 6*(3*5*5+1)+6*(4*5*5+1)+3*(4*5*5+1)+1*(6*5*5+1)=1516

Number of connections: 10*10*1516=151600

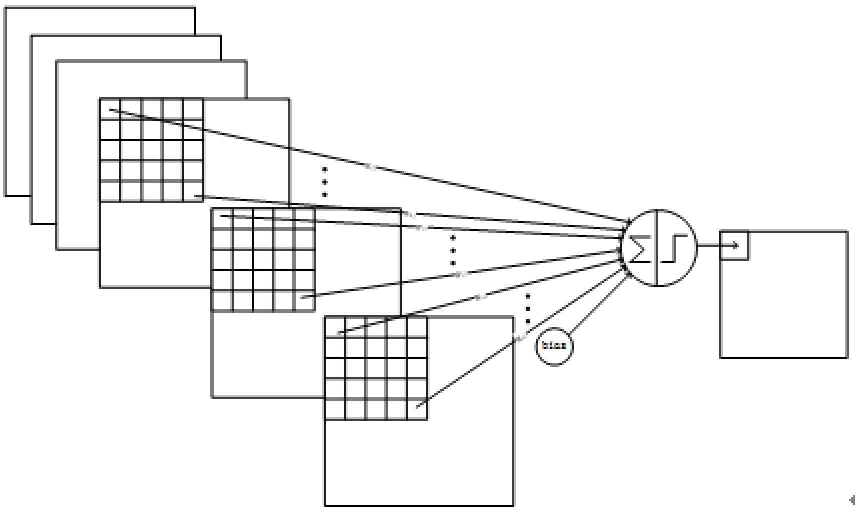
After the first pooling is the second convolution, the output of the second convolution is C3, 16 10x10 feature maps, and the size of the convolution kernel is 5*5. We know that S2 has 6 14*14 feature maps, how can we get 16 feature maps from 6 feature maps? Here is a special combination of S2's feature maps to get 16 feature maps. The details are as follows.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

The first 6 feature maps of C3 (corresponding to the 6 columns in the first red box above) are connected to the 3 feature maps connected to the S2 layer (the first red box above), the next 6 feature maps are connected to the 4 feature maps connected to the S2 layer (the second red box above), the next 3 feature maps are connected to the 4 feature maps not connected to the S2 The last one is connected to all the feature maps in the S2 layer. The size of the convolution kernel is still 5*5, so there are 6*(3*5*5+1)+6*(4*5*5+1)+3*(4*5*5+1)+1*(6*5*5+1)=1516 parameters in total. And the image size is 10*10, so there are 151600 connections.



The convolutional structure of C3 connected to the first 3 graphs in S2 is shown in the following figure.



The above figure corresponds to the parameter 3*5*5+1, and a total of 6 convolutions are performed to obtain 6 feature maps, so there are 6*(3*5*5+1) parameters. Why is this combination used? The paper says there are two reasons: 1) to reduce the parameters, and 2) this asymmetric combination of connections is beneficial to extract multiple combinations of features.

5.5 S4 layer - pooling layer (lower sampling layer)

Input: 10*10

Sampling area: 2*2

Sampling method: 4 inputs are summed, multiplied by a trainable parameter, and a trainable bias is added. Results are obtained by sigmoid

Sampling type: 16

Output featureMap size: 5*5 (10/2)

Number of neurons: 5*5*16=400

Number of connections: 16*(2*2+1)*5*5=2000

The size of each feature map in S4 is 1/4 of the size of the feature map in C3

S4 is the pooling layer, the window size is still 2*2, a total of 16 feature maps, the 16 10x10 maps of C3 layer are pooled in 2x2 units to get 16 5x5 feature maps. There are 5x5x5x16 = 2000 connections. The connections are made in a similar way to the S2 layer.

5.6 C5 layer - convolutional layer

Input: all 16 cell feature maps of the S4 layer (fully connected to s4)

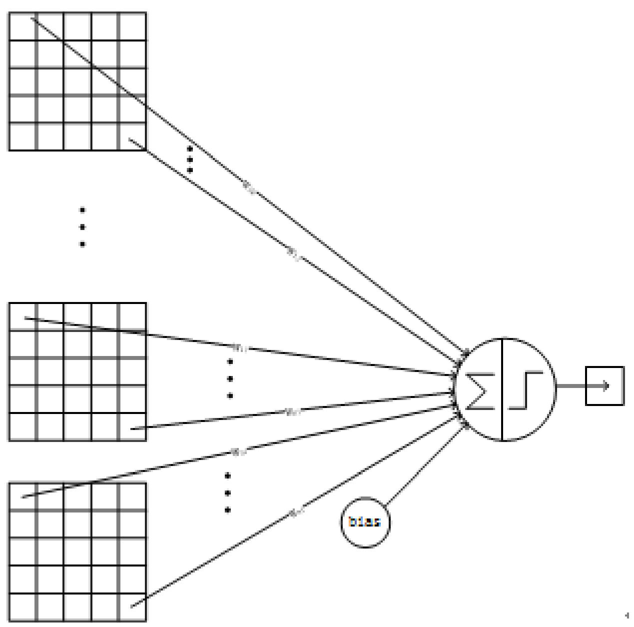
Convolution kernel size: 5*5

Convolution kernel type: 120

Output featureMap size: 1*1 (5-5+1)

Trainable parameters/connections: 120*(16*5*5+1)=48120

Layer C5 is a convolution layer. Since the size of the 16 graphs in layer S4 is 5x5, which is the same size as the convolution kernel, the size of the graph formed after convolution is 1x1. 120 convolution results are formed here. Each of them is connected to the 16 graphs of the previous layer. So there are (5x5x16+1)x120 = 48120 parameters and also 48120 connections. the network structure of C5 layer is as follows.



5.7 F6 Layer - Fully Connected Layer

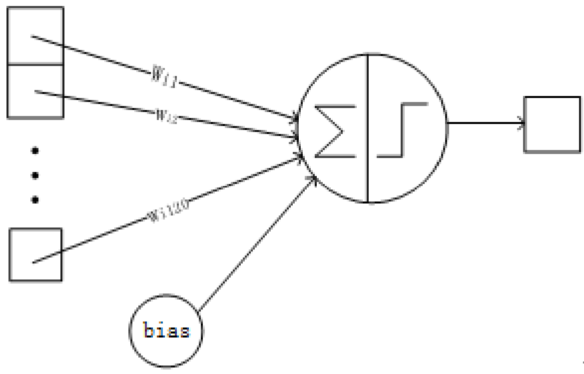
Input: c5 120-dimensional vector

Calculation: Calculate the dot product between the input vector and the weight vector, add a bias, and the result is output by the sigmoid function.

Trainable parameters:84*(120+1)=10164

Layer 6 is a fully connected layer. layer F6 has 84 nodes corresponding to a 7x12 bitmap, -1 for white and 1 for black, such that the black and white of the bitmap for each symbol corresponds to an encoding. The training parameters and the number of connections for this layer are (120 + 1) x 84 = 10164. The ASCII encoding diagram is as follows.

The F6 layer is connected as follows.



5.8 Output Layer - Fully Connected Layer

The Output layer is also a fully connected layer with 10 nodes representing the numbers 0 to 9 and the result of network identification is the number i if the value of node i is 0. A radial basis function (RBF) network connection is used. Assuming that x is the input of the previous layer and y is the output of the RBF, the RBF output is calculated as follows

$$y_i = \sum_j (x_j - w_{ij})^2.$$

The value of the above equation wij is determined by the bitmap encoding of i. i goes from 0 to 9 and j takes values from 0 to 7*12-1. The closer the value of the RBF output is to 0, the closer it is to i, i.e., the closer it is to the ASCII-encoded graph of i, indicating that the current network input recognition results in character i. This layer has 84x10=840 parameters and connections.

6. LEARNING EXPERIENCE AND FEELINGS

6.1 The research practical training has improved my programming and understanding of algorithms

Due to the curriculum of my major, I was often able to understand the mathematical derivation in the process of self-learning algorithms, but sometimes it was difficult to understand the concept and idea of model design, which led to the inability to adjust the parameters to make the model work for me. The teacher's explanation of algorithms and models, especially the introduction of design ideas and development changes, helped me solve this problem to a large extent.

6.2 The research training deepened my knowledge of traditional machine learning and gave me a complete understanding of the logic and ideas of mainstream neural network algorithms

I was unfamiliar with the overall framework of machine learning because of the "look up what you need" mode, but the teacher's explanation helped me to connect the knowledge points and provide a good knowledge framework. At the same time, the vivid classroom demonstration gave me a full and clear understanding of the logic of neural networks.

6.3 The research practical training has provided new ideas for my research projects

In addition to completing our research training task, many of the concepts and ideas in target recognition have provided ideas for my own research projects that I would like to address. For example, in web text sentiment analysis, it is sometimes necessary to analyze the sentiment of emoticons. By using the idea of Haar-like features, we may be able to recognize the targets of simple structured images like expressions and perform sentiment analysis, such as whether the corners of the mouth are raised or not, and whether there are dimples at the corners of the mouth to determine whether the emotion of the expression is positive or negative.

6.4 The process of research training has provided me with a new perspective on things

During my scientific research, I have gained a more practical and tangible understanding of some mathematical methods, such as the computation of pixel matrices to achieve the processing of images, the use of knowledge of graph theory and data structures to solve clustering problems, and so on. It allows formal science to project itself in the natural and social dimensions.