Design of the Intelligent Interview System

Bolin Chen[1], Ruian Yan[2]
[1]*School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, Henan, China*
[2]*iFLYTEK Co., Ltd., Hefei, Anhui, China*

Abstract: Intelligent interview systems have attracted increasing attention as an effective solution for improving the efficiency and objectivity of interview evaluation. This study presents the design and implementation of an intelligent interview system that integrates natural language processing, computer vision, and large language models to support automated interview interaction and assessment. The proposed system adopts a modular architecture and enables multimodal data acquisition, including text, speech, and facial video information. A large language model is utilized to generate interview questions and conduct interactive dialogue, while visual and speech features are extracted to analyze interviewee behavior and emotional states. By jointly analyzing multimodal information, the system provides objective feedback on interview performance. Experimental results indicate that the system supports stable real-time interaction and effective feature extraction, demonstrating its applicability to intelligent interview simulation scenarios.

Keywords: Intelligent Interview System; Large Language Model; Multimodal Interaction; Computer Vision

1. Introduction

Interviews play a critical role in personnel recruitment and talent evaluation, serving as a key mechanism for assessing candidates' professional competence, communication ability, and behavioral characteristics. However, conventional interview processes are typically conducted by human interviewers and depend largely on subjective judgment, which may lead to inconsistency, bias, and limited scalability [1,2]. As recruitment demands continue to grow, there is an increasing need for automated and intelligent interview solutions that can provide standardized and efficient evaluation. In recent years, advances in artificial intelligence have significantly promoted the development of intelligent human–computer interaction systems. Techniques such as deep learning, speech recognition, and computer vision have enabled machines to perceive and analyze multimodal information with improved accuracy. At the same time, large language models have demonstrated strong capabilities in natural language understanding and generation, making it possible to conduct coherent and context-aware dialogue. These developments create favorable conditions for building intelligent interview systems that simulate real interview scenarios while reducing human labor costs.

To address these challenges, this paper proposes an intelligent interview system that combines large language models with multimodal perception technologies [3]. The system is designed to support interactive interview dialogue, real-time data acquisition, and automated performance evaluation. By integrating text interaction, speech analysis, and facial behavior recognition within a unified framework, the proposed system aims to enhance interview realism and evaluation objectivity. The main contributions of this study can be summarized as follows: (1) the design of a modular intelligent interview system architecture based on multimodal data fusion; (2) the integration of a large language model to enable adaptive interview questioning and interaction; and (3) the implementation and evaluation of key functional modules in a practical interview application environment.

The remainder of this paper is organized as follows. Section II reviews related work on intelligent interview systems and multimodal analysis. Section III introduces the overall system architecture. Section IV describes the key technologies and implementation details. Section V presents experimental results and analysis. Finally, Section VI concludes the paper and discusses future work.

2. System Architecture

The system is developed and maintained based on a browser–server (B/S) architecture. The back-end services are implemented using the Spring framework, while data storage is supported by the DM database [4]. The front end is designed and implemented with the Vue framework to provide interactive user interfaces. Overall, the system is organized into four logical layers, namely the presentation layer, business logic layer, technical layer, and data layer. The overall system architecture is illustrated in Figure 1.

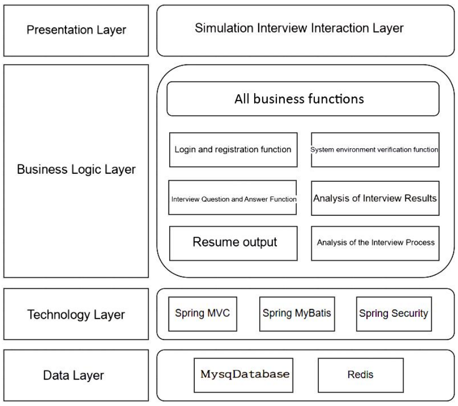

Figure 1. Overall System Design

The interview simulation system operates using a layered architectural model. The presentation layer serves as the primary interface for user interaction and provides interactive pages for functions such as user authentication, interview question answering, and résumé submission. This layer collects user inputs, including account credentials and response content, and forwards them to the business logic layer for processing. At the same time, interview scores and evaluation feedback are visually presented to users through the presentation layer.

The business logic layer functions as the central control unit of the system. Upon receiving requests from the presentation layer, it coordinates operations across the technical layer and the data layer. For user authentication, Spring Security is invoked to verify user identity and validate account information against the user database. During interview question processing, the business logic layer retrieves relevant questions from the data layer and performs real-time analysis of response validity. The service layer exposes RESTful interfaces through Spring Boot, which dispatch client requests to the business logic layer for execution. Business components are assembled and managed using the Spring framework to support flexible and efficient processing.

The technical layer handles HTTP request–response processing and data transmission through the Spring MVC framework [5]. Database interactions are streamlined using Spring MyBatis, while system security is ensured by Spring Security, which provides authentication and access control mechanisms.

The data layer is responsible for storing structured information, including user profiles and interview questions, in the DM database. To improve system efficiency, Redis is employed to cache frequently accessed data. Database access is optimized through MyBatis, which supports create, read, update, and delete (CRUD) operations as well as object–relational mapping. Domain objects are derived from database tables to facilitate ORM-based data management. In addition, the system integrates the iFLYTEK Spark large language model through configuration files. With well-defined interactions among layers, the architecture ensures functional completeness while maintaining operational efficiency and system security.

3. Technical Implementation

3.1 Description of Core Functional Module Design

Considering functional requirements, system stability, and future scalability, a modular design strategy is adopted for system implementation. The interview simulation system is divided into four major functional modules: the comprehensive business management module, the online question-and-answer module, the résumé generation module, and the job recommendation module. The overall functional structure of the system is illustrated in Figure 2.

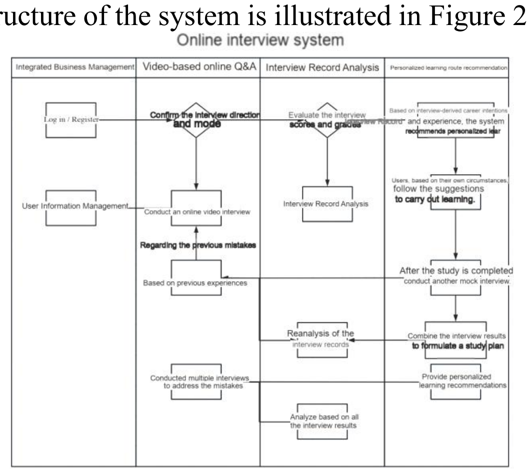

Figure 2. Overall Functional Diagram of the Simulated Interview System

The interview simulation subsystem serves as the core component of the entire platform and acts as the primary entry point for all interview-related operations. Through this module, users access the system and initiate the complete workflow of the simulated interview process.

The comprehensive business management module serves as the foundational administrative component of the system and covers the essential operations required for users to interact with the platform.

Login and Registration: This function enables users to complete account registration and authentication, which constitutes the initial step for obtaining system access. During registration, users are required to provide relevant personal information, while the login process verifies user identity to ensure that only authorized users can access the system. User Information Management: This function is designed to manage user profile data, including name, gender, age, and contact information. Users can view and update their personal information at any time, allowing the system to deliver more personalized services based on the maintained user profiles.

The video-based online question-and-answer module is primarily designed to simulate the interview interaction process and assist users in improving their interview skills and response capabilities.

Environment Testing: Before initiating the interview session, users can verify the operational status of their network connection, microphone input, and camera display through the environment testing function. This step is essential to ensure that the interview process is not disrupted by hardware or connectivity issues. By identifying and resolving potential problems in advance, users are able to enter the interview in a stable and prepared state.

Voice-Based Interaction: Through this function, users engage in voice-based communication with the system, which simulates an interviewer by presenting various interview questions. Users respond verbally via a microphone, enabling a highly realistic interview experience. This interaction mode helps users enhance their spoken expression, real-time reasoning, and ability to respond effectively under interview conditions.

Answer Editing: The answer editing function allows users to revise and refine their responses during the interview process. It can be used to correct omissions or improve unclear expressions after an initial response. In addition, when speech recognition encounters limitations, users may manually input their answers, providing an alternative response method to ensure continuity of the interview process.

To further enhance the relevance and practical value of interview questions, the system incorporates an intelligent question generation mechanism based on specific job orientations. The corresponding workflow is illustrated in Figure 3.

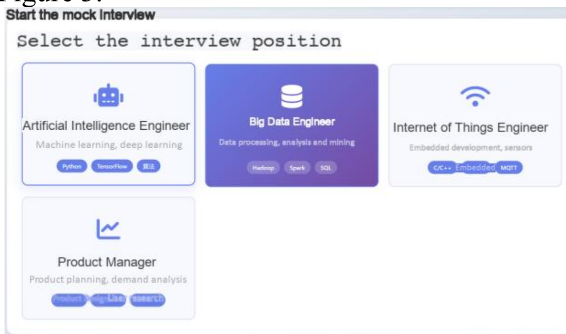

Figure 3. Demonstration of Intelligent Question Generation Based on Job Direction

During the interview preparation stage, users can select a target job category, such as software development or marketing, through a drop-down menu. Based on multi-industry question banks and supported by natural language processing and machine learning techniques, the system automatically generates mandatory interview questions that align with the core skill requirements of the selected position.

Users are able to manage and review their interview records, while the system performs personalized and fine-grained analysis of interview performance by integrating information such as educational background, internship experience, project involvement, and professional certifications. Based on this analysis, the personalized learning path recommendation module provides targeted suggestions to optimize subsequent interview performance. In addition, the system allows users to edit résumé components, including education, work experience, and skill descriptions, in order to highlight individual strengths. As user profiles are updated, the recommendation results are dynamically adjusted to better align with career objectives and evolving labor market demands.

In addition, the system analyzes users' simulated interview performance, demonstrated skills, and résumé information through algorithmic methods to recommend job positions with a high degree of compatibility. This approach is intended to improve job search efficiency and increase the likelihood of successful employment outcomes.

3.2 Front-End Design and Implementation

(1) Implementation of the login interface:

When users log in, the system first verifies their account and password. Upon successful authentication, permissions are determined based on user roles: administrators have full management rights over user information, job skills, question databases, and online interactions, while regular users can only participate in online interview Q&A, engage in online communication, and view or edit their personal information. The login process involves connecting to the network, accessing the login page, entering account credentials for verification, and, if successful, gaining access to the system; if authentication fails, users should check the network or contact the administrator to activate their permissions. This process is illustrated in Figure 4.

Figure 4. Overall Functional Diagram of the Simulated Interview System

The user login interface and its implementation are illustrated in Figure 5.

Figure 5. User Login Interface Implementation Implementation of the Online Q&A Module

(2) Implementation of the Online Q&A module

The online Q&A module consists of three components: system environment verification, online interview Q&A, and simulated interview result analysis. Before entering the interview, users first complete an environment check. They then engage in one-on-one interactive Q&A with an AI interviewer based on a contextual encoding mechanism. The system can proactively and dynamically generate questions according to the resume and job requirements while recording responses. It also performs semantic analysis of the candidate's questions and matches answers, generating optimal feedback from the knowledge base to efficiently assess professional competence. The implementation flow of the system responses is shown in Figure 6.

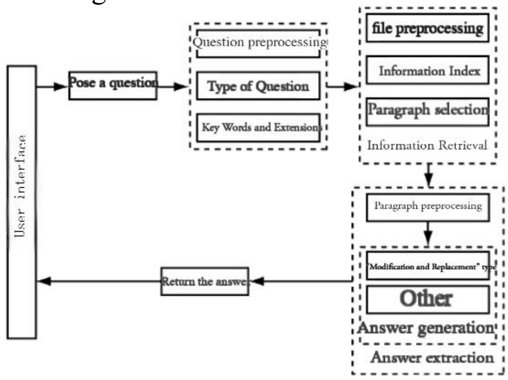

Figure 6. Implementation Flowchart of the Online Q&A Module

The system features a comprehensive interview process control mechanism, allowing flexible configuration from question generation and response timing to evaluation feedback, ensuring a standardized and efficient workflow. Additionally, the platform supports real-time voice and video interactions, integrating speech recognition and emotion analysis to achieve natural, smooth human–machine dialogue and instant feedback. For text input, technologies such as federated learning and Q&A retrieval are employed to provide efficient and accurate interview responses. The system can intelligently analyze candidates' answers, perform semantic matching and scoring, and deliver feedback in either text or voice form. It also supports real-time interaction and assessment by interviewers, enhancing both interview efficiency and objectivity. The workflow is illustrated in Figure 7.

After the interview, the system leverages high-performance artificial intelligence algorithms to automatically perform an in-depth comparison between the user's responses and the standard answers, generating precise scores and detailed interview evaluations. This process is illustrated in Figures 8 and 9.

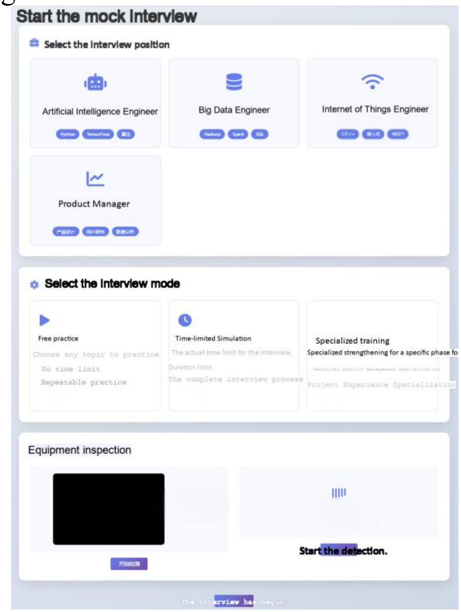

Figure 7. Interview Process Control Mechanism

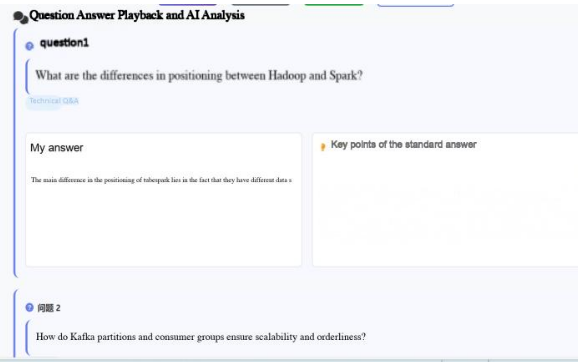

Figure 8. Score Statistics Page

Figure 9. Interview Evaluation Report

The system provides a comprehensive assessment across multiple dimensions, including professional knowledge, communication skills, and logical reasoning ability. In the evaluation report, results are presented under three sections: "Strengths", "Weaknesses" and "Improvement Suggestions". It highlights key merits in the user's responses, accurately identifies existing problems and shortcomings, and offers practical strategies for improvement, enabling users to gain a thorough understanding of their interview performance and clearly identify directions for future development[6].The system response flow for this feature is shown in Figure 10.

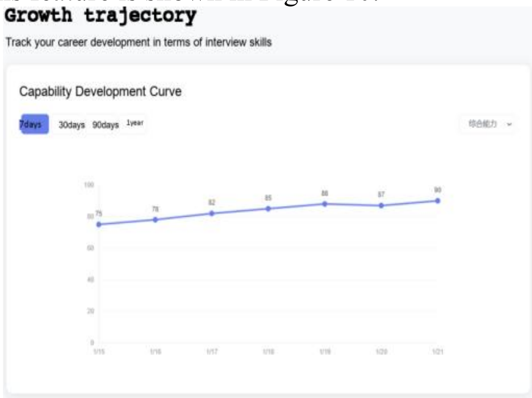

Figure 10. Interview Content Records

The interview analysis module can accurately dissect users' responses, promptly identifying issues such as vague content, insufficient detail, low confidence, or inadequate preparation. It provides clear feedback and targeted recommendations regarding expression, content, and attitude, helping users recognize their shortcomings, optimize response strategies and mindset, enhance professionalism and persuasiveness, and continuously improve interview skills, thereby boosting overall performance and success rates [7].

(3) Learning recommendations

Once the interview records are successfully generated based on the Q&A interactions, the system analyzes the records and provides personalized learning recommendations. This module has a relatively straightforward task, and its specific workflow is shown in Figure 11.

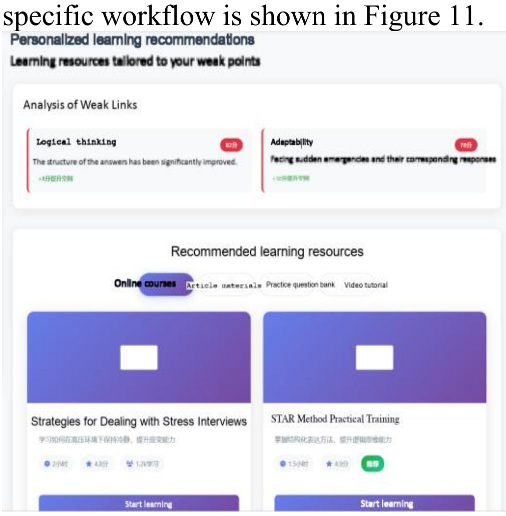

Figure 11. Personalized Learning Recommendations

Analyzing interview records holds significant value: the system can extract candidates' strengths through in-depth examination of interview content, identify shortcomings in experience, skills, and communication, and help users clarify their core competencies and areas for improvement. Based on this analysis, users can optimize their resumes, adjust job-seeking strategies, and enhance relevant abilities [8]. Furthermore, interview analysis results serve as a key basis for job recommendations, improving position matching and job search efficiency. Finally, this chapter systematically presents the overall implementation of the simulated interview system by detailing the processes and outcomes of the integrated business, online Q&A, interview record management, and recommendation modules.

(4) User management module

The platform supports user registration, login, and multi-role access control to ensure data security and operational compliance. Users can manage their resumes online, editing and updating them at any time, based on which the system generates corresponding interview content. Additionally, the platform automatically records the results of each interview session, allowing for historical review and performance tracking, thereby helping users achieve continuous improvement and targeted skill enhancement. The detailed outcomes are shown in Figure 12.

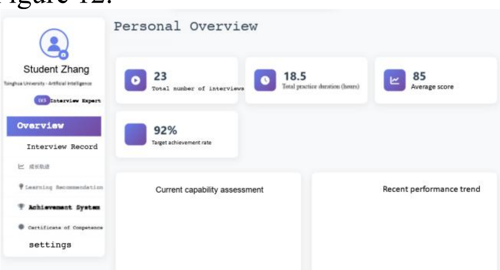

Figure 12. User Management Module Diagram

In this section, basic data is presented, including numerical values such as "23", "18.5" and "85" as well as percentage data like "92%". These correspond to different statistical dimensions, such as total number of interviews, total practice duration (in hours), and average interview score, providing a multifaceted overview of the user's performance and related metrics.

It also includes a capabilities and trends module, featuring sections for "Current Ability Analysis" and "Recent Performance Trends." This module aims to analyze the user's present skill level while illustrating changes in their recent performance in relevant areas, helping users understand their current capabilities and development trajectory.

3.3 Implementation of the Computer Vision Module

(1) Video Acquisition Layer: Video capture is implemented using FFmpegFrameGrabber based on JavaCV, enabling stable acquisition of high-quality video streams at 1080P and 30 fps. It is compatible with mainstream cameras and video conferencing interfaces, providing a reliable data foundation for subsequent analysis.

(2) Face Detection Layer: The MTCNN algorithm is employed for face detection and alignment, using a cascaded network to balance high accuracy with relatively fast processing. It is robust to variations in lighting and occlusions, supports simultaneous detection of multiple faces, and provides precise regions of interest (ROIs) for subsequent processing.

(3) Feature Extraction Layer: The FaceNet model is used to map faces into 128-dimensional feature vectors, where the distance between vectors of the same person is small and that of different individuals is large. This provides a foundation for identity verification and other subsequent recognition and extension functionalities.

(4) Expression Analysis Layer:A CNN-based expression classification model, trained on the FER-2013 dataset, is used to recognize seven basic facial expressions in real time. The results serve as a key basis for evaluating the candidate's emotional state and on-the-spot performance.

3.4 Integration of Large Language Models

To enable intelligent Q&A and scoring functionalities, the system requires integration of high-performance large language models (LLMs). This project conducted a comprehensive comparison of several mainstream LLM solutions, evaluating them across four key dimensions: model scale, deployment method, response performance, and output quality.

(1) LLaMA 2-7B (Meta): This model has 7 billion parameters and supports local, private deployment, offering strong controllability and data security. Its response time can be maintained under 500 milliseconds, with balanced performance in general knowledge

Q&A, logical reasoning, and code generation, achieving an overall four-star capability rating. It is suitable for scenarios that require high-quality Q&A and have access to local GPU inference resources [9].

(2) ChatGLM-6B (Zhipu AI): With approximately 6 billion parameters, this lightweight bilingual model's main advantage lies in its support for local deployment and faster response times (under 300 milliseconds), while requiring relatively modest hardware resources. It performs well in Chinese contexts, delivering Q&A quality comparable to LLaMA 2-7B, though its English proficiency is somewhat weaker. It is particularly suitable for medium-scale concurrent applications where low latency is critical for real-time interaction [10].

(3) Tongyi Qianwen API (Alibaba Cloud): This solution provides services via cloud API, with the model's exact parameter count not publicly disclosed. Its main advantage lies in exceptionally high Q&A quality (five-star), leveraging large-scale cloud model clusters to achieve superior knowledge breadth, comprehension depth, and generation fluency. The response time is approximately 800 milliseconds, but it depends on network conditions and carries potential data sovereignty risks. It is suitable for projects where Q&A quality is the primary concern, local deployment is unnecessary, and network connectivity is reliable.

Considering the above factors, Tongyi Qianwen was selected as the large language model for this project.

3.5 Key Technical Challenges and Limitations

This study holds both theoretical and practical value in exploring the impact of anthropomorphic design features on organizational attractiveness in AI interview contexts; however, several technical and research limitations remain:

First, scenario-based experiments cannot fully replicate the complex psychological states present in real interviews. Current technologies still fall short in high-fidelity emotion elicitation and multimodal emotion perception. Future work should integrate more realistic emotion induction models along with physiological and behavioral data collection methods to enhance ecological validity.

Second, due to the current application of AI interviews primarily in campus recruitment, the sample is concentrated on student populations. The model's generalizability to social recruitment scenarios requires improvement through broader sample coverage and inclusion of data across different career stages.

Third, in terms of research content, this study only focused on external appearance anthropomorphism and role-based anthropomorphism. Future research could expand to additional anthropomorphic and technological features, investigate the effects of anthropomorphic design on interview-related emotions such as anxiety, and examine how individual differences among candidates (e.g., personality types) may moderate the model's applicability boundaries.

4. Conclusion

This study successfully developed a Java-based intelligent interview system covering the full process of "interview initiation—multimodal interaction—real-time evaluation—result feedback". The system supports collaborative interaction across text, voice, and video modalities, integrating ASR and computer vision technologies for face detection and expression recognition, thereby establishing a multidimensional interview assessment framework. The user interface follows a goal-oriented design, with core operations requiring no more than three steps; on average, fresh graduates can complete the full interview process within eight minutes, demonstrating strong usability and contextual adaptability. To address deployment challenges of large language models on RISC-V edge hardware, a lightweight solution combining quantization, distillation, and hardware adaptation was proposed, enabling stable operation of the LLaMA 2-7B model on development boards, with single-pass inference under 500 ms and memory usage controlled within 6 GB, meeting real-time interaction requirements. The system further constructs a multimodal evaluation framework integrating textual semantics, voice features, and visual emotion, achieving a comprehensive assessment accuracy of 85.7%, significantly higher than traditional single-dimension methods. Standardized testing confirms that the system meets design targets for response speed, resource consumption, concurrency, and stability, supporting scalable applications such as SME recruitment and university career guidance, thereby demonstrating high practical deployment value.

References

[1] Ni Ying, Wang Shihan, Sun Jian, et al. Simulation and modeling of complex interactive behaviors at intersections based on multi-agent imitation learning. Journal of Tongji University (Natural Science Edition), 2025, 53(9): 1391–1402.
[2] Li Ming, Wang Hua, Zhao Qiang. Research progress on interview evaluation systems based on deep learning. Chinese Journal of Computers, 2023, 46(10): 2145–2158.
[3] Zhang Yu, Liu Hang, Wang Jun. A survey of artificial intelligence applications in recruitment and interview assessment. Journal of Software, 2023, 34(8): 3721–3736.
[4] Zhou Chen, Li Tao, Chen Si. Research on interview behavior analysis methods based on multimodal data fusion. Computer Engineering and Applications, 2024, 60(12): 45–52.
[5] Wang Lei, Zhao Xin, Sun Kai. Intelligent interview evaluation model integrating speech and visual features. Pattern Recognition and Artificial Intelligence, 2024, 37(6): 561–570.
[6] Liu Yang, Chen Jie, Huang Rui. Research on multimodal emotion recognition methods for human–computer interaction. Computer Research and Development, 2025, 62(3): 589–601.
[7] Brown T, Mann B, Ryder N. Language Models are Few-Shot Learners//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: NeurIPS, 2020: 1877-1901.
[8] He K, Zhang X, Ren S. Deep Residual Learning for Image Recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: CVPR, 2016: 770-778.
[9] RISC-V Foundation. The RISC-V Instruction Set Manual. San Jose: RISC-V International, 2022.
[10]Bradski G, Kaehler A, Pisarevsky V. OpenCV 4.8 Computer Vision Library Reference Manual. Beijing: OpenCV Organization, 2023.