Face Recognition Technology Analysis Based on Deep Learning Algorithm

Li Liang
Jiangsu Vocational Institute of Commerce, Nanjing, Jiangsu, China

Abstract. With the Explosive Development of Deep Learning Technology Face Recognition and Other Recognition Technologies Mostly Adopt Deep Learning Algorithm for Recognition. Although the Deep Learning Algorithm Has High Recognition Accuracy, It Has a Huge Demand for Computing. in the Mobile Terminal, We Can Use Artificial Intelligence Chips That Can Accelerate Deep Learning Operations to Complete Relevant Operations. Deep Learning Has Fixed Modes, Like Convolution. Ai Chips Can Significantly Improve the Efficiency of Deep Learning Operations by Optimizing the Corresponding Operation Modes. in This Way, Mobile Terminals Can Quickly Implement Complex Deep Learning Operations, Such as Face Recognition Based on Deep Learning. One Representative of Ai Chips is the Tensor Processing Unit of Google, Which is Able to Accelerate the Tensor Flow of the Deep Learning System, Which is Far More Efficient Than Gns. the Tpu Provides 1,530 Times the Performance Improvement and 3,080 Times the Efficiency (Performance/Watt) Improvement over the Same Cpu and Cpu. Traditional Face Recognition Algorithms Include Face Recognition Technology Based on Pca(Principal Components Analysis) and Face Location Technology Based on Ad Boost. Although the Traditional Face Recognition Technology is Fast, the Detection Effect is Much Different from the Deep Learning Technology. on the One Hand, the Accuracy of the Traditional Face Recognition Methods Represented by Pca is Far Lower Than That of the Deep Learning Algorithm. on the Other Hand, for the Recognition of Massive Users, the Traditional Pca Face Recognition Technology is Not Competent.

1. Introduction

Face Recognition System Mainly Includes Four Core Links: Face Detection, Face Alignment, Face Feature Extraction and Face Recognition. the Image or Image Sequence is Acquired through the Camera to Judge Whether There is a Face in the Input Image. in the Case of a Face, the Size and Position of the Face Are Determined to Obtain the Face Image [1]. Then Face Image Detection Key Points and Face Alignment Processing; Next, Face Image is Extracted to Get Face Feature Information. Finally, the Facial Feature Vectors Are Identified and Classified to Get the Classification of the Face.

The Purpose of Face Detection is to Analyze the Input Image and Determine Whether There is a Human Face in the Input Image. If There is, the Position and Size of the Face Can Be Obtained. At Present, There Are Many Face Detection Algorithms, Including Template Matching Model, Support Vector Machine Model and Ad Boost Model, Etc. Considering the Effect and Running Speed of Face Detection, This Paper Selects the Face Detection Location Method Based on Deep Learning Algorithm.

Face Alignment Refers to the Automatic Positioning of Key Facial Feature Points, Such as Eyes, Nose Tip, Mouth Corners, Eyebrows and Contour Points of Various Parts of the Face, on the Detected Face Image. the Input is Facial Appearance Image and the Output is Facial Shape, Namely Feature
Point Set. At Present, There Are a Lot of Face Alignment Algorithms, Including Asm, Aam, Clm and a Series of Improved Algorithms [2]. However, Esr and 3d-Esr Are More Popular. Face Alignment Method is Often Used in Face Organ Location or Organ Tracking, Facial Expression Recognition, Face Caricature Image Generation and So on. In This Paper, the Position of Eyes, Nose Tip and Other Key Points in the Face Image is Located, and the Face Image is Rotated, Cropped and Aligned According to the Position and Distance of Key Points [3].

The Purpose of Feature Extraction is to Extract the Main Information of Facial Features [4]. To Be Able to Represent the Features of Human Face Well, the Following Conditions Need to Be Met: the Face is Less Affected by Light, Noise and Other External Interference Factors That Have Nothing to Do with the Face; as Many Facial Features as Possible and Relatively Complete Information; Avoid High Dimensional Data.

Matching Recognition is the Last Step of Face Recognition System. Choose the Appropriate Classification Method, and Match the Face to Be Recognized with the Known Face Samples in the Single Sample Database, So as to Get the Final Face Recognition Results [5]. Among Them, It Needs to Be Distinguished from Face Verification, Which is to Judge Whether Two Face Images Are the Same Person, While Face Recognition is Multiple Face Verification, Which is a One-to-Many Matching and Recognition Process [6].

The network structure of the convolutional neural network can be divided into three parts according to the function. The first part is feature extraction, which is the most important part of the convolution neural network [7]. It is composed of periodic alternation of convolution, nonlinear transformation and pooling. Convolution is through convolution kernels and the target image to do operation, resulting in some characteristics of the image, nonlinear transform of the convolution is the stage to get secondary processing characteristics according to certain principles, so that it can enhance the characteristics of the model expression ability, by pooling operation can be more low resolution images, its purpose is to obtain certain shift invariance, let one of the main features of the image is more outstanding, to improve the robustness of image recognition, the pooling of commonly used functions, there is a maximum average pooling nuclear pool is to use a particular area of the average or maximum value instead of its original value, after feature extraction layer, After forming the feature vector, the obtained feature graphs are connected to the final output layer through the full connection layer [8]. Among them, the whole connection layer improves the expression ability of things through the mutual combination of features [9].

The caspearl1 data set is a large-scale Chinese face image database. The data set contains a total of 30,900 images with a size of 360*480 for 1,040 people, which are the sub library of frontal images and the sub library of attitude images, respectively [10, 11].

In this paper, the pose image sub library of the data set is selected for the experimental test of multi-pose face recognition. The sub library contains a total of 21,840 images of 1,040 people. As shown in figure 1, each person has 21 attitude change images, including 3 attitude changes (head up and left and right flip up+Yaw, right and left flip Yaw, head down and right and right flip down+Yaw and 7 horizontal depth rotation attitude changes under each elevation posture, respectively, 150, 220, 300, 450, 670.

Figure 1 Cas-Peal-R1 Dataset of Different Pose Samples

With the construction and development of ping city, a strong video monitoring network system has been established all over China. In video monitoring, the identification information of pedestrians can be quickly and accurately determined, has become a hot issue in the research field. Among them,
identity recognition technology, as one of the important research directions in the fields of information security and artificial intelligence, has become a key technology to be developed. In recent decades, the safe and convenient way for identity recognition is the biometric technology, which refers to the identification technology that USES the personalized biological characteristics of the human body for identity authentication, such as face recognition, iris recognition, fingerprint recognition, palm print recognition, voice print recognition, and signature recognition. Compared with traditional authentication methods, the advantage of identity authentication based on biometric technology is the use of personalized information of the human body, which cannot be stolen and forged.

2. Ease of Use
As Alpha GO beat GO champion lee se-dol, artificial intelligence began to become a worldwide hot topic, one of deep learning is an important method to realize the artificial intelligence, but it is not the original scientists a new way, it's just an extension of traditional neural network, and also can be said to be the further development of neural network in the era of big data. At present, it has excellent performance in image recognition, natural language processing, audio processing, finance and other fields.

Depth study of the emergence and development is not smooth, has experienced a long and tortuous process, the initial neural network just USES a simple linear weighted summation to simulate the input and output of the conversion process, but the process requires people to set up the connection weights, more human factors to join and cannot achieve the best effect, in order to make neural network can be independent to learn to update weights, then appeared the perceptron model, a model of machine learning now had a great effect. However, the perceptron model can only solve the problem of linear reparable, but it cannot solve the problem of linear reparability. In the 1980s, distributed expression and back propagation algorithms emerged. The core idea of distributed expression is that knowledge and concepts in the real world should be expressed by multiple neurons, and each neuron can participate in the expression of multiple features, which greatly strengthens the model's ability of feature expression and enables the neural network to have a better performance for linear inseparable problems. The back propagation algorithm also greatly reduces the complexity of neural network training. Until now, back propagation algorithm is the main algorithm for neural network training. With the improvement of computer performance and the emergence of cloud computing and GPU, computing capacity is no longer a difficult problem in neural network research. At the same time, with the rapid development of Internet +, we can easily obtain massive data. With the solution of these problems, deep learning has begun. Subsequently, in the image classification competition held by ImageNet, the deep learning algorithm also achieved amazing results, and made breakthroughs in natural language processing, speech recognition and other fields.

Deep learning to achieve such remarkable achievements, features of the strong power of expression has a close relationship, the traditional manual extraction is based on the characteristics of a person's thoughts, and the characteristics of the selected subjectivity is too strong, so is not very good response performance of a thing, at the same time, it takes time and energy to consider how to set up characteristic, if for different tasks, and going to design features. In the process of feature extraction, deep learning does not need human participation at all, and good features can be obtained through its own learning. Studies show that the processing of information in human brain is hierarchical, and the hierarchical structure of deep learning is just corresponding to it. Compared with shallow model, deep learning method has stronger characterization and generalization ability for highly unstructured and complex distributed data such as images.

The success of the depth model is inseparable from two important basic conditions. One is the massive training and test data set, because only as much data as possible can accurately describe a model. Originally GPU was only used for graphics display. In recent years, due to its outstanding performance in parallel computing of massive data, it has become the computing hardware basis of deep learning. Currently, mainstream deep learning frameworks like Cafe and then so flow all support GPU.
Deep learning can be divided into unsupervised learning model and supervised learning model. Unsupervised deep learning models mainly include the depth confidence network based on restricted Boltzmann machine and the depth network based on automatic encoder. Supervised learning depth model includes multi-layer perceptron and deep convolutional neural network.

Neural network is a system with learning function. When we input the training sample into the network, the network can automatically learn and adjust the weight, so as to meet the minimum error between the output and the sample label. The trained model is like a function whose parameters are known, and each input will give a specific output, which is used for the prediction purpose of the model. General neural network is composed of input layer, hidden layer and output layer, each layer contains a certain number of neurons, the specific number of neurons to be set according to the specific problem to be analyzed. The structure of neurons is shown in figure 1.

The output of the neuron is as follows:

$$h_{w,b}(x) = f\left(w^Tx + b\right) = f\left(\sum_{i=1}^{3} w_i x_i + b\right)$$  \hspace{1cm} (1)

The neuron model contains input, weight, threshold, activation function and output. The input is multiplied by the corresponding weight and then summed up, and then added a threshold value, and the result is fed into the activation function to get the output. The process is similar to electrochemical changes between neurons in the brain.

Multi-layer neural network is a network structure composed of one or more hidden layers, as shown in figure 2, which is a three-layer neural network, including a hidden layer. Neural networks learn to perform specific functions. Among them, back-propagation algorithm is the most commonly used learning algorithm. The learning process is divided into forward propagation and reverse propagation.

In the 1960s, biologists Hubel and Wiesel looked at how neurons in the cat's brain responded to images projected at precise positions on a screen in front of the cat. People call it the receptive field.
Because the receptive field is only locally active, it is a good indicator of local associations between images. Influenced by this idea, Fukushima proposed what is believed to be the first CNN engineering implementation network. The basic structure of CNN consists of input layer, convolution layer, pooling layer, full connection layer and output layer.

The network structure of the convolutional neural network can be divided into three parts according to the function. The first part is feature extraction, which is the most important part of the convolutional neural network. It is composed of periodic alternation of convolution, nonlinear transformation and pooling. Convolution is through convolution kernels and the target image to do operation, resulting in some characteristics of the image, nonlinear transform of the convolution is the stage to get secondary processing characteristics according to certain principles, so that it can enhance the characteristics of the model expression ability, by pooling operation can be more low resolution images, its purpose is to obtain certain shift invariance, let one of the main features of the image is more outstanding, to improve the robustness of image recognition, the pooling of commonly used functions, there is a maximum average pooling nuclear pool is to use a particular area of the average or maximum value instead of its original value, after feature extraction layer, After forming the feature vector, the obtained feature graphs are connected to the final output layer through the full connection layer. Among them, the whole connection layer improves the expression ability of things through the mutual combination of features.

The first convolution operation is $1*1+1*0+1*0+1*0+1*0+1*0+0*1+0*0+1*1=4$. In this way, the eigenvalue of the first position in the image is obtained. Then the convolution kernel moves to the right and down one pixel in order to carry out the convolution operation. And you end up with a convolution feature of size 3 by 3 of the convolution kernel.

![Figure 4 the Picture About Convolution Operation](image)

Compared with multi-layer perceptron, the model of convolutional neural network has fewer parameters due to the feature of weight sharing in the convolutional layer, so it is easier to train and the model has stronger generalization ability. Meanwhile, pooling makes the model more robust. In the convolution layer, we extract some specific features of the image through the convolution kernel. Different convolution cores extract different features. Therefore, the number of convolution kernels is also important for feature extraction. If there are too few convolution kernels, the extracted features will be relatively small, which is not conducive to the distinction between various categories. Too many convolution kernels will make the model more complex and increase the cost of training. For a specific classification task, it is now mainly based on multiple trials to select a better result.

3. Prepare Your Paper Before Styling

Facial features are highly non-rigid and have a lot of details reflecting individual differences. Face feature extraction is to check each pixel in the face image to determine whether the pixel is representational. Feature extraction of face image is to transform face data described by pixel into advanced description of shape, texture, motion and so on.

Based on the shallow study of traditional face image is characterized by choosing completeness of low-level features of shallow model to describe, such as the Gabor features, LBP, sift features, etc., and as hardware performance has increased in recent years, deep learning to use its very high flexiblity, strong fitting, strong adaptability, gradually emerging in different areas. At present, face recognition methods based on deep learning have achieved fruitful results. In feature extraction of face
images, not only the intra-class changes of face images caused by external factors such as expression and posture, but also the inter-class changes caused by individual differences should be considered, both of which are nonlinear and highly complex. Feature extraction based on deep learning is a cognitive learning that simulates human visual perception system, which can extract facial depth features with more representational power and further improve the influence of intra-class and inter-class changes in the face.

FaceNet is Google's 2015 classic network for face recognition under different stance is to use the same samples with high cohesion, samples of different person has the characteristics of low FGC, puts forward the convolutional neural network combined with three Yuan loss function approach, namely by using convolution neural network to extract the depth of the human face, then use three Yuan loss function will feature vector map to European space, to calculate different samples after mapping feature vector distance, for face recognition. The objective of the ternary loss function is to train and learn network parameters so that the sample spacing of the same person is always smaller than that of different people.

Face recognition methods based on FaceNet are mainly divided into the following steps.

1) Extract feature vectors from face images using FaceNet network. There are two main points about the selection of network. First, there is the problem of dimensionality selection of feature vectors. The larger the dimension is, the easier it is to distinguish different images. However, too large a training model is not easy to converge, and the calculation is slow during the test, which takes up a lot of space. Experiments show that 128-dimensional features can balance this problem well. Second, the selection of CNN model, the model of high precision is often more parameters, calculation. Smaller, less accurate models are best for mobile devices. High-precision, high-computational models can be used on the server.

2) Use the ternary loss function. Some scholars adopt the binary loss function. The objective of the binary loss function is to map the facial features of the same individual to the same point in space, while the objective of the ternary loss function is to map the facial features of the same individual to the same region, and make the distance within the class less than the distance between the classes. Triplet loss function, the triplet is a triple, forms from the training data set randomly choose a sample, the sample are recorded as x - a (Anchor), and then randomly selected from a sample and x - a belong to the same class and different kinds of samples, corresponding to the two sample are recorded as x - p,) and x - n (Negative), and thus form a (a, x - x - p, x - n) triples, as shown in figure 4. The purpose of triples loss function is to obtain a set of network parameters through learning and training, so as to make the feature spacing of homogeneous samples as small as possible, and at the same time make the feature spacing of heterogeneous samples as large as possible. In order to ensure the convergence speed of training, the triad data is trained by selecting the farthest homogeneous face image and the nearest heterogeneous face image.

3) Identification and verification. By calculating the distance between feature vectors after the face image is mapped to the Euclidean space and setting the threshold, the face classification results are obtained.

![Figure 5 Illustration of Triplet Loss Function](image)

In the practical application of deep learning based face model, on the one hand, it requires a very large amount of data to train the learning network for the good robustness of the network model, and it also requires the powerful computing ability of the computer, which hinders the extensive application of the deep learning network. On the other hand, for specific face recognition tasks, the training
sample size is very limited, especially for face recognition based on single sample, so it is impossible to train an effective face model based on deep learning. Therefore, a pre-trained deep learning-based face model is transferred to a specific face recognition task by using deep transfer learning method to solve the face recognition problem based on single sample.

Transfer learning refers to the transfer of knowledge learned from previous training to new target tasks to make use of, which breaks through the traditional requirement that training data and test data must be distributed together in machine learning. For different data in the same field, when the data are relatively similar, although there are differences between the data, the knowledge learned from previous training can also be transferred to the target task to make use of. Abstract, similar tasks in training study can represent the characteristics of the image, will learn some common characteristics, therefore, can be found using the source domain of characteristics, its migration to the target task, by using the prior knowledge of the source domain, can be in a very small amount of sample data for the training model, and study this pre training model in the target domain can be used as the feature extraction mechanism. Formally, deep transfer learning is defined as, when given a source domain $D_S$ and source task $T_S$, and a target domain $D_T$ and target task $T_T$, transfer learning is to apply the knowledge learned in $D_S$ and $T_S$ to task $T_T$. For face recognition, $D_S=DT$ and $T_S$ does not equal $T_T$(although it is closely related). Specifically, TS has been learned from a very large data set training, but now it is necessary to learn TT in a specific target task (a small number of data sets, $n_S >> n_T$), where the knowledge learned in TS helps to learn effective $T_T$.

Deep learning network migration learning is generally divided into two categories. First, for the new face recognition task, the pre-training model of convolutional network trained in the source domain (a very large amount of data) is directly used as the feature extraction machine to extract the features of the new face recognition task. Second, for the pre-training model of convolutional network, the network parameters are fine-tuned in the new face recognition task to achieve face recognition of the target task.

In face recognition, the accuracy of face recognition is very sensitive to the pose. Because this article is aimed at problem of face recognition in video surveillance, face posture change often exceed 30°, and known face samples containing only a single front view, the lack of information. Therefore, considering the smooth and gradual attitude change of the same person in video monitoring, and the robustness of face image based on VGGNet to the situation of small attitude change, this paper proposes the method of multi-confidence statistical criterion to solve the problem of multi-attitude face recognition under the single sample.

Algorithm is the core idea, first of all could be divided into different gesture grade of unknown sample library collection (15°, 30°, 45°, 60°). And other bigger posture change, then use the single sample library to identify sample library 15° posture collection for identification of forecast, and the sample after identification prediction for seed samples, extended to the single sample library, the seed samples with high similarity between adjacent posture level samples and smaller Angle difference, for example, 30° posture samples relative to the single view sample library Angle difference is 30°, but relative to 15° new seed samples of Angle difference is only 15°, and then reuse these seeds sample set for identification of 30° posture prediction, calculation of unknown samples of confidence in its collection. The high confidence samples are extended to the seed sample set. Similarly in 45°, 60° posture collection will be high confidence level samples is extended in seed sample set, and using small Angle difference between adjacent level posture collection and characteristics of the high similarity to the gradual and progressive expansion of seed samples set; Finally, the obtained seed sample set was used to construct the classification model, and the remaining non-seed samples and
other samples with larger attitude changes in the attitude set were identified and predicted to complete face recognition.

For the face recognition problem in the case of large attitude change, it is no longer face recognition based on the single perspective sample in the single sample target set, but the seed sample set after the gradual and progressive recognition and prediction of the high similarity features between adjacent attitude samples is used to identify and predict other larger attitude change sample sets. Therefore, the face recognition problem based on single sample is transformed into the multi-pose face recognition problem in the case of multiple books. In this paper, the seed sample set was obtained by constructing the multi-confidence statistical criteria, and the seed sample was used to construct the classification model to identify the samples with large attitude changes. The seed sample set obtained in this paper not only increases the number of training samples, but also covers a rich variety of postures, so as to solve the impact of multiple postures on face recognition accuracy in a single perspective.

Face pose estimation is to approximately calculate the deflection Angle of a two-dimensional face image corresponding to the three-dimensional rectangular coordinate system XYZ in a two-dimensional plane. To be specific, the change of face pose refers to the combination of the deflection Angle of the face image in the three-dimensional rectangular coordinate system XYZ direction, so the face pose estimation process mainly estimates the deflection Angle a, b and c of the two-dimensional face image in the three directions, as shown in figure 6.

![Figure 6: Face Pose Description Illustration](image)

In terms of face key point’s extraction, it is able to automatically find the location of the eyes, nose, mouth, face contour and other iconic features on the face. In this paper, we only focus on the shape information of the face, not the contour information of the face. Taking human face key point location accuracy and efficiency into consideration, this paper selects SDM (Supervised Descent Method) facial feature point detection Method.

SDM method is a regression-based method. The basic principle is to give an initial shape for a given face image. Through continuous iteration, the initial shape is returned to the position close to or even equal to the real shape. SDM is an optimization method based on machine learning to minimize the nonlinear least squares objective function. By learning a series of descending directions and the scales in this direction, the objective function can converge to the minimum value at a very fast speed, which avoids the problem of solving Jacobean matrix and Hessian matrix. Compared with gradient descent method or Newton method, SDM method greatly improves the running speed and simplifies the complexity of the algorithm.

4. Using the Template
As one of the more convenient and safe directions in biometric technology, face recognition is not only widely used in real life, but also plays an important role in the development of auxiliary disciplines such as pattern recognition and artificial intelligence. In this paper, the face recognition system is mainly the face feature extraction module. With the application of deep learning method, the recognition accuracy of face recognition technology based on deep learning has been improved. At present, face recognition technology has been relatively mature and widely used in intelligent security, Internet finance and other fields. Face recognition operation is simple, and the mobile phone supporting face recognition function is more and more popular, face recognition technology can be effectively applied in the field of remote education.

How to describe a face image is a problem to be solved in feature extraction. In this paper, the deep learning feature extraction method based on VGGNet is adopted to extract the features of the whole face image, which can obtain the depth features with more feature representation power and further improve the influence of intra-class and inter-class changes in the face. In addition, the feature extraction method based on VGGNet is compared with other classical facial recognition network features in an experiment to reflect the advantages of VGGNet in representing facial features. Finally, use the VGGNet characterization of facial features were in posture, facial expression is carried out on the face data set experimental test, the results show that based on the network characteristics of VGGNet face image can be more than a good characterization of expression, as well as facial profile on more gesture data sets turn water depth (Yaw) and 300° recognition effect is very good, but for larger sample recognition effect is to improve posture change.

References

[1] Pan J , Wang X S , Cheng Y H . Single-Sample Face Recognition Based on LPP Feature Transfer[J]. IEEE Access, 2016, 4, pp. 2873-2884.
[2] Wu, Shuai. Quaternion-based improved LPP method for color face recognition[J]. Optik - International Journal for Light and Electron Optics, 2014, 125(10), pp. 2344-2349.
[3] Lu J , Tan Y P , Wang G . Discriminative Multimanifold Analysis for Face Recognition from a Single Training Sample per Person[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1):39-51.
[4] Wen Y , Zhang K , Li Z , et al. A Discriminative Feature Learning Approach for Deep Face Recognition[M]// Computer Vision – ECCV 2016. Springer International Publishing, 2016.
[5] Jin Y , Lu J , Ruan Q , et al. Complete discriminative feature learning: A new approach for heterogeneous face recognition[C]// 2014 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2014.
[6] Jin Y , Lu J , Ruan Q . Coupled Discriminative Feature Learning for Heterogeneous Face Recognition[J]. IEEE Transactions on Information Forensics and Security, 2017, 10(3):640-652.
[7] Seo J J , Kim H I , Ro Y M . Pose-Robust and Discriminative Feature Representation by Multi-task Deep Learning for Multi-view Face Recognition[C]// IEEE International Symposium on Multimedia. IEEE, 2016.
[8] Choi Y , Kim H , Ro Y M . Two-step Learning of Deep Convolutional Neural Network for Discriminative Face Recognition under Varying Illumination[J]. Electronic Imaging, 2016, 2016(11):1-5.
[9] Koc M , Barkana A . Discriminative common vector approach based feature selection in face recognition[J]. Computers & Electrical Engineering, 2014, 40(8):37-50.
[10] Wu B , Wu H , Zhang M M Y . Scalable Angular Discriminative Deep Metric Learning for Face Recognition[J]. 2018.
[11] Sungatullina D , Lu J , Wang G , et al. Multiview discriminative learning for age-invariant face recognition[C]// IEEE International Conference & Workshops on Automatic Face & Gesture Recognition. IEEE, 2013.