Research and development of autism diagnosis information system based on deep convolution neural network and facial expression data

Received 31 August 2019
Revised 16 December 2019
Accepted 23 January 2020

Wang Zhao and Long Lu
School of Information Management, Wuhan University, Wuhan, China

Abstract

Purpose – Facial expression provides abundant information for social interaction, and the analysis and utilization of facial expression data are playing a huge driving role in all areas of society. Facial expression data can reflect people's mental state. In health care, the analysis and processing of facial expression data can promote the improvement of people's health. This paper introduces several important public facial expression databases and describes the process of facial expression recognition. The standard facial expression database FER2013 and CK+ were used as the main training samples. At the same time, the facial expression image data of 16 Chinese children were collected as supplementary samples. With the help of VGG19 and Resnet18 algorithm models of deep convolution neural network, this paper studies and develops an information system for the diagnosis of autism by facial expression data.

Design/methodology/approach – The facial expression data of the training samples are based on the standard expression database FER2013 and CK+. FER2013 and CK+ databases are a common facial expression data set, which is suitable for the research of facial expression recognition. On the basis of FER2013 and CK+ facial expression database, this paper uses the machine learning model support vector machine (SVM) and deep convolution neural network model CNN, VGG19 and Resnet18 to complete the facial expression recognition.

Findings – In this study, ten normal children and ten autistic patients were recruited to test the accuracy of the information system and the diagnostic effect of autism. After testing, the accuracy rate of facial expression recognition is 81.4 percent. This information system can easily identify autistic children. The feasibility of recognizing autism through facial expression is verified.

Research limitations/implications – The CK+ facial expression database contains some adult facial expression images. In order to improve the accuracy of facial expression recognition for children, more facial expression data of children will be collected as training samples. Therefore, the recognition rate of the information system will be further improved.

Originality/value – This research uses facial expression data and the latest artificial intelligence technology, which is advanced in technology. The diagnostic accuracy of autism is higher than that of traditional systems, so this study is innovative. Research topics come from the actual needs of doctors, and the contents and methods of research have been discussed with doctors many times. The system can diagnose autism as early as possible, promote the early treatment and rehabilitation of patients, and then reduce the economic and mental burden of patients. Therefore, this information system has good social benefits and application value.

Keywords Facial expression data, FER2013, CK+, Deep convolution neural network, VGG19, Resnet18, Autism, Diagnostic information system

Paper type Research paper

1. Introduction

Facial expression recognition is an important social cognitive skill. Emotions are expressed by facial expressions. Therefore, recognition and understanding of facial expressions is the

This research has been possible thanks to the support of projects: National Natural Science Foundation of China (No. 61772375) and Independent Research Project of School of Information Management Wuhan University (No: 413100032).

0737-8831
DOI 10.1108/LHT-08-2019-0176

basis of communication and interpersonal relationships with others. Abnormal expression is a prominent manifestation of autism, and it is also one of the criteria for the diagnosis of autism. Doctors can diagnose autism by responding to abnormal facial expressions in children.

Autism, also known as autism or autism disorders, is a representative disease of generalized developmental disorders. In recent years, the incidence of autism in children has become higher and higher, experiencing a transition from rare diseases to epidemics. At present, research on autism is still in its infancy at home and abroad, and research methods and tools are still developing.

The main symptoms of autism include impaired social and interpersonal communication, language retardation, repetitive behavior and sensory dysfunction. It is difficult for autistic patients to correctly recognize faces and explain facial emotions. They have different emotional expressions from ordinary people, and they cannot correctly perceive and understand some basic expressions such as anger (Yan, 2008).

At present, the diagnostic methods for autism spectrum disorders include: traditional standard DSM-IV-TR (Segal, 2010) and ICD-10 (Organization W H, 1992), various autism diagnostic assessment scales such as "Childhood Autism Rating Scale (CARS)", "the autism child behavior scale (ABC)" and autism behavior rating scale and questionnaire interviews (Wang and Lu, 2015). Most of these methods rely on doctors' direct observation of the patient's expression, speech and behavior based on their experience. Diagnostic results are easily disturbed by external factors such as hospital level, physician's subjective level, patient's education level, age and so on. There are relatively large subjective factors, resulting in a certain degree of missed diagnosis and misdiagnosis. It takes about 1–2 h for each autistic patient to diagnose, so doctors have a lot of work to do. The best period of treatment for autistic patients is before the age of six. Early diagnosis is of great significance for the rehabilitation of autistic patients.

The purpose of our research and design is to train the model and make a facial expression recognition system based on the normal expression, so as to verify the abnormal expression. This system can test the facial expression of autistic children and judge the difference between autistic children and normal children.

In this study, FER2013 and CK+ were used as the main facial expression training samples. At the same time, we collected the facial expression image data of 16 Chinese children as a supplementary sample of facial expression. With the help of VGG19 and Resnet18 algorithm models of deep convolution neural network, according to the hospital autism diagnosis scale and diagnosis process, this paper studies and designs an information system for the diagnosis of autism by facial expression data. After the actual test of recruiting testers, the recognition rate of the system is 81.4 percent. It can effectively distinguish whether the expression of children is normal or not. It provides a practical information system for the diagnosis of autism. This paper will continue to collect more children's facial expression data from different countries and regions as training samples to further improve the recognition rate of facial expressions.

The autism diagnosis information system designed in this study has the following important significance:

(1) Autism can be diagnosed as early as possible by using this system. The best time to treat autism is before the age of six. The earlier the diagnosis of autism is made, the less the treatment cost and the higher the probability of recovery. Early diagnosis is of great value in alleviating the burden on families and society of autistic patients. The system can be published in the form of app or web pages and disseminated through the Internet. The system can be installed and used on different devices, such as computers, mobile phones, tablets, etc. It has good applicability. Through this

system, autism can be diagnosed conveniently, and time can be saved for the early treatment of autism patients, especially those in underdeveloped areas.

(2) It can make the diagnosis of autism more objective. The whole diagnosis process is completed by the system. Because artificial intelligence technology is used to recognize facial expressions without human intervention, the diagnosis results are objective and accurate.

(3) Reduce the intensity of doctors' work. Before the system was used, it took an hour for doctors to diagnose an autistic patient. By using this system, doctors can save a lot of time and pay attention to the treatment of autism.

(4) The facial expression database used in the training of this system contains different races in the world. Therefore, this system can not only diagnose children in different countries and regions but also diagnose suspected autism patients all over the world.

(5) This research designs the system according to the actual business. The early design of the system adopts the suggestions of several doctors, so it is designed and manufactured according to the actual needs of doctors. Although there are some papers on autism diagnosis by facial expressions at home and abroad, there are still few autism diagnosis systems developed which can be used in practice.

(6) This paper uses the latest in-depth learning technology to improve the accuracy of facial expression recognition. Previous traditional techniques and methods have low recognition rate of facial expressions. In recent years, with the development of artificial intelligence technology and the improvement of computing speed, the convolutional neural network has greatly improved the accuracy of facial expression recognition, which is the innovation of this research in technology.

2. Facial expression database and its recognition technology

2.1 Facial expression database

Facial expression is an important way for people to express their emotions. In the social process, facial expression is an important way to judge the attitude and inner feelings of the other party (Lanlan, 2018). Mehrabian (2008) found that in a conversation, the change of facial expression played the most important role. Of these, 55 percent are facial expressions, 38 percent are voice and only 7 percent are words (Mei and Hu, 2015). Compared with voice, expression can convey more abundant information. Recognition and understanding of facial expressions is very important for communicating with others (Shen *et al.*, 2013). In 1972, Ekman demonstrated through empirical research that human beings have six basic facial expressions: happiness, sadness, anger, fear, disgust and surprise (Ekman, 1992). In subsequent studies, neutral expression has also been added to the basic expression, and it is generally believed that there are seven basic expressions in facial expression.

With the continuous development of computer software and hardware technology, people have a deeper understanding of facial expression recognition technology. In order to better study facial expression recognition technology, many international research institutions have established standard facial expression databases, the main facial expression databases are as follows:

(1) JAFFE

The database stores facial expression data of Japanese women. It contains 213 facial images of ten Japanese women. There are seven types of facial expressions, namely neutral, happy,

sad, surprise, anger, disgust and fear. The resolution of each image is 256 × 256 pixels. Everyone has seven kinds of pictures of facial expressions.

(2) CK+

The expression database was collected under laboratory conditions. It includes African Americans, Asians and South Americans. The resolution of each image is 640*480 pixels. It contains 593 expression sequences of 123 people, 69 percent of whom are female and 31 percent are male. Each sequence begins and ends with neutral expression, which includes the process from calm to strong expression. CK+ is a facial expression data set with many applications. The reliability of various facial expression evaluation experiments using this database is very high. It includes seven types of facial expressions: anger, contempt, disgust, fear, happy, sadness and surprise.

(3) FER2013

There are 35,887 facial images in the library, and there are seven facial expression types: angry, disgust, fear, happy, sad, surprise and neutral. The resolution of each image is 48*48 pixels. All the images are gray images. There are three sample sets: 28,709 images in the training set; 3,589 images in the validation set and 3,589 images in the test set.

(4) MMI

The expression database can be divided into two parts: one is a dynamic data set composed of more than 2,900 video sequences. The other part is a static data set consisting of a large number of high resolution images. There are seven types of expression in the library.

(5) AFEW

All the facial images in the database are edited from the movies and contain seven basic facial expressions.

(6) SFEW

The expression library is a static frame image extracted from the AFEW data set, which contains seven basic expressions.

2.2 Facial expression recognition process

The process of facial expression recognition includes two stages as shown in Figure 1: One is the training stage and the other is the recognition stage. The training and recognition stages can be divided into three parts: the pretreatment of facial expression images, the extraction of facial expression features and the classification of facial expressions. The training stage is to train the model in order to achieve the purpose that the model can be used. The recognition stage is to recognize and classify the expression of the test image (Du, 2018).

The two stages of expression recognition process include the following processes: First, face detection is carried out on the image in the expression database, including the location,

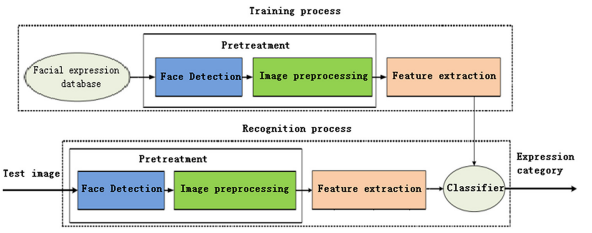

Figure 1.
Facial expression recognition process

alignment and clipping of the face area. This is the basis of the follow-up process. Only when the expression area is accurately obtained, the following series of work will be more accurate. After the face area is detected, the image needs to be preprocessed in order to eliminate the noise caused by the influence of acquisition equipment and environment and avoid the interference of feature extraction. Then it is the feature extraction step, which aims to extract the features that can represent the essence of expression from the preprocessed facial images. In this process, in order to avoid the high dimension of feature extraction and affect the efficiency of the algorithm, we need to reduce the dimension of extracted features in order to extract the most representative expression features. Finally, the extracted facial features are classified to determine which type of facial expression is.

2.3 Facial expression recognition technology

Facial expression recognition technology mainly includes traditional machine learning technology and deep learning technology. The two technologies have similarities and different characteristics.

(1) Traditional machine learning technology

Facial expression recognition algorithm based on traditional machine learning includes three steps: image preprocessing, facial expression feature extraction and feature classification.

First, for the convenience of feature extraction, it is necessary to preprocess the image, which can effectively avoid the interference of various noises and leave the key information needed by the face. The pretreatment process includes image gray processing, face alignment, face size tailoring, data enhancement, brightness, pose normalization, etc. (Li and Deng, 2018).

Second, the traditional feature extraction methods include directional gradient histogram feature, Gabor filter feature, local directional pattern feature and enhanced local binary algorithm. Because these methods are artificial design, time-consuming and laborious, and have certain limitations and often have better effect in feature extraction in small sample image set, most of the current studies are based on deep learning feature extraction method.

There are many basic machine learning methods for expression classification, such as support vector machine (SVM), hidden Markov model (HMM) and k-nearest classification algorithm.

(2) Deep learning technology

Facial expression recognition algorithm based on deep learning also needs image preprocessing. The difference is that it often combines feature extraction and feature classification into an end-to-end model, which greatly simplifies the process of facial expression recognition. In addition to end-to-end learning, deep learning algorithm can be used to extract facial expression features, and then other independent classifiers can be used. For example, SVM or random forest algorithm is used to process the extracted features and classify them.

In this paper, we construct a facial expression recognition model based on deep learning technology, extract facial expression feature data of children and classify them into groups, so as to diagnose autism.

2.4 Driving role of facial expression data

Research on facial expression recognition has been applied in a series of life scenarios. In children's education, advanced human-computer interaction, medical diagnosis and other aspects have played an important role (Cai, 2018).

In distance education or classroom teaching, teachers can better improve students' learning quality by observing students' emotional changes in the classroom and adjusting

teaching plans in time. Advanced human-computer interaction can make human-computer interaction more harmonious. For example, intelligent robots can automatically respond to the facial expressions of their interlocutors. In medical diagnosis, facial expressions also play an important role in the prevention and diagnosis of diseases. For example, this article is to diagnose autism by analyzing children's facial expressions.

3. Autism and facial expression diagnosis

3.1 Autism and its development

Autism is a neurodevelopmental disorder, which is collectively referred to as autism spectrum disorder (Duan *et al.*, 2015).

Since Kanner, an American child psychiatrist, first reported autism in 1943, the incidence of autism has risen rapidly worldwide. In the 1980s, about 3–5 out of every 10,000 people suffered from the disease, while in 2000, 6.7 out of every 1,000 children suffered from the disease (Vismara and Rogers, 2008). According to the National Center for Health Statistics, the probability of autism among children aged 3–14 in the United States reached 2.76 percent in 2016 (Zablotsky *et al.*, 2017).

There is no statistical survey on autistic children in China. However, according to the data of the report on the development of China's autism education and rehabilitation industry II, the number of people with autism in China is estimated to exceed 10 million, of which 2 million are autistic children. At the same time, it is growing at the rate of nearly 200,000 annually (Beijing Wucai Deer Autism Research Institute, 2017).

Autism brings serious financial burden to both society and family. Families with autistic children, on the one hand, spend a lot of time caring for their children, while working hours are reduced so that work income is reduced. On the other hand, the cost of family rehabilitation treatment for autistic children is huge, which increases the family's financial burden (Wu and Chen, 2018). According to the survey on the occupational and economic burden of preschool autistic children's families, 33 percent of parents of autistic children reported that their caregiving problems seriously affected their careers, and their annual income was significantly lower than that of ordinary families, with an average loss of income of 30,957 yuan per year. Meanwhile, the average annual cost of autistic children's families for children's education and training is significantly higher than that of ordinary families (Yang and Wang, 2014). The society and the government also need to invest a lot of money in the rehabilitation education of autistic children. At the same time, autism also brings high subjective load and depression to the families of patients, which has a negative impact on their quality of life (Singh *et al.*, 2017; Wang *et al.*, 2018). It can be seen that the incidence of autism in children is relatively serious, and the harm to society and family is enormous.

3.2 Diagnosis of autism through facial expressions

3.2.1 Facial expression recognition disorder. Autistic children have facial expression recognition obstacles, which are mainly manifested in their inability to recognize facial expressions (Liu *et al.*, 2015). It is easy to distinguish autistic children from normal children by observing their facial expressions. Therefore, we combine facial expression recognition technology to extract facial expression response feature vectors and use artificial intelligence technology to distinguish normal group and autistic group based on these facial features.

3.2.2 The principle of diagnosing autism through facial expressions. A large number of studies have pointed out that autistic patients have deficiencies in facial expression recognition and understanding. This is the core source of impaired social function in autistic patients (Yang *et al.*, 2017). Autistic children are more difficult to identify other people's emotional behavior, and it is difficult to make appropriate judgment and response

(Shen *et al.*, 2013). Overseas research on facial expression recognition ability of autistic patients has been carried out not only in children but also in adults. Most studies believe that the ability of facial expression recognition of autistic patients is low. Baron-Cohen *et al.* (1997) used standard facial expression maps to study the recognition of different emotional types in autistic adults. It was found that autistic adults had better recognition of some basic facial expressions, such as happiness, but relatively complex facial expressions such as surprise recognition were difficult to recognize.

At present, the main diagnostic criteria of autism are: IDC-10, DSM-IV, the autism child behavior scale (ABC), the children autism rating scale (CARS) and the Clancy behavior scale (CABS) (Wang, 2007).

After consulting a large number of literatures and investigating the actual situation of the hospital, now the hospital mainly uses CABS (filled by parents), ABC (filled by parents) and CARS (filled by doctors) to diagnose autism. After a detailed review of the test items of the three scales, these scales all contain the test items to judge autism through children's facial expressions. There were 14 items in the CABS scale, of which the seventh item was inexplicable laughter and the tenth item was not looking at each other's face. Avoiding eye contact was related to expression. There were 57 items in the ABC scale, of which the seventh item was non-communicative smile, the seventeenth item did not respond to other people's facial expressions, and the twenty-fourth item was active avoidance of eye contact with others. Fifteen items of the CARS scale, the third of which is emotional response, pleasure and unhappiness and interest, are expressed by changes in facial expression and posture. These scales basically include the items of autism detection by children's facial expressions, which show that the diagnosis of autism can be more accurate by facial expressions. With the progress of artificial intelligence technology, facial expression recognition technology can objectively and effectively reflect the mental health of children and can be used in early diagnosis of autism (Yanbin *et al.*, 2018).

We also communicated with doctors of Hubei Maternal and Child Health Hospital, Wuhan Children's Hospital and Guangzhou Women and Children's Medical Center many times, and actually checked the process of using the above autism diagnostic scale to diagnose children. The doctor observes the tester's reaction to determine whether the tester is autistic after requesting the tester to make the corresponding expression. Doctors point out that facial expression is an important part of autism diagnosis. In terms of system design, they put forward requirements and suggestions for the process of diagnosing autism through facial expression.

4. Research and development of autism diagnosis information system

4.1 Facial expression database selection

The expression databases in this study mainly come from two public expression databases CK+ and FER2013. In addition, 16 Chinese children's expression data were collected as supplementary samples. The two public expression databases are standard and international and have been widely used, including facial expression data of adults and children. Each sample in the database contains seven expressions: angry, disgust, fear, happy, sad, surprise and neutral. Because children's facial expressions are different from adults, in order to improve the recognition rate of children's facial expressions, we collected facial expression data of 16 children aged 5 to 8 in China. Seven expressions were collected from each child. We combine Chinese children's facial expression data and public expression database as our system's facial expression database.

4.1.1 FER2013 facial expression database. The reason for choosing FER2013 expression database is that it has more samples and is more mature than other expression databases. It has advantages in model training. At the same time, it has been used in many studies (see Plate 1).

4.1.2 CK+ Facial expression database. CK+ facial expression database was selected because it was collected in the laboratory, so its accuracy is relatively high (Lucey *et al.*, 2010) (see Plate 2).

4.1.3 Facial expression data of Chinese children. At present, the mature facial expression databases at home and abroad are mainly based on adult male or female facial expression images. Therefore, it is urgent to establish a facial expression database for children.

Facial images of children are quite different from those of adults. Children have rounder faces, larger eyes and less prominent bones. Because of these differences, children's facial features are less obvious and more difficult to recognize than adults. Because of the particularity of children, it is very difficult to collect children's facial images. In order to improve the recognition rate of children's facial expressions, we cooperated with Amy Education School in Zhengzhou. Sixteen healthy children as volunteers were recruited to collect facial expression data. Each of them collected seven kinds of expressions, totaling 112 pictures. These children are between 5 and 8 years old, including 8 boys and 8 girls. The acquisition environment is quiet and there is no external interference. High-definition cameras are used to collect facial expression images, which are processed professionally. Before collecting facial expression data, parents have been informed of the purpose of collecting facial expression data. After questioning with parents, all the children who participated in the collection of facial expression data had no history of autism.

We loaded the expression data into the training sample library. The purpose of collecting Chinese children's facial expression data is to increase the number of Chinese children's facial expression samples in training samples and improve the recognition rate of the system for children's facial expression. The collection process and the collected children's facial expression data are shown in Plate 3.

4.2 Network topology

According to the network environment and equipment of the information service platform, the network topology can be divided into four levels. The network topology diagram is shown in Figure 2.

The first layer is the application layer, which consists of users, computers and various smart devices. Smart devices include smart tablet computer, smartphones and other electronic devices. Users access and use the information service platform through computers and various smart devices.

Plate 1.
FER2013 facial expression database

Plate 2.
CK+ Facial expression database

The second layer is the communication layer, mainly based on the internet network environment, providing access channels for users and systems.

The third layer is the application server layer, which is composed of firewall and application server and has an ontology display system for autism. The application server manages various business functions, handles various business requests submitted by users and can access the database server for various data exchange.

The fourth layer is the database server layer, which stores all kinds of data and knowledge resources of the information service platform.

4.3 System architecture

The smart diagnosis system of autism adopts client/server architecture. The client includes different versions of programs suitable for computers and smartphones. The system architecture diagram is shown in Figure 3.

The client includes three main modules: user interaction, image acquisition and face detection. User interaction module is responsible for human-computer interaction. According to the requirements of the autism diagnostic scale, users who diagnose are required to make appropriate expressions and feedback by prompting pictures and voice guidance. Through the camera, the image acquisition module can dynamically capture facial expression images. At the right time, the system will collect facial expression images and transmit them to the face detection module. Face detection module recognizes the valid face features and compresses the image and transfers it to the server through the internet or mobile Internet.

The server includes six main modules: image processing, feature extraction, group classification, automatic diagnosis, training model and data management. The image processing module can receive the expression image transmitted by the client and then process the expression image and transfer it to other modules on the server side. The feature extraction module receives the facial expression images provided by the image processing module and extracts the facial expression features. The group classification module is

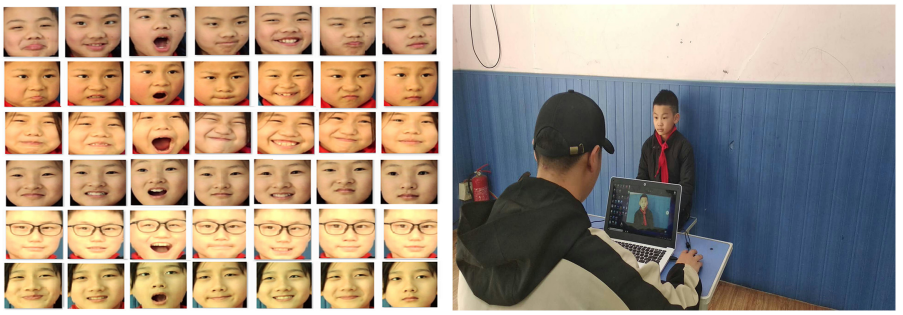

Plate 3.
Collection of facial expression data of Chinese children

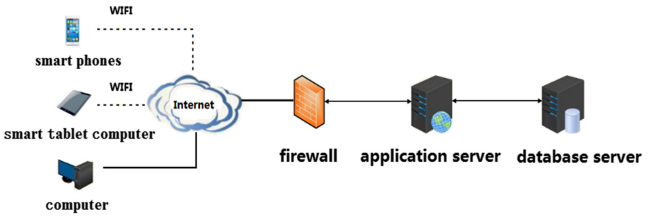

Figure 2.
Network topology diagram

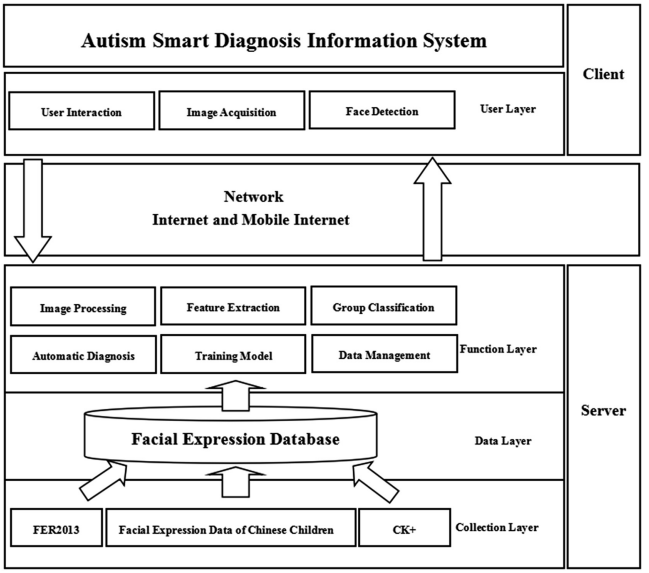

Figure 3.
System architecture

responsible for group classification and correctly classifies the expression images into the most matching expressions among the seven kinds of expressions. The automatic diagnosis module gives the diagnosis of autism by comparing the facial expressions that the tester is required to imitate and the facial expressions that the tester actually makes. Model training module is the core module of the system, which is responsible for recognizing and processing the newly collected facial expression images. The data management module mainly manages facial expression data, including storing and reading facial expression images transmitted by the client.

The system server stores facial expression feature files, which are formed by feature extraction of facial expression database. The expression feature file is HDF5 file format. The expression recognition system running on the server can read the expression feature file at any time. If new facial expression samples are collected, the model can be retrained and the facial expression feature files can be updated.

The client collects the tester's facial expression data by high-definition camera and transmits the facial expression data to the server by JSON file according to TCP communication protocol. The facial expression recognition system running on the server processes the collected facial expression data and then feeds the recognition results back to the tester through the network and stores the recognition results and facial expression data in the server database. Facial data and diagnostic system are stored on a server, and the recognition results and facial data are stored in the SQL Server database. The diagnostic system reads data from the database through SQL structured query language. The response time of the whole database operation and communication process should not exceed 5 s.

4.4 System architecture

4.4.1 VGG19 model. Researchers from the Oxford University and the Google Brain have jointly developed the convolutional neural network VGG. VGNet consists of 11, 13, 16 and 19 layers of neural networks [20].VGNet constructs 16–19 layers of neural networks by stacking small convolution cores of 3×3 and maximum pooling layers of 2×2 repeatedly. VGGNet

has strong scalability and greatly reduces the error rate when extending. When migrating to other image data, it has good generalization ability and simple structure.

4.4.2 ResNet18 model. ResNet was proposed by Kaiming He and others of Microsoft Research institute. They have successfully trained 152 layers of neural networks by using ResNet unit. The structure of ResNet can accelerate the training of the neural network, and the accuracy of the model has been greatly improved.

4.4.3 Graphic of deep learning framework. The deep learning framework used in this paper for facial expression recognition is shown in Plate 4.

The whole process includes image input, image preprocessing, model building, model training, model testing and output of expression recognition results. There are two kinds of deep learning algorithms used in this paper: VGG19 and ResNet18. ResNet18 solves the problem of network performance degradation caused by the high depth of VGG19. By training the two models and synthesizing the two convolutional neural network models, the facial expression features of autistic children can be extracted accurately.

4.4.4 Image preprocessing. The purpose of image preprocessing is to achieve uniform normalization of the final input image. The process is shown in Figure 4.

Converting an image to a grayscale image can reduce the computational complexity of the latter pixel level and also reflect the overall and local distribution and characteristics of the image. Then, image transformation is used to enhance data by zooming, rotating, cutting and translating, and the image is located in the center of the window. The contrast and brightness of the image can be improved by histogram equalization to reduce the influence of illumination on expression feature learning. In order to make the image uniform, it is planned to transform the image size into the same size by normalizing the image size. Finally, the mask is used to remove the occlusion of non-face areas.

4.4.5 Model training. Before model training, we need to enhance the image data. We choose SGD random gradient descent algorithm as the optimization method. The batch size is still 128 by default, and the learning rate is set to 0.01 initially. In addition, the initialization of

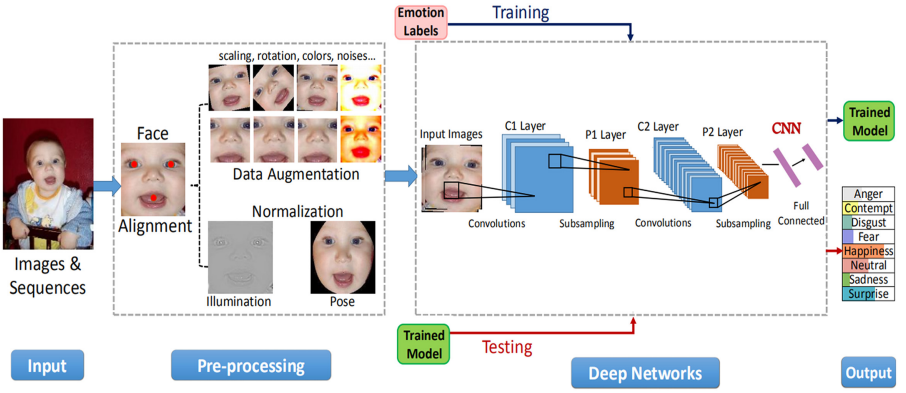

Plate 4.
Framework of deep learning

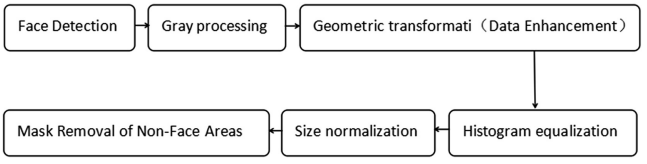

Figure 4.
Image preprocessing

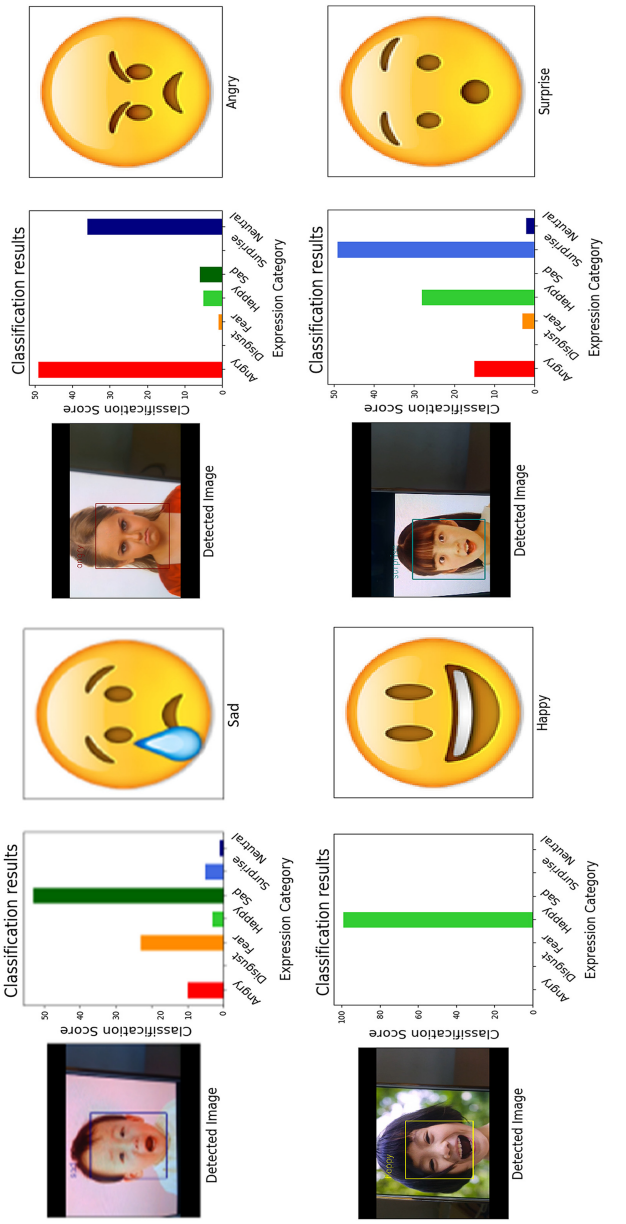

Plate 5.
Facial expression recognition results

network parameters is also very important. We have adopted a random initialization method to train the two network algorithms. The core code of Python is as follows:

```
# Model training
def train(epoch):
    if epoch > learning_rate_decay_start and learning_rate_decay_start >= 0:
        frac = (epoch - learning_rate_decay_start) // learning_rate_decay_every
        decay_factor = learning_rate_decay_rate ** frac
        current_lr = opt.lr * decay_factor
        utils.set_lr(optimizer, current_lr)  # set the decayed rate
    else:
        current_lr = opt.lr
    for batch_idx, (inputs, targets) in enumerate(trainloader):
        if use_cuda:
            inputs, targets = inputs.cuda(), targets.cuda()
        optimizer.zero_grad()
        utils.clip_gradient(optimizer, 0.1)
        optimizer.step()
        correct += predicted.eq(targets.data).cpu().sum()
```

4.4.6 Recognition results. Through the trained model, we use some children's facial expressions pictures and videos to test, and get the probability of various expressions and the final prediction results of the model. As shown in Plate 5, the histogram shows the probability of each type of facial expression, and the histogram of maximum probability is the final recognized facial expression. After testing, the recognition rate of children's facial expression reaches 81.4 percent, which can effectively distinguish whether children's facial expression is normal or not.

5. System validation

5.1 Testing environment

In this study, two kinds of mobile phones, personal computers and servers are selected as test environments. The hardware and software environments are shown in Table I.

5.2 Diagnostic procedure and interface of diagnostic system

The diagnostic process is shown in Figure 5. First, the system randomly displays one of the seven kinds of facial expressions for the tester to imitate. The system will prompt the tester

Testing equipment	Hardware environment	Software environment
OPPO R17 mobile phone	CPU:SDM670 RAM:8GB	Android
IPhone 8 mobile phone	CPU:A11 RAM:2GB	iOS
Personal computer	CPU:Intel i7 RAM:16GB	Windows 10
Server	CPU:Intel W2133 RAM:16GB	Windows Server 2019

Table I. Testing environment

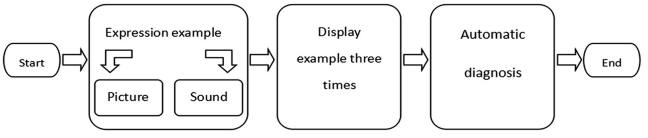

Figure 5. Automatic diagnostic procedure

to imitate the facial expression by pictures and sounds. For example, the system displays happy cartoon smiling faces, plays happy children's songs and induces children to make happy expressions. The system displays the same expression example three times and collects the tester's expression data at the same time. Then the system compares the expression examples and the actual collected expression data and gives the diagnosis results.

5.2.1 Diagnostic procedure. The diagnostic process is shown in Figure 5. First, the system randomly displays one of the seven kinds of facial expressions for the tester to imitate. The system will prompt the tester to imitate the facial expression by pictures and sounds. For example, the system displays happy cartoon smiling faces, plays happy children's songs and induces children to make happy expressions. The system displays the same expression example three times and collects the tester's expression data at the same time. Then the system compares the expression examples and the actual collected expression data and gives the diagnosis results.

5.2.2 Interface of diagnostic system. The system diagnostic interface is designed according to the diagnostic process (see Plate 6).

5.3 System testing

5.3.1 Test sample. We recruited ten normal children and ten autistic children and divided them into normal children group and autistic children group for comparative verification. The accuracy of the system is verified by the actual test of the autism diagnosis information system.

The normal group of children was provided by Amy Education School in Zhengzhou, which cooperated with us. Ten healthy children as volunteers were recruited as the normal

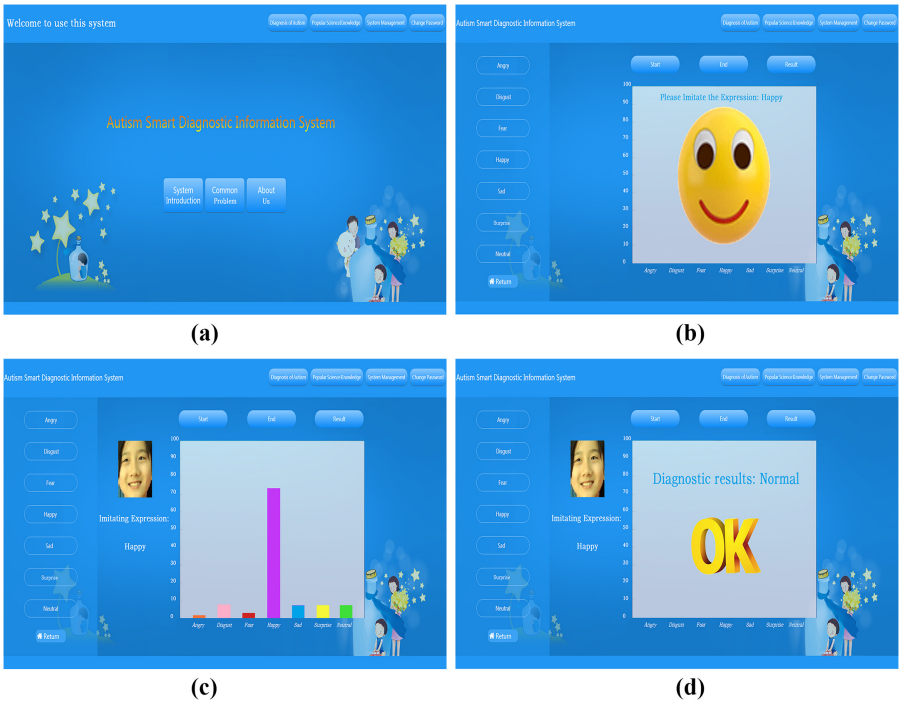

Plate 6.
(a) The main interface of the autism smart diagnosis information system, including system introduction, knowledge introduction of autism and other functions. (b) The facial expression that the system prompts the tester to simulate after starting the diagnostic process. (c) The expression analysis after diagnosis. (d) The result given by the system after three diagnoses

group for testing. These children were between 5 and 8 years old, including 5 boys and 5 girls. Parents were informed of the purpose and content of the experiment before the experiment. Children who participated in the experiment had no history of autism after being asked by their parents.

The autistic children were provided by Guangzhou Children's Care Center, which cooperated with us. Ten volunteers of autistic children were recruited as the autistic children group for testing. These children were aged between 3 and 6 years old, including 5 boys and 5 girls. Parents were informed of the purpose and content of the experiment before the experiment. The selected children with autism were diagnosed by a professional physician.

5.3.2 Test environment and process. All the tests were conducted in quiet classrooms without noise and external factors. Through our autism diagnosis information system, each child was prompted by pictures and sounds to imitate seven kinds of facial expressions and prompted to make corresponding facial responses according to the facial expressions on the pictures. At this time, the camera will capture their facial expressions, and after system analysis, they will be saved in the form of pictures in the computer of the test system (see Plate 7).

5.3.3 Test result. We used the system to test the normal combination and autistic children respectively. Finally, we compared the recognition rate of the two groups.

From Table II, the average recognition rate of each expression is angry 80 percent, disgust 70 percent, fear 80 percent, happy 100 percent, sad 80 percent, surprise 70 percent and neutral 90 percent.

Test child 1 only had a disgusting expression recognition error, and other facial expression recognition was correct, then the average recognition rate of the seven facial

(a) **(b)**

Plate 7.
(a) Test environment for normal children group. (b) Test environment for autistic children group

	Number of test children Correct identification (Y = Yes, No = N)										
Facial expression	1	2	3	4	5	6	7	8	9	10	Average recognition rate %
Angry	Y	Y	Y	N	Y	Y	Y	Y	Y	N	80
Disgust	N	Y	Y	Y	N	Y	N	Y	Y	Y	70
Fear	Y	Y	Y	N	Y	Y	Y	Y	N	Y	80
Happy	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100
Sad	Y	Y	N	Y	Y	Y	Y	Y	Y	N	80
Surprise	Y	N	Y	N	Y	Y	N	Y	Y	Y	70
Neutral	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	90

Table II.
Test results in normal children group

expressions of test child 1 was 85.7 percent. According to this method, the average recognition rates of the seven expressions from test child 1 to test child 10 were 85.7 percent, 85.7 percent, 85.7 percent, 57.1 percent, 85.7 percent, 100 percent, 71.4 percent, 100 percent, 85.7 percent and 57.1 percent, respectively. The average recognition rate is 81.4 percent.

Judging by the 60 percent threshold, there are two test children's facial expression recognition rates at 57.1 percent. This shows that in real environment, the algorithm of the system is affected by the environment and light, and the accuracy will be affected to a certain extent. However, according to the accuracy of 81.4 percent, it can basically meet the preliminary diagnostic requirements of whether the expression is abnormal or not. In the future, more real samples will be added to further improve the accuracy of the system algorithm.

The experimental results show that the errors mainly concentrate on the expressions of disgust and surprise. The main reasons are as follows:

(1) Disgust and surprise have only minor local changes in the faces of the two kinds of expressions, and there is no significant distinguishing feature.

(2) Some of the participants had little change in the two facial expressions, did not have the obvious features of the corresponding categories, approached neutral expressions and were easy to confuse.

From Table III, the average recognition rate of each expression is angry 50 percent, disgust 10 percent, fear 30 percent, happy 60 percent, sad 60 percent, surprise 20 percent and neutral 10 percent.

The average recognition rates of the seven expressions from test child 1 to test child 10 were 28.5 percent, 28.5 percent, 28.5 percent, 42.9 percent, 42.9 percent, 57.1 percent, 28.5 percent, 42.9 percent, 28.5 percent and 14.3 percent, and the average recognition rate is 34.3 percent.

Table III.
Test results in autistic children group

	Number of test children Correct identification (Y = Yes, No = N)										
Facial expression	1	2	3	4	5	6	7	8	9	10	Average recognition rate %
Angry	Y	N	N	N	Y	Y	N	Y	Y	N	50
Disgust	N	N	N	Y	N	N	N	N	N	N	10
Fear	N	N	Y	N	Y	Y	N	N	N	N	30
Happy	N	Y	Y	Y	N	Y	N	Y	N	Y	60
Sad	N	Y	N	Y	Y	Y	N	Y	Y	N	60
Surprise	Y	N	N	N	N	N	Y	N	N	N	20
Neutral	N	N	N	N	N	N	Y	N	N	N	10

Table IV.
Comparisons of two groups of children's facial expression recognition rate

Facial expression	Normal children group	Autistic children group
Angry	80%	50%
Disgust	70%	10%
Fear	80%	30%
Happy	100%	60%
Sad	80%	60%
Surprise	70%	20%
Neutral	90%	10%
Average recognition rate %	81.4%	34.3%

The experimental results show that the recognition rate of happiness and sadness is higher in the seven expressions. Testing children showed difficulty in identifying complex facial expressions such as neutrality and aversion (see Table IV).

The experimental results showed that the recognition rate of facial expressions in autistic children was significantly lower than that in normal children. All the autistic children who participated in the test had a facial recognition rate of less than 60 percent. Therefore, if the accuracy rate of facial expression diagnosis by the system was less than 60 percent, the tester would have a tendency to suffer from autism. The lower the recognition rate, the higher the tendency of autism.

6. Conclusion

In the era of rapid development of information technology, the processing of a large number of health data has brought new opportunities and challenges to medical research. The incidence of autism is increasing, which has attracted more and more attention from all aspects of society. The use of information technology, especially artificial intelligence technology, to build an autism diagnosis system has become an urgent need for doctors and patients. In this paper, an autism diagnosis system based on deep convolution neural network and expression data is constructed. After testing, it can meet the design requirements of autism diagnosis. The public can download and use the system through the network to diagnose autism conveniently. In addition, we will expand the function of the system, increase the recognition of children's physical movement and realize the diagnosis of autism from multiple perspectives.

Because the average age of children using and collecting facial expression data is between 3 and 8 years old, the system can recognize children aged 3–6 years old. Therefore, through this system, autism can be diagnosed as soon as possible. The earlier the diagnosis and treatment of autism is, the better the rehabilitation effect. Therefore, it is of great significance for the treatment of autism.

Because the training samples of the system adopt the international open facial expression database, which contains the facial expression data of children and adults in different countries and regions, the system can diagnose autism for children and adults in different countries and regions.

Of course, the system also needs to be improved through practical use. Next, we will arrange for the system to be tested in a large number of cooperative hospitals. Next, there are two main tasks to be done. The first is to collect more data of normal and autistic children's facial expressions, improve the recognition effect of the system on children's facial expressions and establish a special database of children's facial expressions. The second is to improve the system function, according to the results of facial expression diagnosis of autistic children for a detailed classification, to distinguish between severe, moderate and mild autism patients, in order to facilitate the treatment of doctors.

This study hopes to be helpful to the diagnosis of autism in remote and underdeveloped areas, so as to promote the early diagnosis and treatment of autistic children and reduce the medical costs and burdens of autistic families and society. Therefore, this study has more important social significance and application value.

References

Baron-Cohen, S., Wheelwright, S., Jolliffe, T. (1997), "Is there a "language of the eyes"? Evidence from normal adults, and adults with autism or asperger syndrome", *Visual Cognition*, Vol. 4 No. 3, pp. 311-331.

Beijing Wucai Deer Autism Research Institute (2017), *Report on the Development of Autism Education and Rehabilitation Industry in China 2*, Huaxia Publishing House, Beijing.

Cai, Y. (2018), "Facial tracking and facial expression recognition based on in-depth learning" Southeast University.

Du, J. (2018), "Research on face expression recognition based on Kernel relieff" Zhengzhou University.

Duan, Y., Wu, X. and Jinfeng (2015), "Research progress on etiology and treatment of autism", *Chinese Science: Life Science*, Vol. 9, pp. 820-844.

Ekman, P. (1992), "An argument for basic emotions", *Cognition and Emotion*, Vol. 6 Nos 3-4, pp. 169-200.

Lanlan (2018), "Research on facial expression recognition method based on multi-feature fusion", Jilin University.

Li, S. and Deng, W. (2018), "Deep facial expression recognition: a survey" arXiv preprint arXiv: 1804.08348.

Liu, Y., Huo, W. and Hu, X. (2015), "Summary of research on facial expression recognition of autistic children", *Modern Special Education*, Vol. 8, pp. 35-39.

Lucey, P., Cohn, J.F., Kanade, T., Saragih, J. and Ambadar, Z. (2010), "The extended cohn-kanade dataset (ckþ): a complete dataset for action unit and emotion-specified expression", *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, 2010, IEEE.

Mehrabian, A. (2008), "Communication without words", *Communication Theory*, Vol. 6, pp. 193-200.

Mei, J. and Hu, B. (2015), "Research and implementation of real-time face expression recognition method", *Information and Technology*, Vol. 44 No. 4, pp. 145-148.

Organization W H (1992), *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*, World Health Organization, Geneva.

Segal, D.L. (2010), "Diagnostic and statistical manual of mental disorders (DSM-IV-TR)", *The Corsini Encyclopedia of Psychology*, Vol. 1 No. 16, pp. 1-3.

Shen, X., He, Z. and Ding, X. (2013), "Computer facial expression recognition training to improve the facial expression recognition ability of autistic children", *Sci-tech Horizon*, Vol. 25, pp. 12-13.

Singh, P., Ghosh, S. and Nandi, S. (2017), "Subjective burden and depression in mothers of children with autism spectrum disorder in India: moderating effect of social support", *Journal of Autism and Developmental Disorders*, Vol. 47 No. 10, pp. 3097-3111.

Vismara, L.A. and Rogers, S.J. (2008), "The early start denver model", *Journal of Early Intervention*, Vol. 31 No. 1, pp. 91-108.

Wang, H. (2007), "Psychological and behavioral characteristics, diagnosis and evaluation of autistic children", *Chinese Journal of Rehabilitation Medicine*, Vol. 22 No. 9, pp. 853-856.

Wang, G. and Lu, M. (2015), "Research on educational games for children with autism spectrum disorders", *Modern Special Education*, Vol. 14, pp. 38-40.

Wang, Y., Xiao, L. and Chen, R. (2018), "Social impairment of children with autism spectrum disorder affects parental quality of life in different ways", *Psychiatry Research*, Vol. 2018 No. 266, pp. 168-174.

Wu, X. and Chen, S. (2018), "Research progress on quality of life and its influencing factors of primary caregivers for autistic children", *General Nursing*, Vol. 16 No. 18, pp. 2206-2208.

Yan, S. (2008), "Experimental study on facial expression processing of autistic children", East China Normal University.

Yanbin, H., Fuxing, W., Heping, X., Jing, A., Yuxin, W. and Huashan, L. (2018), "Facial processing characteristics of autism spectrum disorders: meta-analysis of eye movement research", *Progress in Psychological Science*, Vol. 1, pp. 26-41.

Yang, Y. and Wang, M. (2014), "Employment and financial burdens of families with preschool-aged children with autism", *Chinese Journal of Clinical Psychology*, Vol. 22 No. 2, pp. 295-297, 361.

Yang, J., Xing, H., Shao, Z. and Yuan, J. (2017), "Facial expression sensitivity deficits in patients with autism spectrum disorder: impact of task nature and implications for intervention", *Chinese Science: Life Science*, Vol. 47 No. 4, pp. 443-452.

Zablotsky, B., Black, L.I. and Blumberg, S.J. (2017), "Estimated prevalence of children with diagnosed developmental disabilities in the United States, 2014-2016", *NCHS Data Brief*, Vol. 291, pp. 1-8.

Corresponding author
Wang Zhao can be contacted at: creativesoft@sohu.com