Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 19 No. 20 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijim.v19i20.58429

PAPER

VR/AR-Based Mobile Interaction for Virtual Simulation Training in Higher Education

Xiaoxia Tian[1],
Yanjin Wang[2] (✉)

[1]Hebei Vocational College of Resources and Environmental Technology, Shijiazhuang, China

[2]Hengshui Open University, Hengshui, China

wangyanjinhs@126.com

ABSTRACT

In the context of practical instruction in higher education, traditional simulation training is constrained by factors such as spatial limitations, equipment availability, safety concerns, and high costs, making it difficult to meet the demands for hands-on experience. Mobile interaction technology based on virtual reality (VR) and augmented reality (AR), characterized by immersion, strong interactivity, and portability, provides an innovative pathway for virtual simulation training, emerging as a promising solution to the limitations of conventional models. Although the application of VR/AR in virtual simulation has been preliminarily explored, current approaches exhibit significant limitations: interaction modalities are predominantly restricted to unimodal gestures or voice commands, lacking coordinated multimodal integration; system feedback is often reliant on predefined rules, limiting the ability to deliver adaptive and precise guidance; and model optimization fails to adequately leverage prior feedback, resulting in suboptimal learning efficiency and adaptability. To address these challenges, a reinforcement learning framework for VR/AR-based mobile interaction in virtual simulation training was developed in this study. This framework includes the design of a multimodal feature extraction method combining voice and gesture inputs, the construction of an auxiliary decision-making and feedback mechanism, the formulation of principles for reward function design, and the proposal of a model optimization strategy informed by prior feedback data. The core innovations of this study are threefold. First, multimodal interaction feature extraction—integrating both speech and gesture inputs—was implemented to overcome the limitations of unimodal interaction and to enhance interaction naturalness. Second, a dynamic feedback mechanism based on real-time operational data was established, replacing traditional rule-based feedback systems to improve instructional precision. Third, prior feedback information was embedded within the model optimization loop to accelerate model iteration and enhance adaptability across diverse training scenarios. This study provides technological support for improving the quality of virtual simulation training in higher education and offers novel insights into the integration of VR/AR technology with educational practice.

KEYWORDS

virtual reality (VR)/augmented reality (AR) technology, higher education, virtual simulation training, reinforcement learning framework, multimodal interaction, dynamic feedback

Tian, X., Wang, Y. (2025). VR/AR-Based Mobile Interaction for Virtual Simulation Training in Higher Education. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(20), pp. 168–182. https://doi.org/10.3991/ijim.v19i20.58429

Article submitted 2025-06-13. Revision uploaded 2025-07-28. Final acceptance 2025-08-01.

1 INTRODUCTION

As the emphasis on practical instruction in higher education continues to increase [1–4], the limitations of traditional training models—particularly in terms of spatial constraints, equipment availability, safety risks, and high operational costs—have become increasingly pronounced. Practical training in disciplines such as medicine, engineering, and aviation typically requires expensive equipment and specialized environments [5, 6] and often involves a high degree of operational risk [7, 8], making it difficult to provide students with sufficient hands-on opportunities. Virtual reality (VR)/augmented reality (AR)-based mobile interaction technology [9, 10], characterized by its immersive experience, high interactivity, and portability, offers a promising alternative to these challenges. By simulating realistic training environments, this technology enables students to engage in repeated practice within a virtual setting, thereby transcending the limitations of time and space. As a result, it has attracted growing attention and has begun to see preliminary adoption in the domain of virtual simulation training in higher education.

The investigation into the application of VR/AR-based mobile interaction technology in higher education virtual simulation training holds substantial theoretical and practical significance. From an educational perspective, this technology offers a novel paradigm for virtual simulation training [11, 12], improving both the effectiveness and engagement of practical instruction. It allows students to gain operational experience in a safe and accessible virtual environment, enhancing their hands-on abilities and problem-solving skills, thereby contributing to improved instructional quality in higher education. However, limitations in current research methodologies remain evident. Many existing studies have relied on unimodal interaction approaches, focusing primarily on gesture- or voice-based interaction in isolation [13, 14], with limited attention to the integration of multimodal interaction. For example, some studies [15, 16] have concentrated solely on gesture-based interaction in virtual simulation scenarios while overlooking the synergistic potential of voice and other input modalities. In terms of feedback mechanisms, the responses provided by some systems have been neither timely nor precise, failing to offer targeted guidance based on real-time user operations. As demonstrated in the study by Kim and Lee [17], system feedback was predominantly based on preset rules and lacked dynamic analysis of real-time operational data. Additionally, model optimization efforts have largely neglected the role of prior feedback, resulting in limited adaptability and suboptimal learning efficiency.

The present study focuses on the development of a reinforcement learning framework for VR/AR-based mobile interaction in virtual simulation training. The framework is composed of several key components. First, a multimodal feature extraction method was introduced, incorporating both voice and gesture inputs to enable more natural and efficient human-computer interaction. Second, the principles of system-assisted decision-making and auxiliary feedback were detailed, allowing intelligent decisions to be made based on students' operational behavior while enabling real-time feedback to be delivered adaptively. Third, the design of the reward function was outlined, with the objective of guiding the system through rational reward mechanisms to support correct simulation-based instruction. Finally, a model optimization strategy based on prior feedback was proposed, leveraging historical training data to continuously refine system performance. The significance of this study lies in its ability to address existing limitations in interaction modalities, feedback mechanisms, and model optimization within virtual simulation training.

By constructing a reinforcement learning framework, the overall interaction experience and instructional effectiveness of VR/AR-based virtual simulation were substantially enhanced. This work not only provides technical support and practical strategies for the advancement of virtual simulation training in higher education but also offers valuable reference for the broader integration of VR/AR-based mobile interaction technology into the field of education.

2 REINFORCEMENT LEARNING FRAMEWORK FOR VR/AR-BASED MOBILE INTERACTION IN VIRTUAL SIMULATION TRAINING

In virtual simulation training within higher education, significant variability has been observed in students' knowledge bases, operational proficiency, and learning habits. Some students respond more efficiently to voice commands, while others perform more effectively through gesture-based operations. Novice learners require fundamental and detailed instructional support, whereas more advanced students benefit from targeted and progressive guidance. Existing VR/AR simulation tools generally employ fixed interaction patterns and uniform assistance strategies, which are insufficient for accommodating individual interaction preferences or adjusting the intensity of instructional support based on user performance. These limitations have resulted in fragmented interaction experiences and restricted training efficiency. To address these challenges, a reinforcement learning framework for VR/AR-based mobile interaction in virtual simulation training was proposed in this study. This framework enables the dynamic adaptation of interaction strategies through reinforcement learning, ensures active student engagement and feedback via the mobile interaction module, and ultimately resolves the core issues of inflexible interaction, imprecise guidance, and the absence of a closed-loop optimization mechanism.

2.1 Multimodal feature extraction

To capture and interpret student input behavior in VR/AR-based mobile interaction environments for virtual simulation training, two sensory modalities—voice and gesture—were employed in this study. For voice input, variations in accent, ambiguity of commands, and background noise commonly encountered in higher education training scenarios were considered. A three-stage processing pipeline was established, comprising noise reduction preprocessing, context-aware semantic parsing, and dynamic error correction. Initially, environmental noise was filtered using an adaptive noise suppression algorithm to preserve effective speech signals. Subsequently, a semantic model was constructed based on a domain-specific vocabulary relevant to simulation tasks, enabling the mapping of ambiguous instructions to standardized operational commands. Finally, real-time interaction data from the reinforcement learning framework was referenced. When inconsistencies between recognized commands and subsequent student actions are detected, a dynamic correction mechanism is triggered. This mechanism adjusts recognition outputs by incorporating historical interaction preferences. For gesture input, the non-standard and self-directed nature of student actions necessitates an approach that supports interpretive flexibility. A method combining key motion extraction with contextual scenario fusion was adopted. Core gesture features were identified through skeletal keypoint tracking, and these features were fused with task-specific attributes to construct a scenario-aware gesture classification model. Feedback data from the reinforcement learning framework was concurrently integrated. When gesture recognition fails to align with the intended operation, the recognition threshold is dynamically adjusted. This adaptive tolerance mechanism allows for acceptable degrees of motion deviation, thereby preserving the naturalness of the interaction and avoiding disruption from overly rigid standardization.

2.2 System-assisted decision-making

In higher education virtual simulation scenarios, the proposed framework initially captures students' interactive behaviors and latent instructional needs in real time through multimodal sensing channels, including voice recognition and gesture recognition. These inputs are combined with sub-goal reward mechanisms and student state data to construct a multidimensional decision-making basis. When instructional assistance is required, the system analyzes the multimodal information using the reinforcement learning model. For example, if a student exhibits hesitant gestures and simultaneously inquires about a component's function via voice, a component explanation prompt is prioritized. If two consecutive operational errors are detected without any accompanying voice feedback, the system automatically triggers a demonstration of the corresponding procedural steps. The framework strictly adheres to a "student feedback-driven strategy update" rule: when assistance is accepted, the corresponding strategy is marked as effective and incorporated into future decision references; when assistance is explicitly rejected, the system immediately adjusts the strategy based on current interaction data; in the absence of explicit feedback, the original strategy is maintained while operational data continue to be collected to validate effectiveness. This ensures that system-assisted decision-making remains dynamically aligned with students' real-time needs. The following expression formalizes the system-assisted decision-making process:

$$x = \begin{cases} \varphi, IF \quad x_j = -1 \\ x_j, IF \quad x_j \geq 0 \\ x_\tau, IF \quad x_j = \varphi \end{cases} \tag{1}$$

In this formulation, the system's assistance behavior x is determined by the student's feedback x_j and the system's recommended action x_τ. If the student's feedback x_j indicates rejection (i.e., a negative value), no assistance action is taken. If the feedback is non-negative, indicating acceptance of assistance, the corresponding prompt action x_j is executed. In cases where no explicit feedback is provided—such as student silence or neutrality—the recommended action x_τ is executed by default.

This assistance decision-making mechanism offers significant advantages in virtual simulation training for higher education. On one hand, the integration of multimodal inputs effectively addresses the limitations of unimodal interaction and enhances the precision of assistance decisions. For example, in medical virtual simulation training, a student may simulate a needle-holding gesture while simultaneously voicing the phrase "suture technique." By fusing both gesture and voice input, the system can accurately infer that the student's need pertains to "suture angle guidance" rather than "instrument selection recommendation," thereby preventing ineffective assistance caused by incomplete input information.

2.3 System-assisted feedback

The core principle underlying the system-assisted feedback mechanism is a co-driven model that integrates student feedback and task-based reward functions. In higher education virtual simulation training, once the system executes an assistance behavior—such as providing prompts or recommendations—student responses are classified into three types: "positive," "negative," and "neutral." These responses directly influence the feedback function $d_j(t, x)$. Specifically, a "positive" label yields a positive output value, such as $d_j(t, x) = 1$; a "negative" label results in a negative output, such as $d_j(t, x) = -1$; and in the absence of explicit feedback, the response defaults to "neutral," such as $d_j(t, x) = 0$. Simultaneously, the system tightly integrates student feedback with a task reward mechanism. A sub-goal reward e^s_{th} is designed to provide targeted reinforcement for discrete, critical states within the virtual simulation task, ensuring that the system remains focused on key instructional components. In parallel, a composite reward function $e(t, x)$ is defined by combining the simulation process reward $e_{MDP}(t, x)$ with the sub-goal reward $\lambda e_{th}(t, x)$, thereby establishing a comprehensive evaluation criterion. The feedback function $d_j(t, x)$ is expressed as:

$$d(t,x) = \begin{cases} 1, positive \\ -1, negative \\ 0, neutral \end{cases} \tag{2}$$

The sub-goal reward e^s_{th} is defined as:

$$e^u_{th}(t) = \begin{cases} 1, IF\ t == t_{th} \\ 0, ELSE \end{cases} \tag{3}$$

Letting λ denote the weighting parameter, the composite reward function $e(t, x)$ is defined as:

$$e(t,x) = e_{MDP}(t,x) + \lambda e_{th}(t,x) \tag{4}$$

In the context of interactive learning, the system guides the transition to the next simulation state t^{s+1} through an intervention in the form of an assistance behavior. Based on the current state t and the executed assistance behavior x, a composite reward can be obtained. The policy is then optimized with the objective of maximizing long-term reward, whereby both feedback and reward jointly serve as "navigation signals" for iterative policy updates. Assuming the state modified by student intervention is denoted by t_j, this process can be formally represented as:

$$(t)^{s+1} = \begin{cases} (t')^t, IF\ \ t_j == \theta \\ t_j, ELSE \end{cases} \tag{5}$$

The principal advantage of this feedback mechanism lies in its capacity to accurately capture and dynamically respond to individual student needs, thereby achieving high adaptability to the personalized requirements of higher education virtual simulation. For instance, in a medical simulation scenario, if a system-generated prompt on "venipuncture angle guidance" receives a "positive" response from the student, the corresponding positive output from the feedback function reinforces the priority of this assistance strategy in similar contexts. Conversely, if a recommendation on "surgical instrument selection" receives a "negative" response, the resulting negative feedback prompts the system to revise its strategy—such as switching to a prompt on "instrument function explanation." In cases where "neutral" feedback is recorded, the system avoids premature misclassification while retaining the potential for post hoc validation through operational outcomes. For example, if no response is received for a prompt on "chemical experiment procedural guidance," the system evaluates whether the student's subsequent actions align with defined sub-goals to indirectly assess the effectiveness of the original prompt. This design enhances the robustness of the feedback mechanism and ensures that preferences across diverse students can be reliably identified. In doing so, assistance strategies are continuously aligned with evolving student needs.

2.4 Reward function design

In the VR/AR-based mobile interaction reinforcement learning framework for virtual simulation training, the principle of reward function design is illustrated in Figure 1. The primary consideration in reward function design is the reception and processing of multimodal input signals, which must first be transformed into fuzzy representations. In higher education virtual simulation environments, student-generated interaction signals often exhibit uncertainty. For instance, during mechanical simulation training, the gesture for "gear installation" may deviate in amplitude due to a student's unfamiliarity with the operation, while a verbal instruction such as "tighten the screw" may be pronounced ambiguously under pressure. Directly applying such gesture-angle data or speech waveforms in reward computation can lead to recognition errors. Therefore, these signals need to be converted into fuzzy signals—for example, gesture accuracy may be categorized as "highly accurate," "moderately accurate," or "vague," and speech clarity as "clear" or "unclear." This transformation reduces the impact of signal noise and unifies heterogeneous signal types into a common representational form, thereby providing a robust foundation for subsequent rule application. Let the input multimodal variable be denoted as a, the fuzzy logic set as U, and the corresponding parameters as x and y. The process of obtaining the fuzzy set is expressed as:

$$U_u(a) = \frac{1}{x}\left(-\left|a - y_u\right| + x_u\right) \vee 0 \tag{6}$$

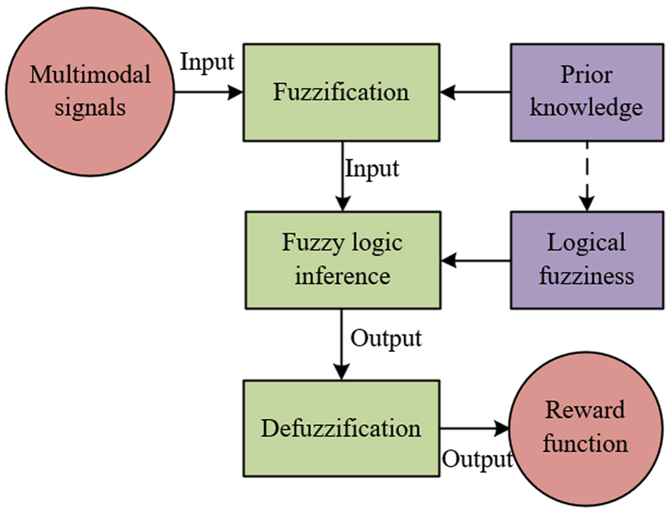

Fig. 1. Principle of reward function design in the reinforcement learning framework for VR/AR-based mobile interaction in virtual simulation training

Subsequently, prior knowledge was employed to formulate fuzzy logic rules for the fuzzified input data. In virtual simulation training for higher education, aligning with instructional objectives is essential. Prior knowledge serves as a distilled representation of instructional experience and procedural standards. For example, in a medical simulation scenario, instructional principles such as "a venipuncture angle of approximately 30° is optimal" and "excessive needle insertion speed increases the risk of simulated vascular damage" can be translated into rules: if "gesture angle fuzzy value is optimal" and "insertion speed fuzzy value is moderate," then the inferred operation quality is "high." These rules ensure that the reward logic remains consistent with pedagogical requirements while covering common procedural contexts encountered during simulation training.

Next, sensor data was transformed into fuzzy variables through fuzzy logic rules. VR/AR device sensors continuously capture key real-time data within the simulation environment, such as the spatial coordinates of student gestures, the force applied to virtual instruments, and the response time to voice commands. These raw data must be converted into fuzzy variables suitable for inference. In architectural surveying virtual simulation training, for instance, gesture stability data associated with "total station operation" can be converted into the fuzzy variable "operational stability," with values such as "stable," "moderately stable," or "unstable." Similarly, voice response latency during "coordinate input" can be transformed into the fuzzy variable "operational efficiency," with possible values such as "efficient," "moderate," or "inefficient." This transformation abstracts concrete sensor inputs into evaluation dimensions directly linked to instructional objectives, making data processing more focused on core teaching indicators such as "operational quality" and "efficiency." Let the designed fuzzy logic rules be denoted by $E^d = I^{\tau}_{k=1}E^d_k$, where the k-th rule is expressed as $E^d_k = U^1_k(a_1) \wedge U^2_k(a_2) \wedge P_k(b)$. For a single-point input, let the merged output be represented by $n_0(b)$, the weighting coefficient by x_n, and the response or activation levels of different modalities by $n_1^{\left(a_1^0\right)}$ and $n_2^{\left(a_2^0\right)}$. When $a_1 = a_1^0$ and $a_2 = a_2^0$, the fuzzy single-point inference yields the following output expression:

$$n_p(b) = \bigcup_{k=1}^{v} x_n \wedge n_0^{(k)} = MIN\left\{n_1^{\left(a_1^0\right)}, n_2^{\left(a_2^0\right)}\right\} \tag{7}$$

Subsequently, fuzzy logic inference was applied to compute the output values of fuzzy variables. Based on the previously derived fuzzy variables and rules, inference was performed to produce a fuzzy output. This process allows the system to synthesize multidimensional data and emulate the comprehensive evaluation logic used by instructors when assessing student performance. For example, in chemical engineering virtual simulation training, if the fuzzy variable for "reagent volume" is categorized as "precise," "stirring speed" as "moderate," and "reaction time control" as "reasonable," then—based on the rule that "all three optimal → high reward level"—the inferred reward output is "high." In contrast, if "volume deviation" is large but "stirring" is correct, the inferred reward output may be "moderate-to-low." This approach avoids the limitations of single-metric evaluation and reflects the multidimensional assessment strategies often adopted in instructional settings, thereby enhancing the alignment between reward generation and the pedagogical complexity of virtual simulation training.

Finally, the defuzzification module was applied to process the fuzzy variables and generate specific reward values. The output of fuzzy inference must be converted into concrete numerical values to serve as effective feedback signals for reinforcement learning. Moreover, a standardized numerical range was employed to ensure comparability of reward values, allowing the system to clearly differentiate between outcomes such as "efficient and correct," "inefficient but correct," and "incorrect operation," thereby precisely refining the assistance strategy. In this framework, the reward function used to address the system assistance problem is composed of two parts: a positive reward and a negative reward. The positive reward is designed to represent the system's ability to complete tasks within the predefined step limit j_{MAX} efficiently. The negative reward penalizes assistance behaviors that result in errors or exhibit high deviation. The positive and negative reward mechanisms are formally expressed as:

$$E^1 = 1 - \frac{j}{j_{MAX}} \tag{8}$$

$$E_s^2 = -\left(\frac{F_h - F_s^2}{F_h}\right) \cdot MAX\left\{ e^{z_1\left(\frac{MAX\left(\left|M_s|s|\right|\right)}{M_{MAX}}\right)}, e^{z_2\left(\frac{MAX(O_s)}{O_{MAX}}\right)}, \right\} \tag{9}$$

where, F_h denotes the target assistance behavior of the system, and F_s^c represents the actual assistance behavior executed at time step s.

2.5 Model optimization based on prior feedback

The core principle of the proposed prior feedback is the dynamic adjustment of exploration strategy weighting through a dual-channel exploration mechanism, thereby enabling efficient and safe instructional assistance. The corresponding structure is illustrated in Figure 2. Within this mechanism, the "dual channels" represent two modes of exploration: conservative exploration based on historical feedback and aggressive exploration targeting new scenarios. Prior feedback acts as a regulatory switch to balance the two. In higher education virtual simulation training, the mechanism adapts flexibly according to students' learning stages. For example, when assisting beginners in tasks such as "circuit soldering," if prior feedback indicates that 80% of responses to "basic operation prompts" were labeled as "positive," the system increases the weight of conservative exploration. This may involve prioritizing previously validated strategies such as "soldering temperature guidance" while reducing reliance on untested strategies like "innovative soldering angle suggestions," thereby avoiding errors caused by excessive exploratory behavior. Conversely, for more advanced students, if prior feedback reveals an increased proportion of "neutral" responses to standard prompts, the mechanism automatically increases the weight of aggressive exploration. At the same time, exploration boundaries are restricted by safety-related knowledge embedded in the prior feedback, ensuring that even new strategies do not encourage operations that violate simulation constraints or pedagogical safety standards. Specifically, noise $PI(\omega^x, \zeta, \delta^{x2})$ was introduced into each output dimension of the assistance behavior, and the distance between the system-generated assistance behavior X_s^e and the experience-referenced behavior X_s^r is calculated as:

$$f_s^x(X_s^r, X_s^e) = \sqrt{\frac{1}{V}\sum_{u=1}^{V}(X_{us}^r - X_{us}^e)^2} \tag{10}$$

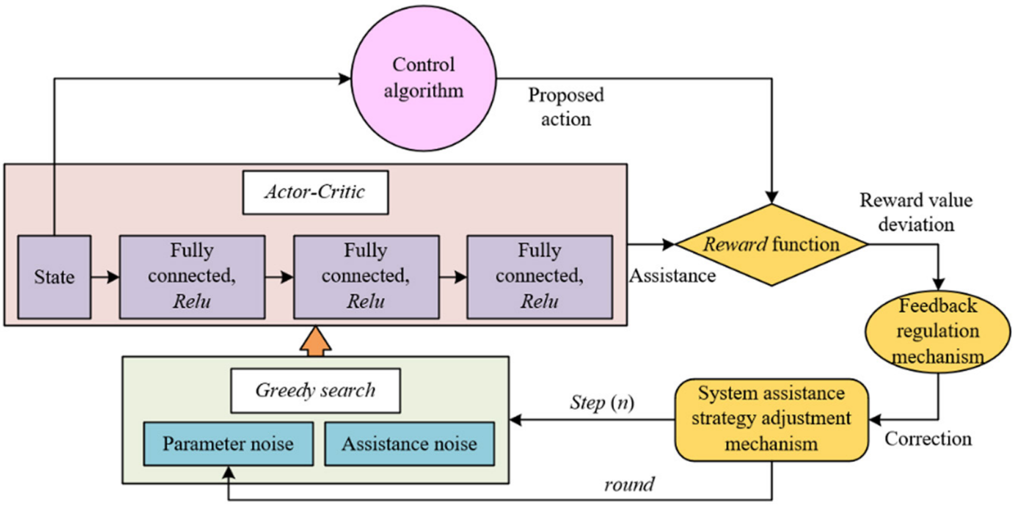

Fig. 2. Principle of the feedback regulation mechanism in the VR/AR-based mobile interaction reinforcement learning framework for virtual simulation training

In higher education virtual simulation training, the parameters governing the assistance strategy require long-term optimization. However, direct parameter adjustments may introduce instability into the strategy. To mitigate this, the system performs updates cyclically over "training episodes." During each episode, the intensity of Gaussian noise is adjusted based on the distribution of episodic rewards obtained from prior feedback. In the early stages of training, prior feedback typically reflects high reward variance. Under such conditions, the system increases the noise intensity to broaden the exploration range, enabling rapid identification of effective parameter regions through diverse parameter configurations. As training progresses, prior feedback may indicate that "a prompt interval of one minute yields the highest short-term reward." In response, the system reduces noise intensity to constrain parameter variations, thereby maintaining training rhythm and preventing excessive policy oscillation. At the conclusion of each episode, short-term rewards are used as evaluation metrics. If an upward trend in reward values is sustained across three consecutive episodes, the current exploration direction is reinforced through prior feedback consolidation, ensuring that strategy optimization remains aligned with instructional efficiency and pedagogical safety objectives. The average short-term reward over a given time interval is computed as:

$$\overline{E}_l = \sum_{s=0}^{S} \varepsilon^{S-s} E_s^2 + E^s \tag{11}$$

Let the updated noise parameters be denoted by δ_{l+1} and σ_{l+1}^e. The parameter noise adjustment process is expressed as:

$$\begin{cases} \delta_{l+1} = 1.01\delta_l, f_l^o(\omega(t\,|\,\varphi), \omega(t\,|\,\tilde{\varphi}_s)) < \sigma_{l+1}^e \\ \delta_{l+1} = \delta_l / 1.01, f_l^o(\omega(t\,|\,\varphi), \omega(t\,|\,\tilde{\varphi}_s)) \geq \sigma_{l+1}^e \end{cases} \tag{12}$$

3 EXPERIMENTAL RESULTS AND ANALYSIS

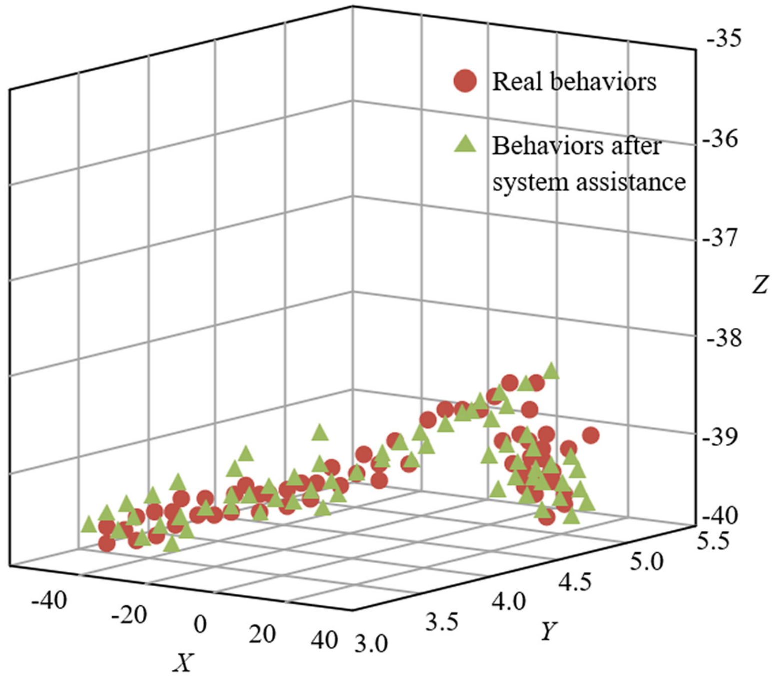

Fig. 3. Comparison of student training behavior locations before and after system assistance

As illustrated in Figure 3, the three-dimensional spatial distribution of student behaviors reveals a pronounced convergence between real behaviors and those exhibited after system-assisted intervention, within the operational space defined by the *X*, *Y*, and *Z* axes. Specifically, within the core simulation region delineated by *Y*-axis values between 3.5 and 5.0 and *Z*-axis values between –38 and –39, a dense alignment of post-assistance behaviors (green points) with real behaviors (red points) was observed, indicating a substantial increase in spatial overlap. Similarly, in the *X*-axis operational range of –20 to 20, green points were more closely clustered around the central aggregation region of the red points. These experimental findings strongly validate the systemic advantages of the proposed VR/AR-based mobile interaction reinforcement learning framework. The observed effectiveness is attributable to the deep integration of its constituent technical modules: (a) Multimodal feature extraction enables precise capture of operational details, providing high-dimensional, fine-grained behavioral data to support informed assistance decisions. (b) Intelligent assistance decision-making and feedback mechanisms dynamically adjust guidance strategies based on real-time behavioral data, correcting operational deviations. (c) The reward function is designed to promote behavioral convergence toward real-world operational patterns, thereby constraining the system's assistance trajectory within pedagogically reasonable bounds. (d) Model optimization via prior feedback leverages historical training data to iteratively improve assistance precision. Ultimately, reduced spatial dispersion and enhanced clustering of post-assistance behaviors in core training regions provide strong empirical evidence that the framework effectively guides student actions towards a reasonable mode in real scenarios.

As illustrated in Figure 4, the system prompt error curve demonstrates that, across the majority of experimental groups, the error values remained within a low fluctuation range of 0 to 3, with only occasional transient spikes reaching approximately 4 to 5. These sporadic peak values were predominantly associated with complex operational scenarios or atypical student behaviors. The occurrence of such peaks is of particular significance, as they highlight both the challenging aspects of virtual simulation and the robustness boundaries of the proposed framework. Notably, even under extreme conditions, the system exhibited the ability to respond using preconfigured strategies, thereby offering empirical guidance for future optimization. Furthermore, Figure 5 depicts the system recommendation error curve, which reveals that in most experimental groups, the recommendation error remained consistently within the 0 to 5 range. Only a few groups exhibited transient peak values, with the maximum approaching 15. This pattern of "low-amplitude stability with localized evolutionary spikes" underscores the systemic advantage of the framework in supporting data-driven self-optimization.

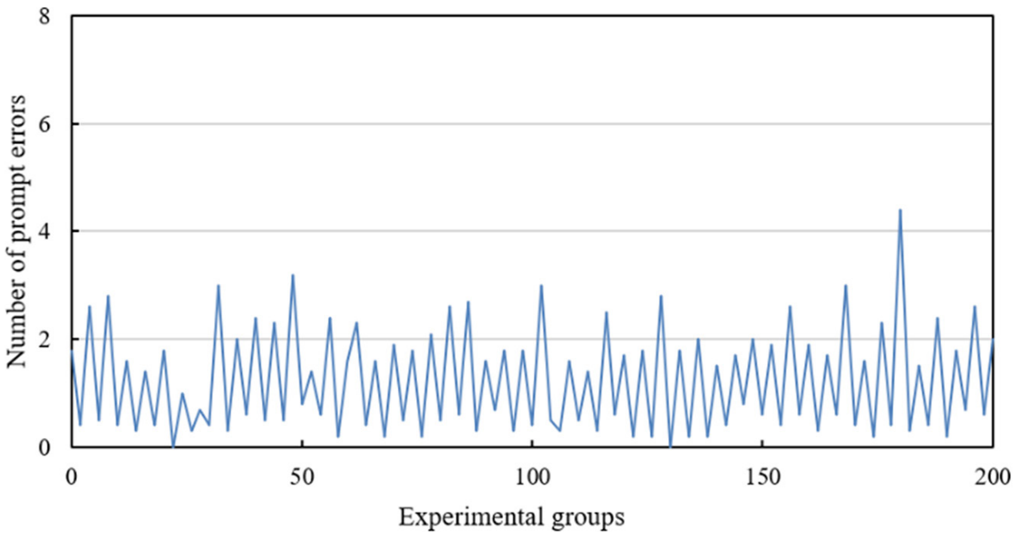

Fig. 4. System prompt error curve

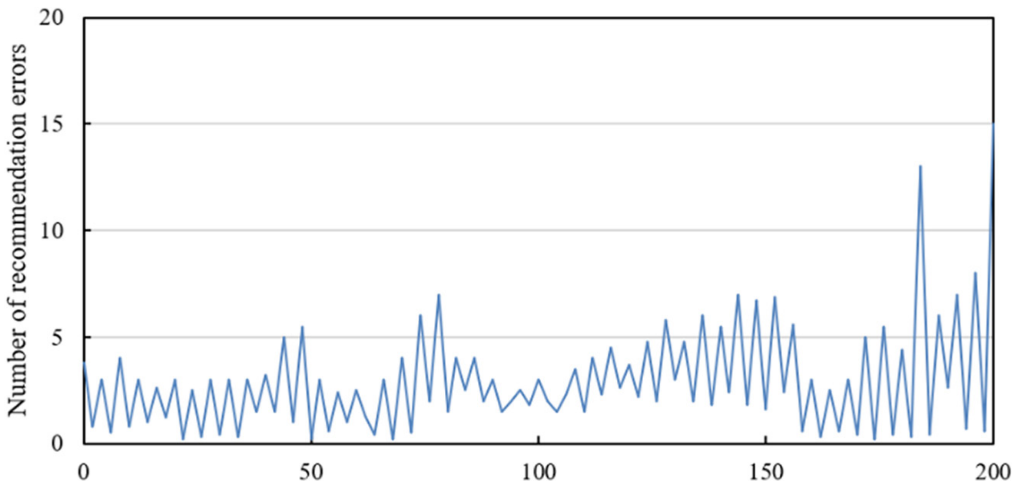

Fig. 5. System recommendation error curve

The low-amplitude and stable fluctuations observed in the error curves of Figures 4 and 5 further reflect the efficacy of the framework's self-evolving closed-loop mechanism based on the data–model–application paradigm. From the perspective of model iteration, the prior feedback optimization mechanism serves as the core engine. Within historical simulation training data, low-error cases are extracted as effective strategy templates, while high-error cases function as negative samples that drive iterative model parameter updates. This data-driven continuous learning process enables the progressive enhancement of the framework's adaptability to complex scenarios across experimental groups, thereby forming an evolutionary cycle of "error experience → model upgrading → error convergence." In the context of educational simulation, the observed low error levels indicate a high degree of precision in the coordination of teaching, learning, and assistance. For students, accurate prompts reduce ineffective guidance, thereby accelerating skill acquisition. From the instructional standpoint, the fluctuation characteristics of the error curves delineate the distribution of procedural challenges within the simulation, thereby assisting instructors in the targeted design of teaching interventions.

Table 1. Changes in the number of successful VR/AR mobile interactions under different reinforcement learning durations

Reinforcement Learning Duration	Student 1	Student 2
0	22	21
400	61	63
600	82	78
800	93	91
1000	82	84

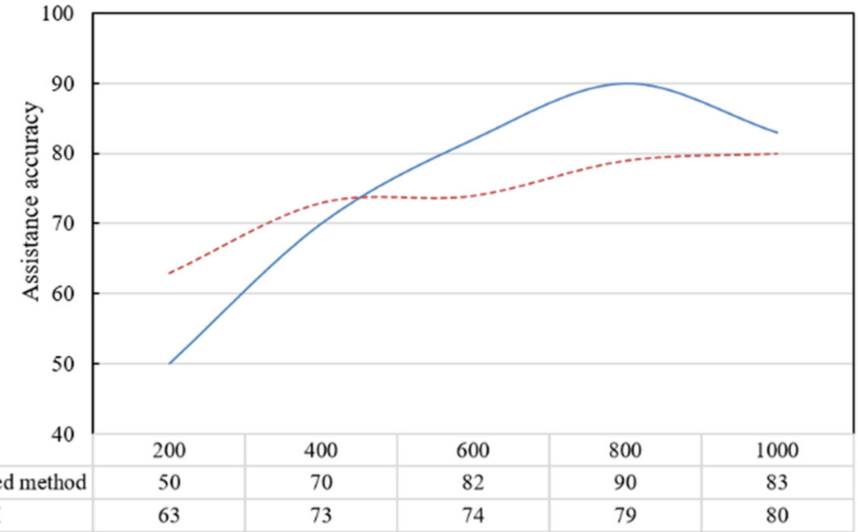

Fig. 6. Comparison of system-assisted accuracy under varying reinforcement learning iterations using different methods

The results presented in Table 1 indicate a dynamic pattern of "increase → peak → adjustment" in the number of successful interactions as reinforcement learning progressed. During the growth phase (duration 0–800), the number of successful interactions increased markedly—from 22 to 93 for Student 1 and from 21 to 91 for Student 2—representing an improvement exceeding 300%. This trend reflects the coordinated functioning of the core modules within the framework. In the adjustment phase (duration 800–1000), a slight decline in success rates was observed for both students. This reduction does not indicate performance degradation but rather reveals a temporary trade-off between increasing scenario complexity and the model's iteration cycle. As the simulation transitioned from routine operations to multi-step, high-degree-of-freedom tasks, the prior strategy knowledge embedded in the model required reconfiguration to accommodate these novel contexts. The 1000-duration stage marked a transitional period from the fading of outdated strategies to the emergence of new ones, thereby substantiating the framework's capacity for dynamic scene recognition and autonomous optimization.

As shown in Figure 6, a clear performance gap was observed between the proposed method and the comparison method, Extended Hidden Markov Model (EHMM), across different reinforcement learning iterations. During the growth phase (50–800 iterations), the assistance accuracy of the proposed method increased sharply from 50% to 90%, representing an 80% relative improvement. In contrast, EHMM exhibited a modest increase from 63% to 80%, corresponding to a 27% gain. This disparity was primarily attributed to the synergistic operation of the modular components embedded in the proposed framework. In the fine-tuning phase (800–1000 iterations), the accuracy of the proposed method slightly declined to 83%, while the accuracy of EHMM experienced a marginal increase to 80%. This short-term divergence should not be interpreted as a failure of the proposed framework; rather, it reflects a transitional interaction between elevated scenario complexity and the model's iteration cycle. As reinforcement learning progresses into high-complexity task scenarios, previously acquired prior knowledge must be restructured to accommodate new requirements. The proposed framework, equipped with a dynamic evolutionary mechanism, maintained a relative advantage during this transitional stage, whereas EHMM encountered performance stagnation due to its reliance on a static model architecture.

4 CONCLUSION

This study centered on the construction and validation of a reinforcement learning framework for VR/AR-based mobile interaction, in which four core modules—multimodal feature extraction, intelligent auxiliary decision-making, reward function design, and prior feedback optimization—were innovatively integrated to establish a closed-loop technical system of perception–decision–feedback–evolution. The proposed framework demonstrated significant effectiveness through multidimensional experimental validation. At the human–computer interaction level, the fusion of voice and gesture inputs enabled the resolution of ambiguities associated with unimodal signals, resulting in substantial improvements in interaction naturalness and intention recognition accuracy. At the level of auxiliary decision-making, dynamic feedback and reward mechanisms facilitated precise adaptation to student behavior, consistently yielding higher assistance accuracy throughout the reinforcement learning process compared to conventional methods. At the model evolution level, the prior feedback mechanism empowered the system to learn from historical data, progressively adapting to increasingly complex virtual training scenarios. In terms of research value, the presented framework addresses a technological gap by integrating VR/AR interaction with reinforcement learning, establishing a reusable and intelligent auxiliary system for virtual training. From a pedagogical perspective, practical pathways are provided for implementing personalized, precise, and adaptive support in higher education virtual training. By enhancing interaction efficiency and assistance precision, the time required for student skill acquisition was significantly reduced, thereby facilitating the advancement of virtual training from tool-based simulation to intelligent enablement.

5 REFERENCES

[1] M. F. Ramadhan, M. Mundilarto, A. Ariswan, I. Irwanto, B. Bahtiar, and S. Gummah, "The effect of interface instrumentation experiments-supported blended learning on students' critical thinking skills and academic achievement," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 14, pp. 101–125, 2023. https://doi.org/10.3991/ijim.v17i14.38611

[2] V. Garrels and P. Zemliansky, "Improving student engagement in online courses through interactive and user-centered course design: Practical strategies," *Nordic Journal of Digital Literacy*, vol. 17, no. 2, pp. 112–122, 2022. https://doi.org/10.18261/njdl.17.2.3

[3] F. Prasetya, A. Fortuna, A. D. Samala, B. R. Fajri, F. Efendi, and A. Nyamapfene, "Effectiveness of distance learning computer numerical control based on virtual laboratory using a metaverse platform to improve students' cognitive ability and practice skills," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 24, pp. 4–21, 2023. https://doi.org/10.3991/ijim.v17i24.45019

[4] A. M. Tessema and L. Nicola-Gavrilă, "Assessment of the experiences of higher education institution on online learning: The case of some selected institutions," *Journal of Research, Innovation and Technologies*, vol. 2, no. 1, pp. 49–56, 2023. https://doi.org/10.57017/jorit.v2.1(3).04

[5] P. Quatresooz *et al.*, "Transforming anatomy and histology practical assignments and lab activities during compulsory distance education," *International Journal of Technologies in Higher Education*, vol. 18, no. 1, pp. 184–193, 2021. https://doi.org/10.18162/ritpu-2021-v18n1-16

[6] H. H. Yu, K. M. Lorenzo-Elarco, M. J. Murro, E. L. McAnany, and H. R. Anderson, "Teaching leadership with popular culture: Practical lessons from Harry Potter," *Journal of Public Affairs Education*, vol. 28, no. 2, pp. 156–181, 2022. https://doi.org/10.1080/15236803.2021.1989735

[7] M. Caliendo, D. A. Cobb-Clark, C. Obst, and A. Uhlendorff, "Risk preferences and training investments," *Journal of Economic Behavior & Organization*, vol. 205, pp. 668–686, 2023. https://doi.org/10.1016/j.jebo.2022.11.024

[8] H. Sun, F. Yang, P. Zhang, and Y. Zhao, "Flight training risk identification and assessment based on the HHM-RFRM Model," *Sustainability*, vol. 15, no. 2, p. 1693, 2023. https://doi.org/10.3390/su15021693

[9] D. Zherdev, L. Zherdeva, S. Agapov, A. Sapozhnikov, A. Nikonorov, and S. Chaplygin, "Producing synthetic dataset for human fall detection in AR/VR environments," *Applied Sciences*, vol. 11, no. 24, p. 11938, 2021. https://doi.org/10.3390/app112411938

[10] S. Oberdörfer, S. Birnstiel, M. E. Latoschik, and S. Grafe, "Mutual benefits: Interdisciplinary education of pre-service teachers and HCI students in VR/AR learning environment design," *Frontiers in Education*, vol. 6, p. 693012, 2021. https://doi.org/10.3389/feduc.2021.693012

[11] K. Frady, "Use of virtual labs to support demand-oriented engineering pedagogy in engineering technology and vocational education training programmes: A systematic review of the literature," *European Journal of Engineering Education*, vol. 48, no. 5, pp. 822–841, 2023. https://doi.org/10.1080/03043797.2022.2141610

[12] K. E. Kreider, "Comparing satisfaction and outcomes in on-campus versus virtual education for nurse practitioner students," *Journal of the American Association of Nurse Practitioners*, vol. 35, no. 2, pp. 135–141, 2023. https://doi.org/10.1097/JXX.0000000000000826

[13] M. Korkiakoski, P. Alavesa, and P. Kostakos, "Preference in voice commands and gesture controls with hands-free augmented reality with novel users," *IEEE Pervasive Computing*, vol. 23, no. 1, pp. 18–26, 2024. https://doi.org/10.1109/MPRV.2024.3364541

[14] K. Amara, C. Boudjemila, N. Zenati, O. Djekoune, D. Aklil, and M. Kenoui, “AR computer-assisted learning for children with ASD based on hand gesture and voice interaction,” *IETE Journal of Research*, vol. 69, no. 12, pp. 8659–8675, 2023. https://doi.org/10.1080/03772063.2022.2101554
[15] L. Lyu, T. Hu, H. Wang, and W. Hou, “SRD method: Integrating autostereoscopy and gesture interaction for immersive serious game-based behavioral skills training,” *Electronics*, vol. 14, no. 7, p. 1337, 2025. https://doi.org/10.3390/electronics14071337
[16] C. Khundam, V. Vorachart, P. Preeyawongsakul, W. Hosap, and F. Noël, “A comparative study of interaction time and usability of using controllers and hand tracking in virtual reality training,” *Informatics*, vol. 8, no. 3, p. 60, 2021. https://doi.org/10.3390/informatics8030060
[17] K. Kim and J. H. Lee, “The effect of feedback in virtual attention training on orienting attention in individuals with sluggish cognitive tempo,” *Journal of Attention Disorders*, vol. 26, no. 12, pp. 1640–1652, 2022. https://doi.org/10.1177/10870547221090664

6 AUTHORS

Xiaoxia Tian from 2011 to 2013 studied at Shijiazhuang University of Economics and received his Bachelor’s degree in 2013. From 2013 to 2016, he pursued his Master’s degree at Shijiazhuang Tiedao University and graduated in 2016. Currently, he works at the Department of Computer and Information Technology, Hebei Vocational College of Resources and Environmental Technology. During his tenure, he has published research papers and monographs and has guided students to win awards in provincial-level skills competitions (Higher Vocational Group). His research interests focus on Computer Application Technology (E-mail: xiaoxia0061@163.com).

Yanjin Wang is a Lecturer at Hengshui Open University; Received Bachelor’s degree from Hebei University of Economics and Business School of Economics and Management; Master’s degree in Education from Chonnam National University, South Korea; and research interest includes Higher education in ideological and political theory and distance education (E-mail: wangyanjinhs@126.com).