

Face Recognition System Based on Capsule Networks

JiangRong Shi

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 3097078251@qq.com

Li Zhao

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 332099732@qq.com

Abstract—This study introduces a technique for facial recognition according to capsule networks. The system utilizes the advantages of capsule networks to model the face features in the image hierarchically, and realizes the efficient recognition of faces. First of all, we know the difference between the capsule network and the convolutional neural network through the study of the operating principle and the structure of the capsule network. Secondly, the Capsule Network is realized through deep research on the algorithm for dynamic routing and the internal operating principle of the capsule. Finally, by conducting experiments on the face dataset and optimizing it with the Adam optimization algorithm as well as the boundary loss and reconstruction loss, the capsule network is promoted to learn more robust feature representations to obtain better face recognition results. The experiments show that the face recognition system based on capsule network can reach 93.5% correct rate of evaluation on WebFace dataset, which achieves a high recognition accuracy. The final results demonstrate the feasibility and effectiveness of capsule networks for face recognition.

Keywords—Capsule Neural Network; Dynamic Routing; Face Recognition

I. INTRODUCTION

In the last few years, face recognition technology has become the most dominant biometric technology today, and its direct, convenient, and contactless features make it easy for users to accept, and it has been extensively used in numerous fields. Deep convolutional neural networks have demonstrated significant potential in a number of fields recently, including image identification. Deep neural networks have, however, been unable to consider the spatial relations of the underlying objects in recent years.

In order to overcome the limitations of deep convolutional neural networks, the capsule network has been widely used in the past. It takes the output of the capsule in the form of vectors, which can not only represent the image according to the intensity of the vectors, i.e., the size, but also describe the direction of the image with vector direction, location, and other information about the image object. Therefore, capsule networks are rapidly developing in deep learning with their unique charm. It is an important topic in image recognition at present and for a long time to come, and many scholars are actively working in this area. The success of capsule networks has many good outcomes in the domain of picture identification, and the technique for recognizing faces based on capsule networks involves a wide range of topics, which is crucial for research given the quick development of IT, the ongoing evolution of society, and the demand for complete recognition [2].

II. RELATED WORKS

A. Face Recognition

One type of biometric technology is facial recognition that recognizes information about the facial features of a person's face as well as determining the likelihood that it is a particular person. Researchers began to study facial recognition as early as the 1960s, and it was not until the 1990s and a half that the technology entered the stage of real primary application, and the technology is so mature that it hasn't been developed yet. Face recognition has been extensively researched as a type of non-contact

biometric identification method along with the advancements in computer vision and pattern recognition. Examples include authentication, monitoring and security detection. It can be quite convenient and automated. The majority of facial recognition techniques rely on conventional strategies for machine learning such as Principal Component Analysis (PCA) and Linear Discrimination (LDA). With the advent of deep neural networks (CNNs), facial recognition technology has advanced significantly in recent years. Deep learning has demonstrated in recent years that high level abstract features may be extracted from the source image with better robustness and accuracy [3]. Face detection and face feature extraction are the two primary components of most deep learning-based face recognition systems. Face features are extracted to create discriminative features, and face detection is used to find and extract these features. Neural networks with convolutions have been the subject of several proposals in recent years for feature extraction from the feature space. Face recognition is also connected to other technologies including 3D face reconstruction, live detection, and multimodal fusion in order to enhance the system's robustness and performance. However, there are still a number of challenges in the field of face recognition, such as changing lighting conditions, changing pose, blocking, and privacy and security. Traditional machine learning methods have seen a dramatic shift within the domain of face recognition in recent years. The performance and application of the face recognition system will be progressively improved along with the ongoing development and advancement of technology, offering society more practical and secure solutions.

B. Convolutional Neural Network

Convolutional neural networks, or CNNs, were first proposed in the 1980s and early 1990s. Yann LeCun is credited with one of the first successful uses of CNNs [16]. LeNet-5 was really utilized by the United States Postal Service and was initially intended for handwritten digit recognition. This model is a major breakthrough

in the application of convolutional neural networks in computer vision. Since the advent of deep learning, Alex Krizhevsky and his colleagues have been able to improve their performance in the 2012 ImageNet Challenge. It introduced innovative designs such as deep structure, massively parallel computation, and ReLU activation function, which has been extensively utilized in the field of computer vision and picture classification [17].

Three layers make up deep convolutional neural networks: the convolutional layer, pooling layer, and fully connected layer. The pooling layer is used to minimize the size of the parameters, while the convolutional layer is used to extract the image's local features. The foundation of convolutional neural networks is a set of filters that can identify characteristics in the input data. The buried topological characteristics are then extracted using pooling and convolution. Convolutional neural networks have been offered as a way to increase system performance and simplify the network's parameters through pooling and weighting procedures.

Deep neural networks have had tremendous success recently in a variety of fields, such as image identification and computer vision. Among the many computer vision tasks that convolutional neural networks have excelled at are picture categorization.

The disadvantages of convolutional neural networks are mainly reflected in two aspects. In the first place, in convolutional neural networks, no attention is paid to the relative positions of different features. Scalar transfers higher level neurons to lower level neurons, but it lacks direction and is unable to convey spatial qualities or the relationship between top and bottom attributes in terms of position. Connection between the objects at the base [1]. Therefore, CNNs have significant limitations in the recognition of spatial relationships. On the other hand, in Convolutional Neural Networks, there is a significant improvement in their robustness, while much valuable information of the model is lost. Convolutional neural network is trained well

when it encounters very similar images in the data set, but when it encounters an image that has flipped, tilted, and other problems related to orientation, whereas convolutional neural network does not work so well. Consider a face, for instance, which is made up of facial contours, eyes, nose, and mouth. When these components are present in convolutional neural networks, there is a strong stimulus, and the components' relative positions and orientations are less significant, so that they can recognize faces from those areas. Except it's not a face for us.

C. Capsule neural network

In 2017, Hinton et al. first proposed the concept of capsule networks, which is one of the most cutting-edge techniques in the domain of recognition and classification of images today [12]. Capsule networks are proposed to solve some problems existing in traditional convolutional neural networks, such as translation invariance, insensitivity to scale changes and pose changes. Capsule networks solve these problems by introducing capsule layers. A Feature Map (FM) is produced in traditional CNNs by adding a filter to a convolution process. However, this method is not able to handle spatial relations and pose information efficiently. In contrast, Capsule Networks can better capture the relationship between target instances by representing each target instance as a vector.

Capsule networks have several advantages over conventional CNNs:

- **Modeling Spatial Relationships:** Capsule Networks can capture spatial relationships between target instances, providing better modeling capabilities for target pose, scale change, and rotation.
- **Stronger representation:** By using vectors to represent the states of target instances, capsule networks can provide richer and more expressive feature representations, thus improving recognition accuracy.
- **Improved robustness:** Capsule networks have better robustness to translation invariance and spatial transformations when

processing images, and can cope with various changes in complex scenes.

- **Strong interpretability:** Since the capsule networks represent the state of the target instance with vectors, it is easier for the network to interpret its learned knowledge.

Unlike convolutional neural networks, which are more mature in various fields, the research of capsule networks is still in its infancy, and most of the capsule neural network research stays on the basis of small samples. In 2018, Deng et al. proposed a method to classify hyperspectral images with a small amount of sample data using capsule neural networks [18]. They introduced a new two-layer restricted training paradigm in HSI classification. Two HSI datasets were mainly used in the implementation and the algorithm was mainly used to describe complex and concise data and the algorithm was trained to examine the robustness and representation of individual models or classifiers. Sabour et al. suggested a novel feature network that comprises of a single layer of convolutional neural networks as a preprocessing layer in order to get around the drawbacks of convolutional neural networks when processing images and a layer of advanced capsule as a prediction vector for image classification [12]. Capsule Neural Network is able to recognize all kinds of features, including posture, size, and orientation and so on, but its dynamic routing mechanism is very cumbersome and needs further improvement. To address this problem, Hinton et al. improved the routing algorithm and the capsule structure, and based on this, proposed a capsule network based on the maximum expectation algorithm matrix. This network has better robustness, but large computation and high complexity. Hahn et al. used a simple perceptron model instead of the traditional dynamic algorithm, which improves the system's performance without increasing the parameters and arithmetic cost. Zhang, Y et al. investigated the performance of capsule networks in processing complex data and proposed an improved capsule network structure [13]. Xiang, S et al. introduced a method called Dynamic Capsule Attention (DCA) which was applied to a visual question and answer task [14]. Tang, H et al. proposed a Recurrent Capsule

Network (RCN) for character re-recognition task [15].

Jiang Hong et al used a convolutional layer in front of the first layer of the capsule of CapsNet, and a filter capsule layer at the end of the network. Compared with Capsnet, this method improves the recognition precision of target image and improves the performance of reconstruction [5]. Zhou Qun improved the capsule network and designed a new capsule network JSSA-Caps Net based on spatial-spectral attention module, and used a capsule network to classify hyperspectral remote sensing images, extracting useful characteristics from the combination of spatial and spectral data to enhance classification performance [7]. Yao Yuqian proposed an algorithm for recognizing expressions based on Enhanced Capsule Network (E-CapsNet) and a recognition of an expression algorithm based on Double Enhanced Capsule Network (E2-CapsNet). It was effectively validated on the expression dataset [8]. Hanqing Zhang et al. designed an algorithm for feature extraction and recognition based on Caps-net + SRNN to be able to overcome the inability of traditional CNNs to deal well with image rotation and blurring due to their information loss in the layer of pooling, and experimentally verified the effectiveness of the neural network model suggested in this document [10]. Chen Shan et al. constructed a capsule graph through dot product attention to obtain the dependency relationship between capsules in the same layer. DPA_Caps Graph not only makes up for the lack of ignoring the sibling features in the original routing process, but also achieves to enhance the model's overall performance by adding jump connections for feature extraction in the feature extraction part, which improves the feature expression of the primary capsules by using dot-product attention instead of dynamic routing, which improves the feature selection ability between the capsule layers, and realizes the improvement of the overall performance of the model [11]. Lou Yue made the first attempt to introduce capsule networks and their improved models into the field of plant recognition for applications including plant organ recognition such as flowers and leaves, preserving detailed pose information (e.g., exact position, rotation,

thickness, inclination, size of the object, etc.), and achieving generalization using less training data [9]. Yang proposed a cross-domain pedestrian re-recognition method based on deep capsule networks. Through the perspective classification training task, the model can learn the effective features of pedestrians in the image, and these features can be directly migrated to the pedestrian re-recognition task, which alleviates the problem of insufficient pedestrian re-recognition generalization capability [4]. SA-Capsnet gives full play to the feature extraction capability of self-injecting networks as well as the capsule-based neural network's dynamic routing mechanism. Dynamic routing is an attention mechanism, which has greater superiority in image regions. Combining the two can achieve complementary functions and improve the network performance. Tests have been carried out with several samples such as MNIST, Mode MNIST, CIFAR10, etc., and the outcomes demonstrate the model's high prediction accuracy [6].

III. REQUIREMENT ANALYSIS OF FACE RECOGNITION BASED ON CAPSULE NETWORKS

A. System Requirements Analysis

The system's objectives are to investigate and analyze the capsule network's theory, comprehend its elements, put the network into practice, and accomplish classification training on face datasets, to carry out the design of visualization interface and to test the working efficiency of capsule network in face recognition system. The visual interface is designed with PYQT5, and various operations are carried out by buttons, such as face data set, facial recognition and facial image addition.

B. System Main Functions

This work develops and implements the Capsule Network Face Recognition System, which is based on the theory and architecture of the capsule neural network. The system mainly contains three functions, and the description of each function is specified as follows:

- Train the face dataset. Select the face data to be trained and train the classification of face data.
- Add a face dataset. Choose which category to add to after selecting the face photos you want to add to the file, and then click the Add button. After the addition is completed, you will be prompted to retrain the added face dataset.
- Face Recognition. Select a photo of the face you like to test, display it on the screen, test it by clicking the Detect button, and the person in the picture and their likelihood of being that person will be displayed.

IV. CAPSULE NEURAL NETWORK

Capsule networks were first discovered in Geoffrey Hinton's academic paper, Transforming Autoencoders. An article titled "Dynamic Routing Between Capsules" was released at the end of 2017 by Geoffrey Hinton and his colleagues [12]. This is a novel neural network model. These days, picture recognition applications are the primary use for this technique.

A component in the brain known as a "capsule" is capable of processing various visual stimuli and encoding information (e.g., position, shape, speed, etc.) very well. In deep learning, a capsule is a structure in the brain that can process different visual stimuli well are structures in the brain that are able to process different visual stimuli well and encode information. In deep learning, a capsule is a set of embedded neurons. Instead of neurons, a network of capsules is comprised of capsules. A capsule is able to represent various features of a particular object in a picture, such as position, size, orientation, texture, etc. A capsule is an independent logical unit. A capsule can produce vectors, the direction of the vector indicates the object's pose, and the length of the vector can represent the degree of similarity of the feature.

A. Capsule network Structure

The encoder in capsule network includes convolutional layer, main capsule layer, and digital capsule layer. The specific process is as follows. The convolution layer is used to detect

the features of the 2D image. The 2D image's features are detected using the convolution layer. ReLU is used to activate the convolutional layer, which has $256 \times 9 \times 9 \times 1$ step size 1 cores. The main capsule layer, which receives the data from the convolutional layer, generates a set of features. The 32 primary capsules in this layer resemble those in the convolutional layer. Digit Capsule Layer This layer contains 10 digit capsules, each corresponding to a digit. Each capsule accepts a $6 \times 6 \times 8 \times 32$ tensor as input. You could view this as an 8-dimensional vector $6 \times 6 \times 32$, or 1,152 input vectors. Inside the capsule, each input vector maps the 8-dimensional input space to the 16-dimensional capsule output space via an 8×16 weight matrix.

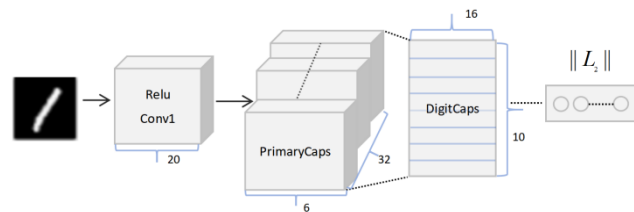


Figure 1. Capsule Network Structure

There are three connecting layers that make up the decoder: the first, second, and third. First, we accept 16 dimensional vectors from the correct digital capsule and learn to translate them into a digital image, Using the loss function — the Euclidean distance between the rebuilt image and the input image—the decoder, a regularizer, learns to reconstruct the 28×28 pixel image after receiving the output of the correct digital capsule. The decoder forces the capsule to learn features that are useful for reconstructing the output image. The ideal reconstruction of an image is one that closely resembles the original. Experiments indicate that the FNN is robust. Through the analysis of the model, we can find out the problem of the model. The rebuilt image goes through three totally connected layers, and the entire decoding process is entirely connected. The rebuilt image goes through three totally connected layers, and the entire decoding process is entirely connected.

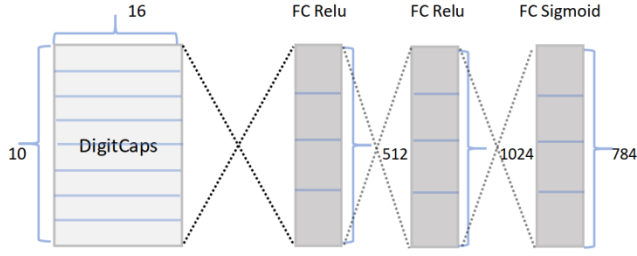


Figure 2. Decoder

B. Dynamic routing

In Capsule Network (CN), dynamic routing mechanism is an important technique used to calculate the weights between different capsules to determine their relationship and the degree of interaction. In simple terms, dynamic routing is used to determine how information is passed from one capsule to other capsules in the next layer. Through the dynamic routing mechanism, the capsule network can establish relationships and interactions between objects at different levels and enhances the model's robustness and accuracy by more effectively capturing the structural details and attitude changes of the objects in the recognition task. The following are the specific steps of dynamic routing.

1) Initialization: First, initial weights are assigned to each pair of neighboring capsules (corresponding capsules between the previous and next layers). These weights can be initialized randomly.

2) Projected Output Vector: Using the weight matrix and current state vector as a basis, an output vector is projected for every capsule. The entities or features that the capsule activates are represented by the output vector. Similarity between vectors. By calculating the dot product of two vectors, this can be accomplished.

3) Route Matching: Using the output vector of the current prediction as input, compute the output with the latter layer of capsules.

4) Update weights: The weights connecting each capsule are adjusted based on how similar they are to the output vectors of the capsules in the layer below. Capsules with higher similarity will get higher weights to have greater influence.

5) Dynamic Routing Generation Selection: The process from step 2 to step 4 is repeated until

the specified number of iterations Lou is reached or the convergence condition is satisfied. The weights of the capsule and the anticipated output vector are modified in each cycle.

6) Output computation: Ultimately, the output vector obtained after dynamic routing is used as input to class or perform other tasks. To get the final classification result, the output vector can be transformed into a probability distribution using the softmax function.

In a capsule network, the following steps are required to perform the operation of a single capsule:

The input vectors are multiplied by a matrix, where v_1 and v_2 are generated from the output of the previous capsule, and in a capsule, W_1 and W_2 are multiplied by v_1 and v_2 , respectively, to obtain new u_1 and u_2 . The formula is as follows:

$$\hat{u}_{j|i} = W_{ij}v_i \quad (1)$$

The input vectors are weighted scalarly, and the weighted vectors are summed.

$$s_j = \sum_i c_{ij} \hat{u}_{j|i} \quad (2)$$

Vector-to-vector no linearization, the result is found with the Squash function and used as input for the next capsule. In the Squash function, v_j is the output vector of capsule j , and s_j is the input vector of capsule j . s_j is also the weighted sum of the output vectors of all capsules in the previous layer, and its value is the capsule j that is currently in. This nonlinear function can be divided into two aspects. The first part represents the scale of the input vector s_j , and the second part represents the direction of the input vector, which is also compressed to the interval $[0,1)$. If the s_j vector is 0, then v_j is 0. If s_j is infinity, then v_j can tend to 1. Generally speaking, the Squash function can be utilized to excite a vector

with a vector or as a means of compressing and redistributing the vector length.

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (3)$$

Through an iterative process of dynamic routing, the capsule network can gradually adjust the weights between capsules to better transfer information and facilitate effective feature learning and pose estimation. This mechanism enables the capsule network to cope with complex spatial relationships and improves the ability to recognize object deformation, rotation, and other situations.

C. Loss function

We already know that the length of the digital capsule layer's output vector is some kind of probability based on the preceding section, how should we construct a loss function and then iteratively update the whole network according to this loss function? Dynamic routing is used to update the coupling coefficients. On this basis, it does not need to be updated according to the loss function, but the other convolutional parameters in the ensemble by value must be updated according to the loss function. The loss value for each capsule vector in a training sample is determined using the following formula and the sum of the ten loss values yields the overall loss. This is a supervised learning, so each training sample is properly labeled. Normally, updating these parameters with a standard loss function backpropagated is sufficient, but the original work used edge loss, which is common in SVMs. The expression for this loss function is:

$$L_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k) \max(0, \|v_k\| - m^-)^2 \quad (4)$$

The k of the formula denotes the classified category, T_k is the classification function (1 for the presence of k and 0 for the absence of k), m^+ denotes the upper boundary, and m^- denotes the

lower boundary. Here, where the mode of v_k is equal to L_k of the vector.

V. EXPERIMENT

A. Selection of face dataset

Some of the commonly used face recognition datasets are as follows. The roughly 13,000 photos in the LFW (Labeled Faces in the Wild) dataset show a variety of real-life subjects. Every picture is tagged with a person. CelebA is a huge data set of celebrities with more than 200,000 images of celebrities. Each image is labeled with 40 attributes, such as gender, hairstyle, etc. FDDB (Face Detection Data Set and Benchmark) is a dataset dedicated to face detection and contains 5,171 images and 16,419 face annotations. WIDER Face is the most widely used data set for face detection, which consists of 32,203 training images and 40,504 test images. Among them, the training images contain 393,703 face instances. MORPH is a dataset for age evolution studies containing 55,134 images covering 13,618 different individuals. It is mainly used for face detection and eye localization tasks. These are just some of the commonly used datasets, and there are many other face recognition datasets, so you can choose the right one according to your specific needs.

For the training task of face recognition based on capsule networks, the face recognition dataset that I have chosen is the CASIA-WebFace dataset. One of the most often used data sets from the Chinese Academy of Sciences Institute for Automation (CASIA) is CASIA-WebFace. 494,414 photos with 10,575 identities make up the CASIA-WebFace dataset. These photographs feature faces in a range of expressions, stances, and lighting settings. Each person has several images in the data set, and each person has a unique and permanent identifier. In the CASIA-WebFace dataset, lighting conditions, poses, and other factors make face recognition more difficult, training difficult, and hardware demanding. Therefore, at the beginning of the training, I just selected face images of 44 people without reducing the photos of each person, and increased the dataset to 8,260 face photos of 110 people in

the later training process. Using this data as my final training dataset.

B. Building a system to realize face recognition

According to the needs of the system, three main interfaces are designed in the system, including selecting pictures for face recognition, selecting face datasets for training, and selecting photos and categories to be added to the dataset in the folder. Firstly, we choose to design the interface through PyQt5, the first interface needs to show the picture that needs to be recognized, the model can be chosen to display the identity of the person in the photo and the likelihood that they are that person, if the selected file is not a picture then it will give a prompt to re-select it. The second interface requires a folder of face datasets that can be selected as well as a training log, which shows the results of each epoch of training, and the lower level of the interface will give hints at the beginning and end of training. The third interface allows you to add a face dataset by choosing the face photographs to add, the category to add them to, clicking the add button, and after the addition is successful, receiving a prompt to retrain.

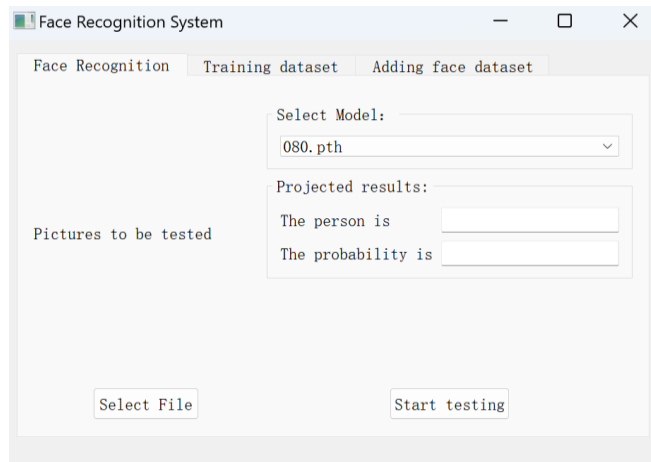


Figure 3. Face recognition

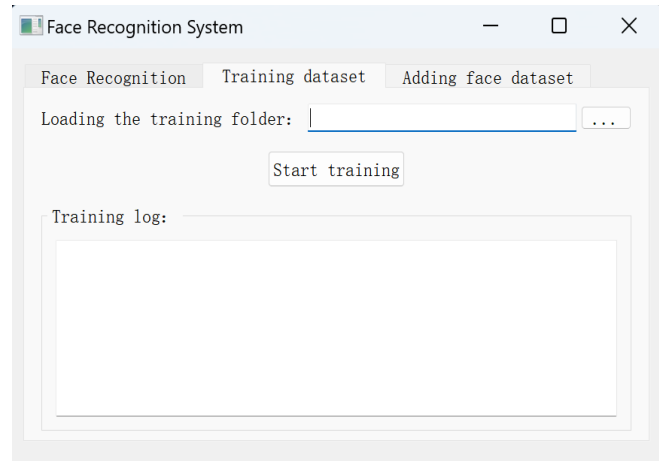


Figure 4. Training dataset

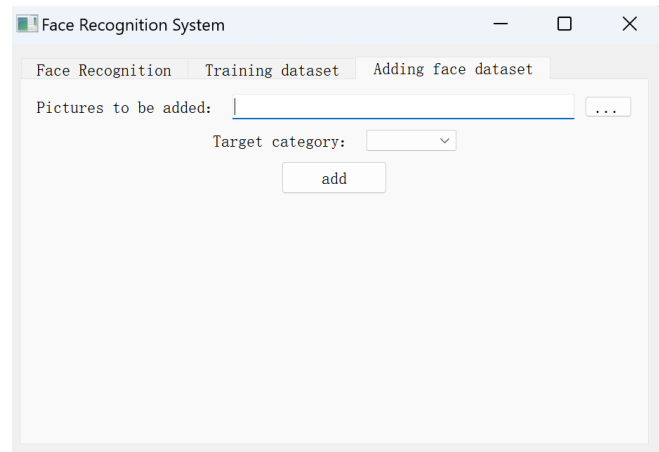


Figure 5. Adding a dataset

C. Training Capsule Network

The deep learning framework Pytorch is used, along with the Adam optimization algorithm for training and the boundary loss and reconstruction loss for optimizing the capsule network. The experimental environment is Python 3.9, the number of iterations is set to 80, the learning rate is 1e-4, and the Batch_size is set to 64. The experimental test part of the photographs can be up to 98.7% accurate, the The correct rate of model evaluation can reach 93.5%. The loss curve and correctness curve are shown below.

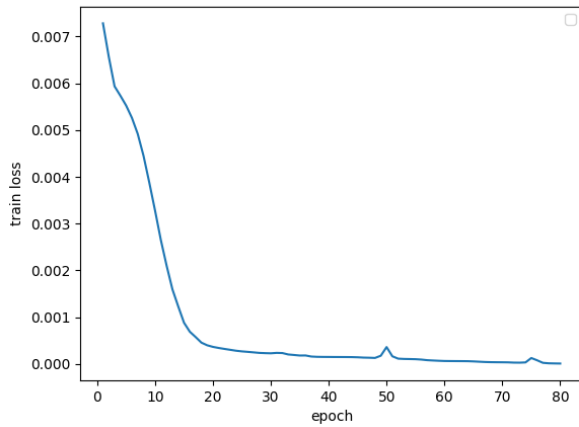


Figure 6. Loss Curve Graph

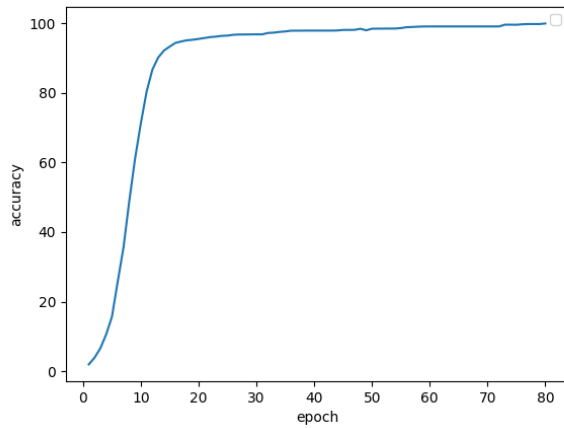


Figure 7. Accuracy Graph

Through this design implementation of capsule network based face recognition system, it is very good to understand the advantages of capsule network image recognition field for image processing. It is better at capturing the spatial relationships found in the picture. The categorization capabilities of the capsule neural network is confirmed by the face recognition results on the CASIA-WebFace face dataset.

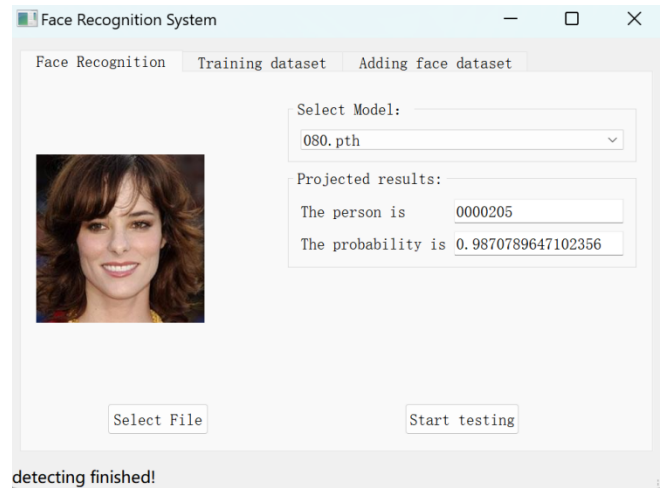


Figure 8. Face test Results

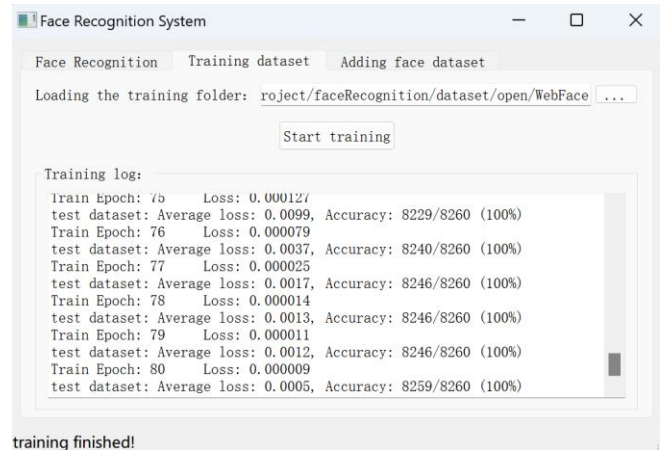


Figure 9. The dataset training process

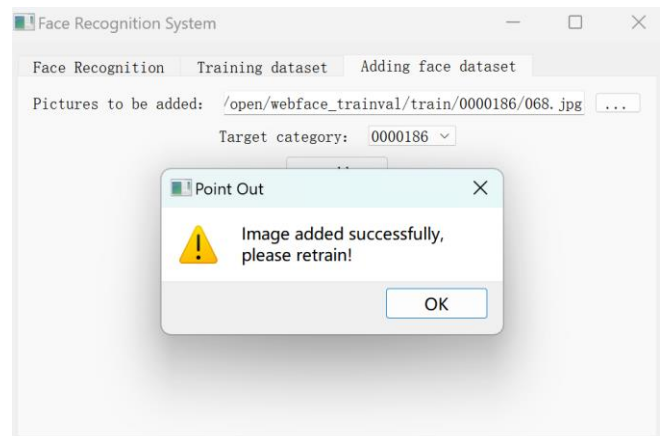


Figure 10. Add photo retraining

VI. CONCLUSIONS

The capsule network based face recognition system consists of three main functions, which are detection of a single face image, training the face dataset, and adding data to the original face dataset. However, the most important thing to realize these is the need to understand and learn the theoretical basis of capsule network, Early knowledge of convolutional neural networks gave rise to a certain view of neural networks, of which convolution operations are crucial. Convolution operations are also present in the capsule network's back. Secondly, after a more profound understanding, through in-depth study of the working principle of the capsule neural network, determine the advantages of the capsule neural network over the convolutional neural network by understanding its structural makeup, and understand the advantages of the capsule network in the realization process. Learn in-depth about dynamic routing algorithms, how they operate inside individual capsules, and how parameters are updated inside individual capsules. Finally, the development of the capsule network is not perfect, the capsule network still exists some shortcomings, the capsule network has to continue to learn.

REFERENCES

- [1] W. L. He. M.L. Zhu. Current status and future analysis of capsule neural network research [J]. Computer Engineering and Application, 2021, 57(03):33-43.
- [2] YANG Jucheng. HAN Shuiie. MAO Lei et al. A review of capsule network modeling [J]. Journal of Shandong University (Engineering Edition), 2019, 49(06):1-10.
- [3] Zheng Yuanan. Li Guangvang. Li Ye. A research review of deep learning in image recognition [J]. Computer Engineering and Applications, 2019, 55(12):20-36.
- [4] YANG Xiaofeng. ZHANG Laifu. WANG Zhineng et al. Cross-domain pedestrian re-identification based on capsule networks [J]. Computer Engineering and Science, 2021, 43(09):1591-1599.
- [5] JIANG Hong. JIA Shuaiyu. YAO Hongge. Capsule network for object recognition in complex realistic scenes [J]. Journal of Xi'an University of Technology, 2019, 39(06):712719. DOI:10.16185/j.jxatu.edu.cn.2019.06.014.
- [6] Liu Linsong. Tong Minglei. Wu Dongliang. SA-CansNet:Self-attentive capsule network[J]. Computer Application Research. 2021. 38(10):3005-3008+3039. DOI:10.19734/j.issn.10013695.2021.03.0092.
- [7] Qun Zhou. Research on hyperspectral remote sensing image classification based on capsule neural network [D]. Northern Nationalities University, 2021. DOI:10.27754/d.cnki.gbfmz.2021.000172.
- [8] Yao YO. Research on facial expression feature extraction and recognition algorithm based on capsule network [D]. Beijing Jiaotong University, 2020. DOI: 10.26944/d.cnki.gbfju.2019.000835.
- [9] Lou Yue. Research on plant recognition method based on improved capsule neural network [D]. Jilin University,2021.DOI:10.27163/d.cnki.gjlnu.2020.000142.
- [10] H.H. Zhang. Research and development of security system based on Cans-Net face recognition [D]. Xinjiang University, 2021. DOI:10.27429/d.cnki.gxjdu.2020.00355.
- [11] Shan Chen. Rencheng Sun. Fengjing Shao et al. Research and improvement of dynamic routing based on capsule networks [J]. Computer Engineering. 2022, 48(05):208214. DOI:10.19678/j.issn.1003428.0062928.
- [12] Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In Advances in Neural Information Processing Systems (pp.3856-3866).
- [13] Zhang, Y., Yang, J., & Davis, L. S. (2018). Capsule network performance on complex data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(7), 1552-1566.
- [14] Xiang, S., Wang, Y., Liu, Z., & Gilmore, J. H. (2019). Dynamic capsule attention for visual question answering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6255-6264).
- [15] Tang, H., Yu, N., Wang, R., & Wang, M. (2019). Recurrent capsule network for person re-identification. In Proceedings of the IEEE International Conference on Computer Vision (pp. 7130-7139).
- [16] Lecun Y. Bottou L. et al. Gradient-based learning applied to document recognition[J]. Proceedings of IEEE, 1998, 86(11):2278-2324.
- [17] Krizhevskv A. Sutskever I. Hinton G E. Imagenet classification with deep convolutional neural networks [C] // Advances in neural information processing systems. 2012: 1097-1105.
- [18] Deng F. Pu S. Chen X. et al. Hyperspectral image classification with capsule network using limited training samples [J]. Sensors, 2018, 18(9):22.