Design and application of facial expression analysis system in empathy ability of children with autism spectrum disorder*

Chen Guo, Kun Zhang**, Jingying Chen, Ruyi Xu, Lei Gao
Faculty of Artificial Intelligence in Education,
National Engineering Research Center for E-Learning,
Central China Normal University, Wuhan, China
Email: zhk@mail.ccnu.edu.cn

·*Abstract*—**Empathy is an important social ability in the early childhood development. One of the significant characteristics of children with autism spectrum disorder (ASD) is their lack of empathy, which makes it difficult for them to feel and understand other people's emotions and to judge other people's behavioral intentions, leading to social disorders. This research designs and implements a facial expression analysis system that could obtain and analyze the real-time facial expressions of children when viewing stimulus materials, and then evaluate the differences of empathy ability between ASD children and typical development (TD) children. The results of this research provide new ideas for the evaluation of ASD children, and also help to develop empathy intervention plans for ASD children.**

I. INTRODUCTION

AUTISM Spectrum Disorder (ASD) is a developmental disability that can cause many social, communication and behavioral challenges [1]. The causes of ASD are still unclear, and there is no special medicine for treatment. Only early and long-term educational intervention can improve the ability and behavior of ASD children. It has caused huge difficulties for their study and development, and become a heavy burden on their families and society. In recent years, the evaluation and educational intervention of ASD have received more and more attention [2][3]. However, due to the increasing incidence of ASD worldwide, the number of educational institutions and the effectiveness of educational intervention for individual differences in ASD children are still facing severe challenges.

In the early childhood development, social disorders are regarded as the most obvious symptom of ASD children. Their social disorders could be manifested before the development of language, such as distracted attention and dull facial expressions. Many studies had shown that the lack of empathy was the main factor leading to social disorders in ASD children. The empathy refers to the ability to sense other people's emotions and to imagine what someone else might be thinking or feeling. It is a very important social ability in early childhood development [4]. Due to the lack of empathy, ASD children have difficulty in feeling and understanding the emotions of others and judging the behavioral intentions of others, resulting in social disorders. Therefore, the evaluation research of empathy is helpful to infer the probability of ASD and provide guidance for empathy intervention programs.

Facial expression is one of the main forms in the process of social interaction and can be used as a means of evaluating children's social skills. For typical development (TD) children, they usually have the ability to capture facial expressions when they are about six months old, and can recognize several basic facial expressions when they are about one year old. Some researches had tried to analyze children's facial expressions through quantitative methods. For example, Rozga et al. used physiological sensors to determine the ability of ASD children to imitate other people's expressions [5]. However, when this kind of physiological sensor was in direct contact with ASD children's facial skin, it might inhibit their spontaneous facial expressions [6][7]. Afterwards, Samad proposed to capture facial images of ASD children through optical image sensors, and analyze the movement of ASD children's muscles related to facial expressions when they received visual stimuli, so as to determine their ability to imitate facial expressions [8]. In the process of data collection, the use of non-invasive sensors could eliminate noise interference caused by sensor intrusion, and could more truly reflect the facial expression changes of ASD children.

In recent years, with the emergence of a large number of high-precision facial expression analysis algorithms, some researchers had used facial expression analysis techniques to evaluate the differences in facial expressions of ASD children and TD children [9]. For example, Coco et al. used a computer vision-based facial expression analysis method to quantify the generation of facial expressions, and proved that ASD children and TD children had different facial expressions in response to external stimuli [10].

Based on the promising work proposed by the above studies, this paper proposed a facial expression analysis system to evaluate the empathy ability of children. By collecting facial expression data of ASD children and TD children when they received visual stimuli, the facial expression analysis system was used for preprocessing and

facial expression analysis to determine whether ASD children and TD children can produce corresponding emotional responses and the corresponding degree of emotional arousal when receiving external visual stimuli, so as to study the difference in empathy ability between ASD children and TD children.

The second section of this paper presented the system framework and introduced the design of each module. The third section presented the system implementation, including the data processing process. The fourth section introduced the system application, including experimental design, as well as the analysis and discussion of experimental results.

II. SYSTEM DESIGN

A. System framework

The main function of the facial expression analysis system was to analyze facial expressions in real time through the camera and visualize the analysis results. It was divided into four modules, which were real-time data loading module, data preprocessing module, expression analysis module and results visualization module. The system framework was shown as in Fig. 1.
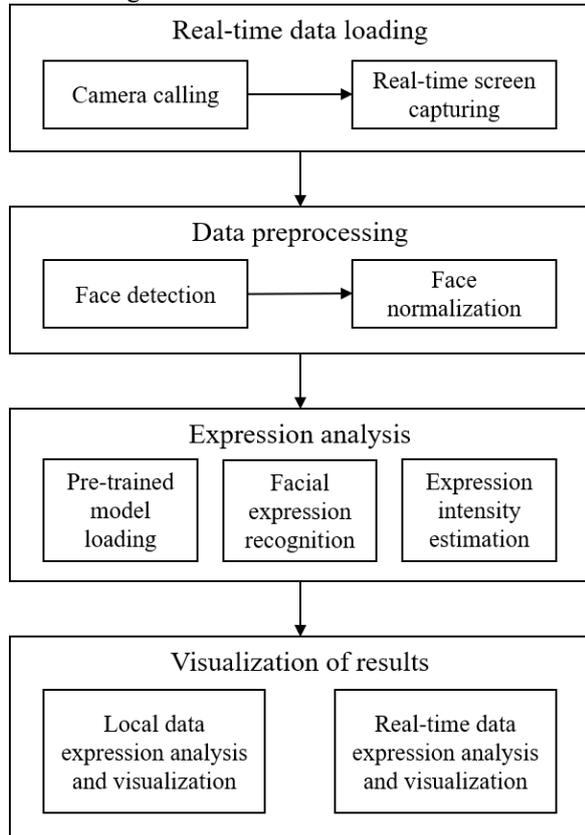

Fig. 1 System framework

B. Real-time data loading module

The real-time data loading module called the currently available camera, read and displayed the frame of the camera, and then input the captured images to the data preprocessing module frame by frame. If there was currently no available

camera, this module would pop up a prompt to inform the user. The flowchart of this module was shown as in Fig. 2.
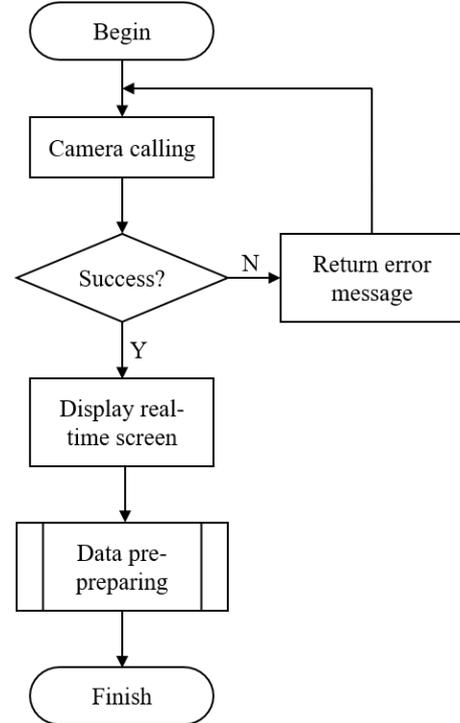

Fig. 2 Real-time data loading flowchart

C. Data preprocessing module

There were usually some noises that were not related to expressions (such as background, clothing, etc.) during the collection of facial expression data. These factors unrelated to expressions would affect the extraction effect of the convolutional neural network on the expression-related features in the images, and affect the expression recognition effect of the model. Therefore, it was necessary to preprocess the expression image before inputting the model. The data preprocessing module of this system included two steps: face detection and face normalization. Face detection was done using MTCNN [11], and face normalization was done using OpenCV image processing tools.

D. Expression analysis module

The expression analysis module had two main functions, namely the loading of the pre-trained model and expression analysis. In order to reduce the waiting time when loading the model, the user could set the storage directory of the model when selecting the function, and the system would complete the loading of the model in advance. The expression analysis function used the model trained in the previous step to perform expression classification and intensity estimation on the input expression samples, and save the results in the corresponding directory. The flowchart of this module was shown as in Fig. 3.
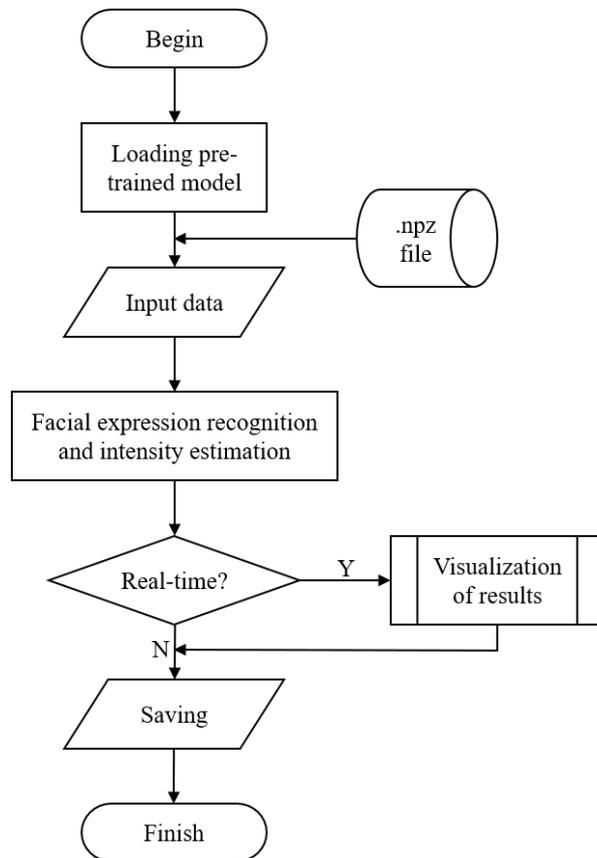
Fig. 3 Expression analysis flowchart

E. Results visualization module

After the expression recognition and intensity estimation were completed, the visualized results would be directly displayed in the interface to enhance user experience. The content included the face frame generated by face detection, the results of facial expression recognition and intensity estimation, etc.

III. SYSTEM IMPLEMENTATION

A. System development environment

The facial expression analysis system was developed on the 64-bit Windows 10 version 1803, using PyCharm 2019 development environment, Python version 3.6.6 and PyQt5 toolkit for coding. Finally, the PyInstaller toolkit was used to package the code files to generate executable files. The hardware environment was Intel Core i7-8700K CPU, with 64 GB memory and 2 TB disk space.

B. Real-time data display

When the user selected the real-time detection function on the main window, the system entered the real-time data display interface. The Video Capture method in the OpenCV library was used to call the camera and display real-time images. If no cameras were available, a prompt would pop up to tell the user. When the system could capture the camera image normally, the user could click the start detection button to start the real-time recognition function.

and the system would input the captured image frame by frame into the preprocessing module for image processing. The real-time display interface was shown as in Fig. 4.
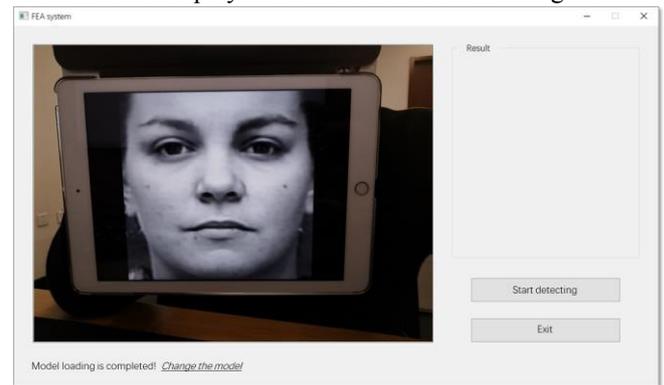

Fig. 4 Real-time data display interface

C. Data preprocessing

The data preprocessing was mainly divided into two steps. First, the face detection algorithm was used to detect the facial feature points and the face marker frames. Second, the face normalization was done using the detection results, to obtain the standard facial expression image. The specific methods of face detection and normalization were as follows:

(1) Face detection

The cascaded multi-task convolutional neural network (MTCNN) was used for face detection. The algorithm was carried out in three steps. First, the Proposal-Net was used to quickly generate face location candidate frames. Second, the Refine-Net was used to optimize the facial candidate frames generated in the first step, eliminate the incorrect frames, and capture the high-quality candidate frames through the NMS algorithm. Third, the Output-Net was used to locate and mark five key feature points of the face while removing overlapping candidate frames. Face detection was performed on the data through MTCNN algorithm, and a series of labeled data with face detection results and 5 key feature points were obtained.

(2) Face normalization

According to a series of face coordinates and feature point information obtained by face detection, most of the noises that were not related to facial expressions in the image could be eliminated after cropping. Because the distance between the participant's face and the collection device could not be fixed during data collection, the size of the face extracted in the face detection step was different. But the deep network model required input images of uniform size. Therefore, it was necessary to normalize the image to obtain an expression image of a uniform size. In addition, in order to retain the areas that were highly related to facial expressions such as the corners of the mouth and the eyes after image cropping and size normalization, the image needed to be geometrically preprocessed so that the processed image could maintain the same proportions as the original image. The specific method was to extract the coordinate points of

the left eye and the right eye and the coordinate data of the face detection frame from the feature point mark data, call the fitgeotrans function in OpenCV to perform geometric transformation and fitting of the coordinate point pair group, align and crop the image using the normalization parameters and the coordinates of the face detection frame. Finally, the face image was normalized to 128×128 pixels.

D. Expression analysis

When the real-time image of the camera was displayed, the user could click the start detection button to input the image sequence into the expression analysis module for expression recognition and intensity estimation using the loaded model, and output the recognition result to the visualization module and save to the specified path. If the system did not detect a face in the captured image, it would output "No face detected" on the screen to prompt the user.

E. Results visualization

When the real-time detection was working, if the facial expression was successfully detected by the analysis module, the result would be displayed according to the intensity and category of the expression. The face frame for face detection was displayed on the real-time screen, and the expression category and intensity were displayed in the upper left corner of the real-time screen. On the right side of the screen, the face detection results were displayed in the form of a histogram and marked under the corresponding category. The figure was drawn using the matplotlib package in the Python toolkit. The horizontal axis represented the Neutral Expression (NE) and six types of basic expressions, which were Anger (AN), Disgust (DI), Fear (FE), Happiness (HA), Sadness (SA), Surprise (SU). The vertical axis represented the intensity of these six types of basic expressions. When the expression intensity was zero, that indicated a neutral expression, the sample was displayed in the NE column by default. The results visualization interface was shown as in Fig. 5.
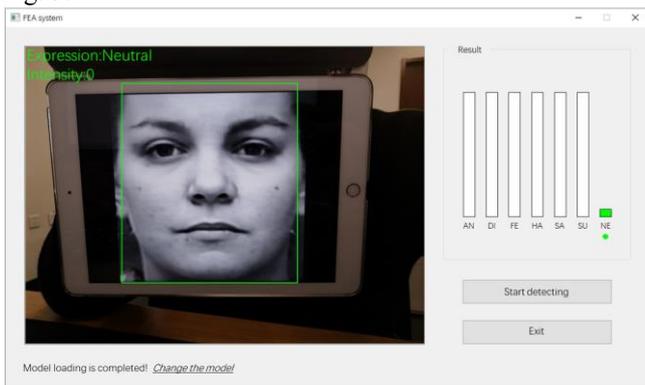

Fig. 5 Results visualization interface

IV. SYSTEM APPLICATION

The facial expression analysis system proposed in this paper was used to evaluate the empathy ability of children. Firstly, dynamic videos of facial expressions of ASD children and TD children when receiving visual stimuli were collected, intercepted and saved in frames, and organized into a data set of children's expressions. Then the facial expression analysis system was used to analyze the children's expression data set to determine whether ASD children and TD children could understand and produce corresponding emotional responses when receiving external visual stimuli, and the degree of corresponding emotional arousal. These would serve as the basis for evaluating children's empathy ability.

A. Experimental materials

Many studies had shown that the facial expression disorder of ASD children was usually manifested as the expression recognition disorder to negative expressions (such as anger, fear, etc.) was often greater than that of positive expressions (such as happiness, surprise, etc.). Liping Gu proposed that children's ability to recognize expression was usually affected by the emotional intensity of pictures, leading to differences in the results of children's recognition of expressions in pictures [12]. Generally, pictures with positive expressions and high emotional intensity were more likely to be recognized, while children had lower ability to recognize negative expression and pictures with low intensity [13]. Therefore, in order to make the collected data have a more obvious distinction, five segments of positive emotion stimulus materials were selected from the initial material library for the experiment. These materials were all from the animated version of the children's themed sitcom, called "Family with Children". Each video segment was selected by experts in the special education field after analyzing and evaluating factors such as video duration, expression type, difficulty in understanding, and emotional arousal. Each selected segment was about 12 seconds long, and the video themes included: being praised, helping others, rehearsing the program, going home from school, and eating snacks.

B. Experimental participates

The children who participated in the data collection of the experiment were aged 3-7 and came from a kindergarten (30 children in TD group) and a special children's rehabilitation institution (30 children in ASD group) in Wuhan. All children in ASD group were clinically diagnosed with ASD (according to DSM-5, 2013), normal or corrected visual acuity, and no other respiratory diseases, childhood schizophrenia, epilepsy and other organic brain diseases.

Before the experiment, the agreements were signed with the kindergarten and institution, and the informed consent forms were signed with their parents to protect the privacy of these children participating in the experiment.

C. Experimental design

In order to avoid the mutual interference of different video materials on emotional awakening, each video segment was played after an interval of 30 seconds from the end of the previous segment. A question session would be

conducted on the content of the video material to determine whether the child understood the material in the video.

When children watched the stimulus, the camera recorded their emotional response. A total of 60 segments of effective video data were collected from 30 ASD children and 30 TD children. The video resolution was 720×576 pixels. The video processing module in OpenCV was used to process the video data and save it as the image sequence by frame. Each image sequence contained about 1200 frames, which recorded the spontaneous changes in expression of each child under same emotional stimulus. The psychologists labeled each sequence with the child's main facial expression category.

In this experiment, a pre-trained model was used to test all frames in the children's data set. The model was trained on the CK+ dataset using the expression analysis method of the tensorflow-gpu 1.13.1 framework. The number of iterations was set to 3000 poaches. The early stopping method was used in the training process to ensure that the model got the best generalization ability and could better fit the data. When the accuracy of the validation set dropped, the training process was terminated and the model was saved for use in the children's facial expression detection.

D. Experimental results and analysis

The facial expression analysis system was used to analyze the collected children's expression data set to explore the difference between the empathy ability of ASD children and TD children. Fig. 6 showed the experimental results of a certain child's expression sequence analyzed by the facial expression analysis system, where Fig. 6(a) was the result of the distribution diagram of the expression sequence, Fig. 6(b) was the expression category and intensity of the current frame, Fig. 6(c) was the expression category and intensity result of the sequence.

It could be seen that in Fig. 6(a), the overall distribution of the data was statistics. Among them, the accuracy rate referred to the percentage of the children's facial expression data label consistent with the expression category predicted by the model in the overall sample. The tracking rate referred to the ratio of the emotional feedback generated by the child on the facial expression which was consistent with the emotion of the stimulus material. The child shown in the picture could better understand the emotion expressed by the stimulus material and could reflect it on facial expressions. Therefore, it could be considered that this child had a good empathy ability. Fig. 6(b) could help users analyze the wrong samples by viewing each frame of image sequence. Fig. 6(c) showed the overall recognition of the child's facial expression sequence. This function could be used to analyze whether the child had produced the corresponding emotion under the corresponding stimulus material in combination with the video material.
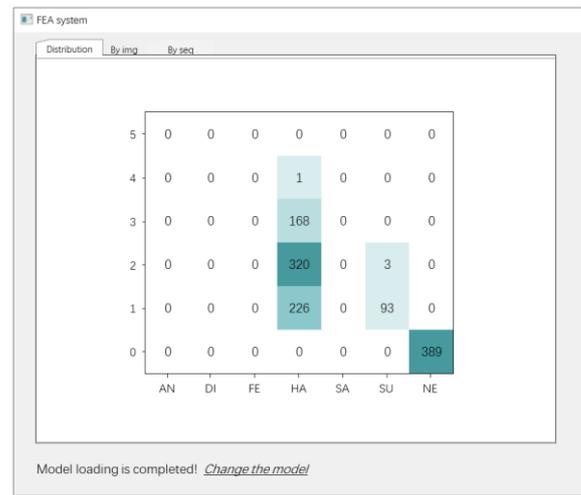
(a) The distribution diagram of the expression sequence result
(Horizontal axis: six types of basic expressions and neutral expression.
Vertical axis: the intensity of expressions.)
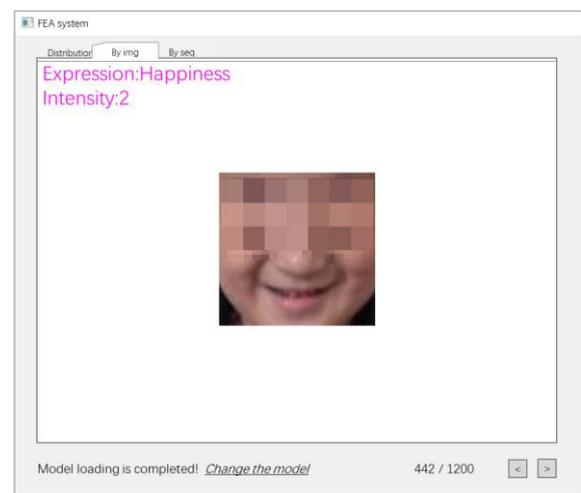
(b) The expression category and intensity of the current frame
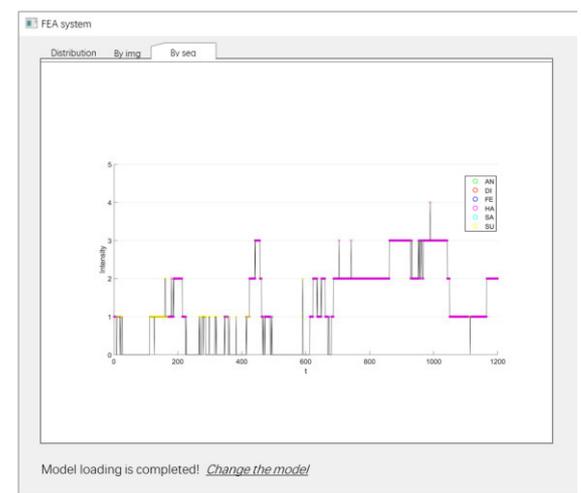
(c) The expression category and intensity result of the sequence
(Horizontal axis: frame sequence number.
Vertical axis: the intensity of expressions.)

Fig. 6 Experimental results of a certain child's expression sequence

TABLE I.
ANALYSIS RESULTS OF ALL CHILDREN'S FACIAL EXPRESSIONS

Group	Accuracy rate	Tracking rate
Autism Spectrum Disorder (ASD)	90.0%	23.3%
Typical Development (TD)	93.3%	80.0%

The results of all children's expression data analyzed by the facial expression analysis system were shown in Table I. It could be concluded that these two groups of children could produce corresponding feedback when they receive the same emotional stimulus materials, but the proportion of children in the TD group that produced the same emotion as the stimulus was significantly higher than that in the ASD group. It could be inferred that there were significant differences in emotional understanding and facial expression recognition between TD children and ASD children.
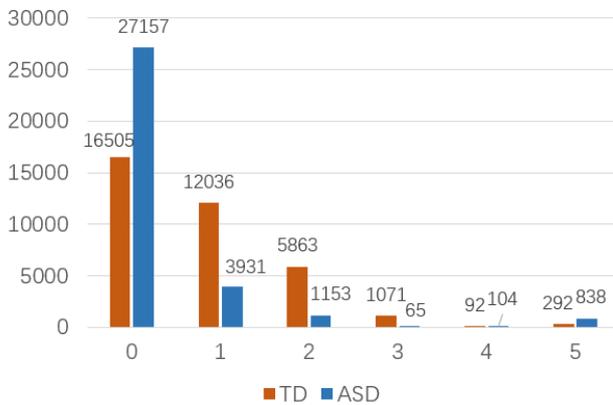

Fig. 7 The statistics of the expression intensity results (Horizontal axis: the intensity of expressions. Vertical axis: the number of samples.)

Fig. 7 was the statistics of the expression intensity results of the TD group and the ASD group. The expression intensity was divided into 6 levels (from 0 to 5). It could be found that in samples with an intensity of 0, the number of samples in ASD group was significantly more than that in TD group. For the samples with expression intensity in the interval [1,3], the number of samples in TD group was significantly more than that in ASD group. It showed that the same stimulus material had higher emotional arousal to TD children than to ASD children, and most ASD children could not produce correct feedback when receiving external emotional stimuli. From the perspective of empathy, there might be two reasons. On the one hand, ASD children might have defects in feeling and experiencing the emotions of others, which made it difficult or even impossible for them to understand other people's emotions. On the other hand, ASD children might have defects in expressing emotions and could not express their true inner emotions through facial expressions. In the interval between expression intensity 4 and 5, the number of samples in ASD group was slightly higher than that in TD group. This might be because many ASD children were often accompanied by symptoms

such as distracted attention and impulsiveness. When viewing stimulus materials, there might be interference factors such as large changes in head posture or occlusion of relevant facial expressions in ASD children, leading to errors in the expression intensity identification of some samples.

In order to evaluate the empathy ability of the two groups of children more comprehensively, the temporal statistical method of expression category and expression intensity was used for further analysis. Specifically, the detailed criteria for evaluating empathy (E) were set as follows: (1) According to the research work in Reference [14], whether child's main facial expression category (C) was consistent with emotional stimuli (S) was determined by analyzing the consistency of the child's facial expression and the emotion of the stimulus. (2) Inspired by Reference [15], the expression awakening duration was used as the time feature to determine whether the expression activation duration (A) exceeded the threshold. In this experiment, the threshold was set to 1/3. When the total duration of samples whose facial expression intensity was greater than 0 exceeded the threshold, the expression was considered to have been activated.

$$E = \begin{cases} 2 & A > \frac{1}{3}, C = S \\ 1 & A > \frac{1}{3}, C \neq S \mid A \leq \frac{1}{3}, C = S \\ 0 & A \leq \frac{1}{3}, C \neq S \end{cases} \qquad (1)$$

For empathy indicators, 0, 1, and 2 respectively represent the low, medium, and high levels of empathy ability. The larger the value, the stronger the empathy ability. According to the above criteria, the children in the ASD group and the TD group could be classified and counted. The results were shown in Table II.

TABLE II.
EVALUATION RESULTS OF CHILDREN'S EMPATHY ABILITY

Expression category consistency	Activation duration	TD group	ASD group	Empathy Level
C=S	A>1/3	22	0	2
	A≤1/3	2	7	1
C≠S	A>1/3	6	3	1
	A≤1/3	0	20	0

It could be seen that in the TD group, 22 children were evaluated as high-level empathy ability, and 8 children were evaluated as medium-level. In the ASD group, 20 children had the low-level results of empathy ability, and 10 children had the medium-level results. The experimental results further proved the effectiveness of this expression analysis system in evaluating children's empathy ability.

V. CONCLUSION

This paper introduced the design and implementation of the facial expression analysis system, which could quickly recognize and estimate the intensity of facial expression data captured by the camera, provide visualization of the results,

providing a great convenience for users to analyze facial expression. This system was used to evaluate children's empathy ability and had achieved good results. It verified the effectiveness of the proposed method and the application value of this system. The results of this research provided a certain basis for the evaluation of children's empathy ability, and also helped to develop empathy intervention programs for ASD children.

Future research work is to further explore and improve the application of facial expression analysis in the evaluation of children's empathy ability, such as expanding the scale of experimental data to obtain more applicable conclusions, and optimizing system algorithms to improve recognition accuracy. It is also possible to combine facial expression analysis with EEG signal and other methods to evaluate the children's empathy ability in multiple dimensions, so as to obtain more accurate evaluation results.

REFERENCES

[1] American Psychiatric Association, "Diagnostic and Statistical Manual of Mental Disorders: DSM-V," Washington, DC: American Psychiatric Publishing, pp. 55-59, 2013. ISBN: 978-0-89042-554-1.

[2] K. Pancerz, W. Paja and J. Gomuła, "Random forest feature selection for data coming from evaluation sheets of subjects with ASDs," 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), 2016, pp. 299-302. http://dx.doi.org/10.15439/2016F274.

[3] A. Kołakowska, A. Landowska, M. R. Wrobel, et al, "Applications for investigating therapy progress of autistic children," 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), 2016, pp. 1693-1697. http://dx.doi.org/10.15439/2016F507.

[4] Xiaoxia Zhang, Ye Wang, Xin Liu, et al, "A Review of Studies on Empathy Development of People with Autistic Spectrum Disorders," Chinese Journal of Special Education, vol. 8, pp. 48-55, 2019. (in Chinese). http://dx.doi.org/10.3969/j.issn.1007-3728.2019.08.009.

[5] Rozga A, King T Z, Vuduc R W, et al, "Undifferentiated facial electromyography responses to dynamic, audio-visual emotion displays in individuals with autism spectrum disorders," Developmental Science, vol. 16(4), pp. 499-514, 2013. http://dx.doi.org/10.1111/desc.12062.

[6] Beall P M, Moody E J, Mcintosh D N, et al, "Rapid facial reactions to emotional facial expressions in typically developing children and children with autism spectrum disorder," Journal of Experimental Child Psychology, vol. 101(3), pp. 206-223, 2008. http://dx.doi.org/10.1016/j.jecp.2008.04.004.

[7] Mathersul D, Mcdonald S, Rushby J A, "Automatic facial responses to affective stimuli in high-functioning adults with autism spectrum disorder," Physiology & Behavior, vol. 109(1), pp.14-22, 2013. http://dx.doi.org/10.1016/j.physbeh.2012.10.008.

[8] Samad M D, Bobzien J L, Harrington J W, et al, "Non-intrusive optical imaging of face to probe physiological traits in Autism Spectrum Disorder," Optics & Laser Technology, vol. 77, pp. 221-228, 2016. http://dx.doi.org/10.1016/j.optlastec.2015.09.030.

[9] Leo M, Carcagnì P, Distante C, et al, "Computational assessment of facial expression production in ASD children," Sensors, vol. 18(11), pp. 3993, 2018. http://dx.doi.org/10.3390/s18113993.

[10] Coco M D, Leo M, Carcagni P, et al, "A Computer Vision Based Approach for Understanding Emotional Involvements in Children with Autism Spectrum Disorders," International Conference on Computer Vision Workshop, pp. 1401-1407, 2017. http://dx.doi.org/10.1109/ ICCVW.2017.166.

[11] Zhang K, Zhang Z, Li Z, et al, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Processing Letters, vol. 23(10), pp. 1499-1503, 2016. http://dx.doi.org/ 10.1109/LSP.2016.2603342.

[12] Liping Gu, Jin Jing, Yu Jin, et al, "Research of the relationship between the ability of emotion understanding and social adaptation in children with high functioning autism," Chinese Journal of Child Health Care, vol. 21(1), pp. 16-19, 2013. (in Chinese)

[13] Hou Hu, Mingfan Wu, Shenghua Hu, "Facial emotion recognition in children with high functioning autism," Chinese Journal of School Health, vol. 35(8), pp. 1146-1149, 2014. (in Chinese)

[14] Anastassiou-Hadjicharalambous X, Warden D, "Convergence between physiological, facial and verbal self-report measures of affective empathy in children," Infant and Child Development: An International Journal of Research and Practice, vol. 16(3), pp. 237-254, 2007. http://dx.doi.org/10.1002/icd.464.

[15] Li B, Mehta S, Aneja D, et al, "A Facial Affect Analysis System for Autism Spectrum Disorder," IEEE International Conference on Image Processing, pp. 4549-4553, 2019. http://dx.doi.org/10.1109/ICIP.2019. 8803604.