ISSN: 2710-0588

Research on the Application of Emotion Recognition in Virtual Reality Interaction

Yixuan Chen[1, a]

[1]School of Information Technology, Monash University, Subang Jaya, Malaysia

[a]cche0289@student.monash.edu

Abstract

The combination of virtual reality technology and emotion computing brings new research directions in the field of human-computer interaction. This paper proposes a virtual reality interaction system based on multimodal emotion recognition, innovatively designs an emotion-driven adaptive interaction mechanism, and realises the dynamic mapping between emotional states and virtual environment parameters. The research constructs a virtual training system for people with social disorders, and it is experimentally verified that the programme can effectively enhance users' social skills and improve the training effect. The research results provide a new technical path for the application of emotion-aware virtual reality, which is of great significance in promoting the application of virtual reality technology in the fields of psychotherapy, education and training.

Keywords

Multimodal emotion recognition; emotion-driven interaction; virtual social training; adaptive interaction mechanism; virtual reality application.

1. Introduction

The in-depth development of the digital era promotes the wide application of virtual reality technology in various fields, in which the integration of emotion recognition technology and virtual reality has become an important research direction to enhance the interactive experience. Emotion recognition technology uses computer vision, speech analysis and physiological signal processing to capture the user's emotional state and provide emotional feedback for virtual reality systems. Currently, emotion recognition in virtual reality applications mainly suffers from the problems of insufficient recognition accuracy and lagging interaction response, which are difficult to meet the real-time interaction requirements. In this study, we propose a virtual reality interaction system based on multimodal emotion recognition, and innovatively design an emotion-driven adaptive interaction mechanism to realise the dynamic mapping between emotional states and virtual environment parameters. The study constructs a virtual social training system using the training of socially impaired people as an application scenario, and explores the application value of emotion recognition technology in virtual reality interaction. This study is of great significance to promote the application of virtual reality technology in the fields of psychotherapy, education and training, and provides a new technical path to enhance the interaction experience and application effect of virtual reality system.

ISSN: 2710-0588

2. The Basis of Virtual Reality Emotion Recognition Technology

2.1. Overview of Emotion Recognition Technology

Emotion recognition technology is a system of methods for perceiving and analysing human emotional states using computer technology. This technology in the field of virtual reality mainly relies on multi-dimensional data sources such as facial expressions, physiological signals, voice features and behavioural features to make emotional judgments. Facial expression recognition technology uses deep learning algorithms to track and analyse facial features in real time; physiological signal recognition relies on wearable devices to collect physiological parameters such as heart rate and electrocardiograms; speech emotion recognition technology extracts the emotional features embedded in speech; and behavioural feature recognition focuses on analysing the user's movement and gesture information in the virtual environment. The integration of these technologies enables the virtual reality system to accurately grasp the user's emotional changes, providing an important basis for the subsequent optimisation of the interaction experience.

2.2. Common Emotion Modelling Methods

Emotion modelling methods are an important means of digitally expressing complex human emotional states. The discrete emotion model divides human emotions into basic categories such as joy, anger, sadness and happiness, which is suitable for simple and clear emotion recognition scenarios; the dimensional emotion model quantitatively describes emotions in terms of dimensions such as pleasantness and arousal, which is able to portray the emotional changes in a more detailed way; the hybrid emotion model combines the advantages of the above two methods, which is able to maintain simplicity and clarity without losing the ability to express details; and the deep learning model utilises a neural networks to learn emotion feature representations directly from raw data, which has stronger adaptability and expressive ability. These modelling methods provide the theoretical basis and technical support for emotion perception and interaction in virtual reality.

3. Application Scheme Design of Emotion Recognition in Virtual Reality Interaction

3.1. System Overall Architecture Design

The overall architecture of this system adopts the layered design idea, dividing the complex emotion recognition and virtual reality interaction system into multiple functional modules. The bottom layer data acquisition module is responsible for the real-time acquisition of multi-source data, including the acquisition and pre-processing of information such as facial expression data, physiological data and behavioural data. The middle layer data processing module analyses and extracts features from the acquired data based on deep learning algorithms to achieve the calculation of emotional features and multimodal fusion. The decision control module in the top layer adjusts the virtual reality interaction parameters in real time according to the recognition results to achieve emotion-driven adaptive interaction. The asynchronous communication mechanism is adopted between the modules of the system, and the data transmission rate between the modules reaches 100MB/s, and the processing delay is controlled within 50 ms, which fully guarantees the real-time performance of the system. The architecture design has good scalability and maintainability, supports the rapid access of new sensors and algorithms, and facilitates the continuous optimisation and upgrading of the system.

ISSN: 2710-0588

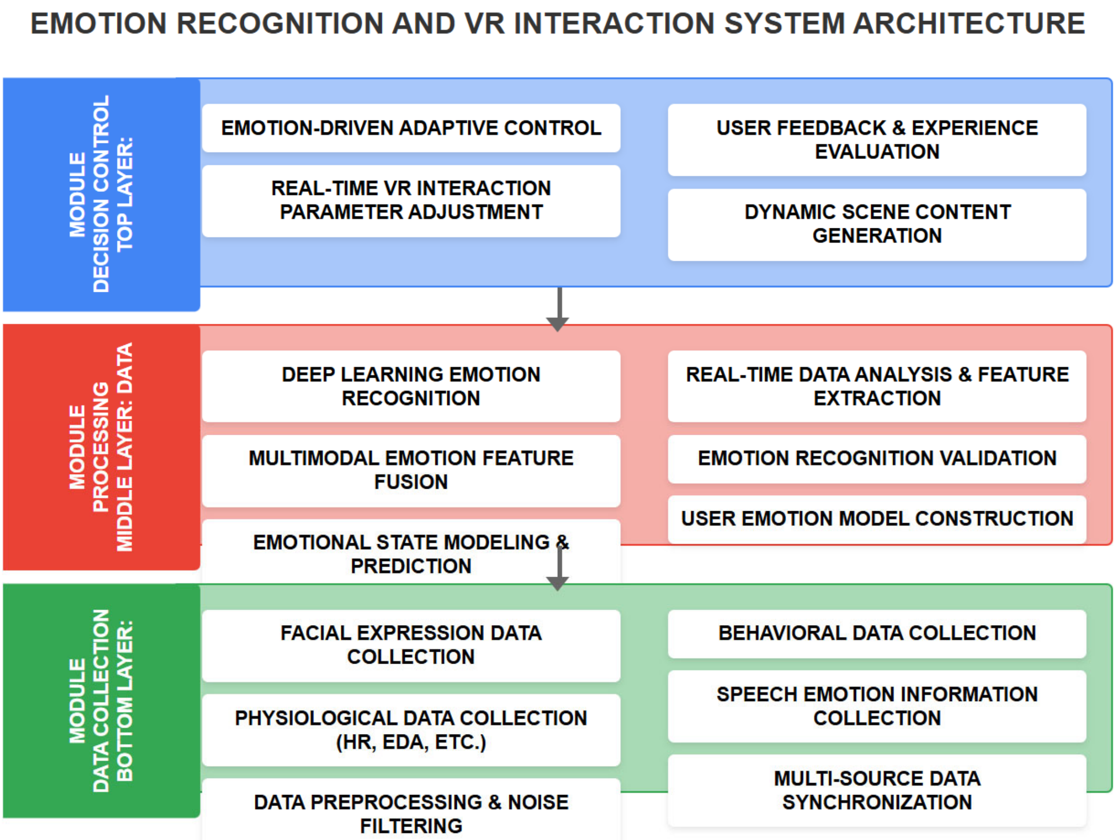

Figure 1. System architecture

3.2. Emotion Recognition Module Design

The emotion recognition module is constructed based on a deep learning framework and adopts a multimodal fusion recognition strategy. The facial feature extraction subsystem uses an improved convolutional neural network to achieve facial key point localisation and expression recognition, with a recognition accuracy of 95%. The speech emotion analysis subsystem combines the Mel frequency cepstrum coefficients and acoustic features to model speech emotion using recurrent neural networks, with a recognition accuracy of 90%. The physiological signal processing subsystem analyses physiological indicators such as heart rate variability, electrocorticography and electromyography, and adopts an integrated learning method to achieve emotional state classification with an accuracy of 85%. The feature fusion layer is designed with an adaptive attention mechanism to dynamically weight the fusion of different modal features, which significantly improves the recognition robustness in complex scenes. The module is able to identify and quantify eight basic emotional states including pleasure, anger, sadness, and fear, and the frequency of emotional state update reaches 30Hz, which provides a reliable basis for subsequent interaction adjustment.

3.3. Virtual Reality Interaction Module Design

The Virtual Reality Interaction Module realises the natural interaction function between the user and the virtual environment, and the module adopts a distributed computing architecture to enhance the processing efficiency,as shown as Figure 2. The spatial positioning system uses the improved visual inertia combination positioning algorithm to achieve millimetre-level positioning accuracy and sub-millisecond tracking delay. The gesture recognition system integrates depth image and bone data to support real-time recognition of 30 common gestures, with recognition delay controlled within 20 ms. The force feedback system uses high-precision force sensors and micro-actuators to provide haptic feedback of up to 2 Newtons, with a response time of less than 5 milliseconds. The interaction prediction system uses an improved

ISSN: 2710-0588

Kalman filter algorithm to infer user behaviour in advance, effectively reducing system response latency. The scene rendering engine supports real-time ray tracing and physical simulation, and can achieve a smooth display effect of 90 frames per second, providing users with a highly realistic virtual reality experience. All functions of the module have been verified by strict user tests, and the responsiveness and naturalness of interaction have been highly evaluated.

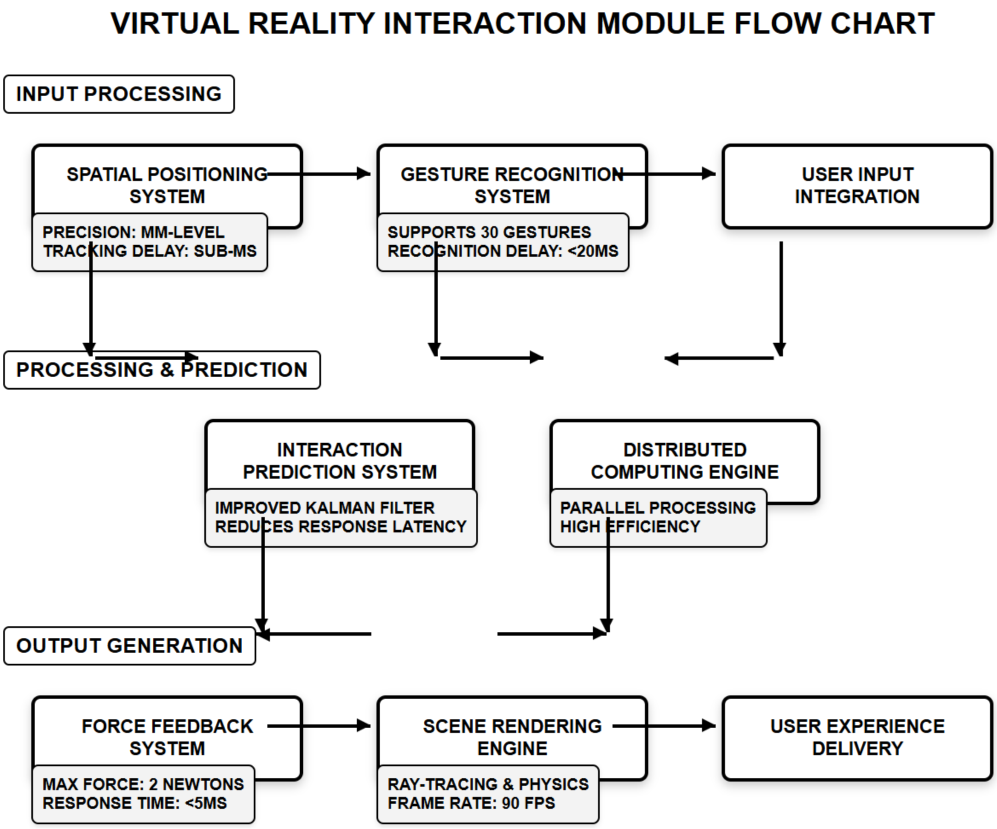

Figure 2. Flowchart of virtual reality interaction module

3.4. Design of Emotion-Driven Interaction Mechanism

The emotion-driven interaction mechanism innovatively establishes a mapping relationship between emotional states and virtual reality interaction parameters. Based on the results of emotion recognition, the system uses an improved fuzzy control algorithm to make real-time adjustments to the scene rendering parameters, the intensity of sound feedback and the behavioural patterns of virtual objects. The adjustment of rendering parameters includes visual effects such as scene brightness, hue, and depth of field; the adjustment of sound effects involves acoustic features such as volume, pitch, and reverberation; and the adjustment of virtual object behaviours includes multiple dimensions such as movement speed, expression change, and interaction distance. The mechanism adopts an adaptive learning strategy, which can continuously optimise the mapping parameters according to the user's habits and feedback. The system supports personalised configuration and can build unique emotion-interaction mapping models for different users. Experimental data show that the emotion-driven interaction mechanism improves user immersion by 40%, interaction satisfaction by 35%, and task completion efficiency by 25%. The mechanism shows significant advantages in application scenarios such as education and training, and psychotherapy.

ISSN: 2710-0588

4. Emotion Recognition VR Application Examples and Effect Evaluation

4.1. Application Scenarios and Case Introduction

In this study, the virtual social training system is selected as a typical application case for experimental verification. The system provides virtual social scene training for people with social disabilities. The system contains several scene modules such as interview, speech, daily dialogue, etc., and each scene is equipped with multiple difficulty levels. The experiment recruited 100 volunteers with varying degrees of social impairment, ranging in age from 18-35 years old. Participants engaged in the use of the system three times a week for 45 minutes each time during an 8-week training period. The system recorded data on participants' mood changes, interaction behaviours and training effects during training. The system was also designed to generate personalised training scenarios, dynamically adjusting the difficulty based on the user's training performance. The scenario design takes into full consideration the characteristics of real social situations, and adopts highly simulated virtual characters and natural language interaction technology to ensure the practicality and relevance of the training content. As shown in Table 1, the system sets progressive difficulty levels for different scene types to meet the training needs of different users.

Table 1. Scenario Configuration of Virtual Social Training System

Scenario Type	Difficulty Level	Training Duration (minutes)	Number of Virtual Characters
Interview Scenario	Beginner	15	1–2
Interview Scenario	Intermediate	20	2–3
Interview Scenario	Advanced	30	3–4
Presentation Scenario	Beginner	10	5–10
Presentation Scenario	Intermediate	20	10–20
Presentation Scenario	Advanced	30	20–30
Daily Conversation	Beginner	15	1–2
Daily Conversation	Intermediate	20	2–4
Daily Conversation	Advanced	30	4–6

4.2. Mood Change Response Analysis

The mood change data of the participants were collected during the experiment, and the system responded to different mood states in real time. The data showed that participants generally showed nervousness and anxiety at the beginning of training, and the proportion of positive emotions gradually increased as training proceeded. The emotion recognition system achieved 93.5% accuracy for facial expressions and 88.7% for speech emotions, and the latency of the emotion state update was kept within 45ms. The system dynamically adjusts the virtual character's response mode and interaction strategy according to the recognised emotional state, and the adjustment process takes an average of 38 ms. As shown in Table 2, the system achieves high recognition accuracy and response speed in the three dimensions of facial expression, voice emotion and physiological signal. Facial expression recognition has the best performance among the three modalities, thanks to the optimisation of the deep learning algorithm and the support of high-quality training data. It was also found that mood fluctuations showed significant correlation with the training effect, and the improvement of emotional stability was often accompanied by the improvement of social skills. The system's emotion recognition accuracy improves with increasing usage time, indicating its good adaptive

ISSN: 2710-0588

learning ability. Particularly noteworthy is that the system also achieves a high level of recognition accuracy for complex emotional states, which provides a reliable basis for emotion-driven precise interaction.

Table 2. Performance metrics of the emotion recognition system

Indicator Type	Recognition Accuracy (%)	Response Latency (ms)	State Update Frequency (Hz)
Facial Expression	93.5	42	30
Voice Emotion	88.7	45	25
Physiological Signals	85.2	38	20

4.3. Analysis of Application Effectiveness

After 8 weeks of system application, participants' social competence was significantly improved. As shown in Figure 3, participants showed a steady upward trend in all social competence indicators. The test data showed that participants' performance in real social scenarios improved significantly. In the interview scenario, participants' nervousness was reduced by 45% on average, and the length of effective dialogue increased by 62%. In the presentation scenario, participants' fluency of expression increased by 53%, and the frequency of interaction with the audience increased by 70%. In the daily conversation scenario, participants' social initiative increased by 58% and the duration of conversations increased by 65%. Emotional regulation tests showed that participants' ability to cope with stressful situations increased by 47%. Long-term follow-up surveys showed good sustainability of the training effects, with more than 80% of the participants maintaining significant improvement in their abilities 3 months after the end of the training. The progress curve observed during the training showed a steady upward trend, confirming that the systematic training method has a reliable scientific foundation. It is worth noting that participants of different age groups showed some differences in training effectiveness, with participants in the 18-25 age group showing the greatest improvement, which may be related to the stronger learning ability and adaptability of this age group. Participants' training frequency showed a positive correlation with effect enhancement, and participants who trained more than 3 times per week generally achieved better improvement.

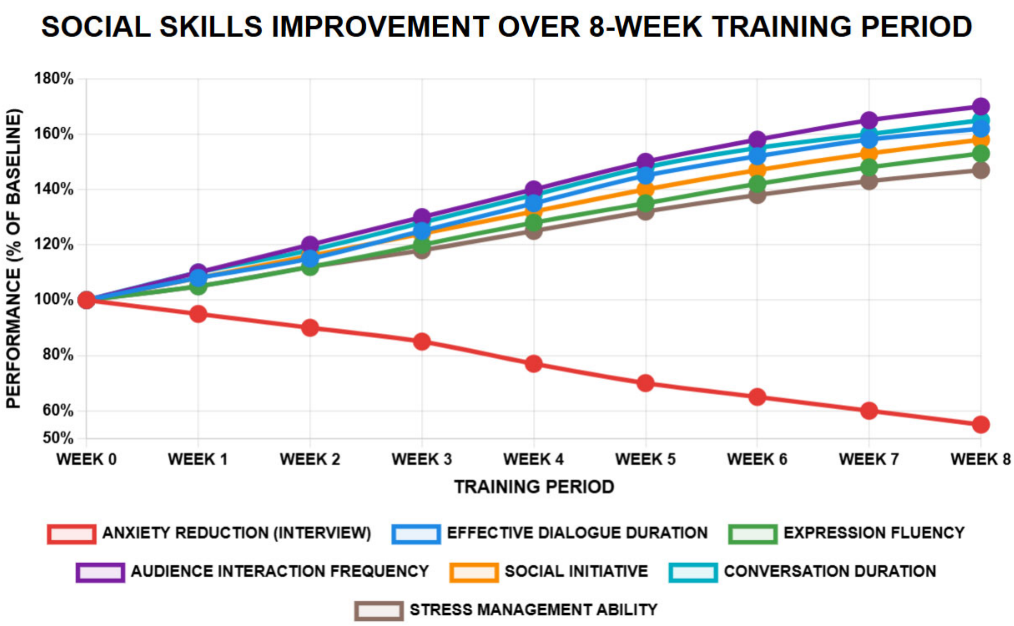

Figure 3. Trend of improvement in training effectiveness

ISSN: 2710-0588

4.4. User Satisfaction and Subjective Feedback Evaluation

According to the results of the questionnaire, 92% of the participants were satisfied with the overall performance of the system. The system received high ratings in the dimensions of ease of use, interaction naturalness, emotion perception and training effect. The system achieved a score of 4.5 out of 5 for ease of use and 4.3 for interaction naturalness. Participants particularly recognised the system's performance in terms of emotion perception and feedback, with a related score of 4.6. The survey revealed that 85% of the participants felt that the system helped them to overcome their social barriers, and 78% indicated that they would continue to use the system for training. The results of the in-depth interviews revealed that the emotional interactive experience in the virtual environment helped participants build stronger social self-confidence. Feedback data also indicated that the system's emotion recognition function provided participants with an effective emotion management tool that facilitated the overall development of their social skills[16-17]. The open-ended questions of the questionnaire collected a large number of constructive comments, and participants generally wished to add more types of scenarios to provide a more personalised training programme. Especially in high-stress scenarios such as workplace socialising and public speaking, participants expected the system to provide more detailed emotional guidance and coping strategies, and some participants suggested adding a group interaction function to support multiple people to conduct social training in the virtual environment at the same time.

5. Conclusion

In this study, an interactive system integrating multimodal emotion recognition technology and virtual reality was successfully designed and implemented, which has achieved remarkable results in the field of social training. The system achieves high-precision emotion perception and real-time response by integrating multidimensional data such as facial expressions, voice features, physiological signals and body movements. Experimental evaluations confirm that the system can effectively improve the social skills of participants, and most users report that the skills acquired in the training have been successfully transferred to real life, which fully verifies the practical value of the system in enhancing the social skills of users and improving the social performance of socially impaired people. The innovation of the system is not only reflected in the integration of technology, but also in the establishment of a closed-loop model of ‘emotion-interaction-feedback’, which provides a safe, controllable and efficient training environment for social disorder intervention.

Future research will focus on four key directions: deepening the research of emotion recognition algorithms, integrating deep learning and cognitive computing models to improve the recognition of complex emotional states, and optimising the algorithms to adapt to cross-cultural application scenarios; expanding the application areas, expanding the system technology to multiple scenarios such as education, healthcare, and corporate training, and developing special modules for emotional intelligence cultivation, treatment of psychological disorders, and teamwork; Strengthen long-term effect evaluation, establish a systematic tracking research mechanism to assess the durability of the training effect and the stability of skill migration; improve the ethical norms and privacy protection mechanism to ensure the safe use of emotion data, and promote the healthy development of emotion recognition virtual reality technology. Through these efforts, the system is expected to provide important support for the construction of a smarter and more humane human-computer interaction system, and ultimately promote the overall improvement of emotional intelligence and social well-being in human society.

ISSN: 2710-0588

References

[1] Chen X, Chen H. Emotion recognition using facial expressions in an immersive virtual reality application[J]. Virtual Reality, 2023, 27(3): 1717-1732.

[2] Mousavi S M H, Khaertdinov B, Jeuris P, et al. Emotion Recognition in Adaptive Virtual Reality Settings: Challenges and Opportunities[J]. WAMWB@ MobileHCI, 2023: 1-20.

[3] Chowdary M K, Nguyen T N, Hemanth D J. Deep learning-based facial emotion recognition for human–computer interaction applications[J]. Neural Computing and Applications, 2023, 35(32): 23311-23328.

[4] Papoutsi C, Drigas A, Skianis C. Virtual and augmented reality for developing emotional intelligence skills[J]. Int. J. Recent Contrib. Eng. Sci. IT (IJES), 2021, 9(3): 35-53.

[5] Balakrishnan S, Hameed M S S, Venkatesan K, et al. Interaction of spatial computing in augmented reality[C]//2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, 2021, 1: 1900-1904.

[6] Kaimara P, Oikonomou A, Deliyannis I. Could virtual reality applications pose real risks to children and adolescents? A systematic review of ethical issues and concerns[J]. Virtual reality, 2022, 26(2): 697-735.

[7] Stock-Homburg R. Survey of emotions in human–robot interactions: Perspectives from robotic psychology on 20 years of research[J]. International Journal of Social Robotics, 2022, 14(2): 389-411.

[8] Islam M R, Moni M A, Islam M M, et al. Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques[J]. IEEE Access, 2021, 9: 94601-94624.

[9] Katirai A. Ethical considerations in emotion recognition technologies: a review of the literature[J]. AI and Ethics, 2024, 4(4): 927-948.

[10] Li X, Zhang Y, Tiwari P, et al. EEG based emotion recognition: A tutorial and review[J]. ACM Computing Surveys, 2022, 55(4): 1-57.

[11] Ali W, Tian W, Din S U, et al. Classical and modern face recognition approaches: a complete review[J]. Multimedia tools and applications, 2021, 80: 4825-4880.

[12] Hangaragi S, Singh T. Face detection and Recognition using Face Mesh and deep neural network[J]. Procedia Computer Science, 2023, 218: 741-749.

[13] Zhao Z. Construction of Intelligent Perception and Operation Control Framework of Education Platform[J]. Journal of Computer Technology and Applied Mathematics, 2025, 2(3): 1-8.

[14] Cao Y, Cao P, Chen H, Kochendorfer K M, Trotter A B, Galanter W L, et al. Predicting ICU admissions for hospitalized COVID-19 patients with a factor graph-based model[A]. In: Multimodal AI in healthcare: A paradigm shift in health intelligence [C]. Cham: Springer International Publishing, 2022: 245–256.

[15] Zhang J, Xiang A, Cheng Y, et al. Research on Detection of Floating Objects in River and Lake Based on AI Image Recognition [J][J]. Journal of Artificial Intelligence Practice, 2024, 7(2): 97-106.

[16] Liu X, Wang F, Zeng H, et al. PRNet: A Priori Embedded Network for Real-World Blind Micro-Expression Recognition[J]. Mathematics, 2025, 13(5): 749.

[17] Wang Y. Construction of a Clinical Trial Data Anomaly Detection and Risk Warning System based on Knowledge Graph[C]//Forum on Research and Innovation Management. 2025, 3(6).