Article

Practice of Innovative Talent Cultivation in Engineering Practical Training Driven by Multimodal Perception Competition Projects

Xinci Wei[1], Yu Zhang[1], Yanfeng Wu[1], Zi Gao[1], Shengcai Yang[1] and Yongjie Zhang[1,2,*]

[1] Guangxi Vocational Normal University, Nanning, China
[2] Internet of Things Electronic Technology Experimental Teaching Center, Nanning, China
* Correspondence: Yongjie Zhang, Guangxi Vocational Normal University, Nanning, China; Internet of Things Electronic Technology Experimental Teaching Center, Nanning, China

Abstract: While students majoring in intelligent technologies in higher vocational colleges can complete the debugging of individual hardware modules, they generally lack the capability for multi-device joint debugging and system architecture design when working on complete multi-modal perception system projects. To address the issues of fragmented practical training content and the insufficient alignment between curriculum instruction and the requirements of competitions and professional roles, this study leveraged an on-campus key laboratory to integrate a multi-modal perception competition project into the "Arduino Hardware Development" practical training module. Two classes of 50 students each underwent comparative teaching over one academic year; the experimental class additionally received competition-oriented tiered practical training, with unified assessment criteria. The results showed that the average hardware testing score for the experimental class was 81.86, compared to 75.1 for the control class, yielding a difference of 6.76 points; the average time per student required for algorithm debugging was 3.5 hours in the experimental class versus 4.2 hours in the control class, representing an approximately 16.7% improvement in development efficiency. This pedagogical model enhances the effectiveness of practical training and provides a practical reference for similar majors in other higher vocational colleges within the region.

Keywords: multimodal perception; discipline competitions; engineering practical training; competition-teaching integration; innovative talent cultivation

1. Introduction

Multimodal perception technology integrates multi-source sensory data such as vision and LiDAR to perceive the environment, is the foundation for the autonomous decision-making of intelligent robots, intelligent connected equipment, and unmanned systems, and is also a key development direction of intelligent manufacturing. In engineering research, heterogeneous multi-agent formation navigation builds a joint state space by means of multimodal observations to achieve formation maintenance and cooperative collision avoidance, and simulation and physical experiments have verified the engineering value of multi-sensor integration [1]. At the software and algorithm level, to address problems such as the computing waste of the traditional frame-sampling mode and the temporal drift among sensors, the event-clock tactile perception algorithm triggers computation only for physical events such as contact and slip, completes inference by fusing visual and acoustic cues, and achieves low-latency and highly stable contact inference on edge hardware [2]. Focusing on talent cultivation in multimodal perception technology, numerous domestic education researchers have carried out explorations of competition-teaching integration and project-driven practical training reform. other research, based on the OBE concept, proposed the competition-teaching

Received: 19 September 2026
Revised: 06 November 2026
Accepted: 19 November 2026
Published: 24 November 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

integration approach of "promoting learning through competitions, empowering through competitions, promoting teaching through competitions, and promoting reform through competitions", advocating that competition requirements be reversely integrated into curriculum objectives, practical training content, and evaluation systems [3]. Research has pointed out that [4] the VR-based blended practical training system can enrich practical teaching carriers, optimize practical training design and improve the education effect through closed-loop training and assessment. This reform integrates multimodal perception with autonomous navigation tasks, incorporates competition content into the curriculum, and verifies the effect through parallel class comparison,It offers references for practical teaching reform of similar majors at other vocational colleges across Guangxi.

2. Overall Architecture Design of Competition-Driven Multimodal Perception Practical Training Education

2.1. Development Status of Technologies for Multimodal Perception Practical Training Teaching

YOLOv8, Vosk, and lightweight Transformer are the development frameworks selected for this project. Undergraduate universities carry out algorithm training based on professional laboratories, but such schemes involve high hardware investment and a high learning threshold, making it difficult to implement large-scale layered teaching in higher vocational colleges.

Integrated Architecture Diagram of Competition-Driven Practical Training Education: Academic competitions related to multimodal perception are themselves a complete engineering project, and the various ideas in project management can be applied to innovative talent cultivation. We integrated these ideas into three-stage practical training teaching, taking foundation consolidation, comprehensive integration, and competition innovation as the three implementation stages, to train students' engineering thinking mode and improve their practical abilities in project planning, troubleshooting, and teamwork. With reference to the actual demands of job positions and the competition scoring rules, we transformed these contents into classroom teaching standards and built a four-layer closed-loop teaching framework, with the hardware training layer as the foundation, the layered training task layer as the carrier, the competition project development layer as the core, and the innovation incubation evaluation layer as the verification. as shown in Figure 1.

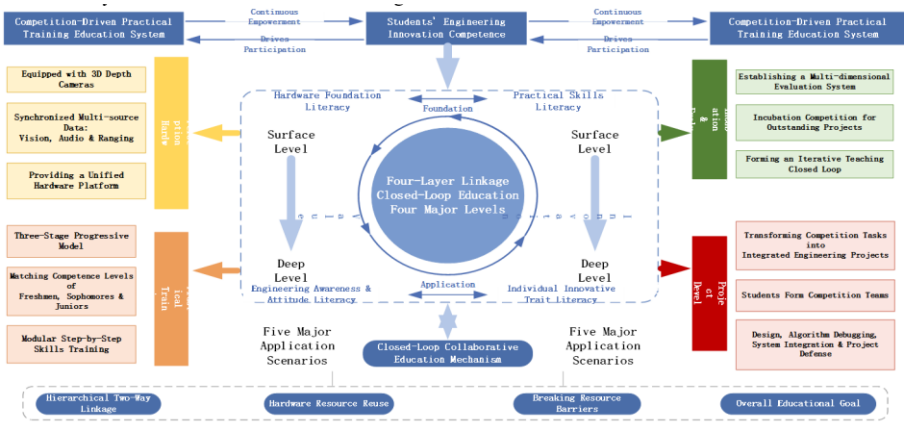

Figure 1. System Architecture Diagram of the Closed-Loop Feedback in Practical Training Education

As can be seen from Figure 1.The whole system establishes a closed-loop cultivation mechanism that can receive various types of feedback and make adjustments accordingly. In this way, the old problems that classroom lecturing, practical training exercises, academic competitions, and innovation practice were mutually independent and disconnected from each other in the past can be properly resolved, the fragmented learning state in which students passively receive knowledge points is changed, the

theoretical knowledge in books is connected with real engineering projects, and the mutual integration of theoretical learning and engineering practical exercises is achieved.

2.2. Design of the Layered Practical Training Model for Multimodal Perception Competition Projects

Theoretical and Technical Logic of the Layered Practical Training Model: The practical training platform we built takes the four-layer teaching framework as the foundation, and the platform is divided into three modules, namely the perception layer, the data layer, and the decision layer, to correspond to the teaching objectives set by the three-stage talent cultivation program. With the perception layer as the basic hands-on module, sensor calibration as the exercise project, and single-type data acquisition as the basic task, students can complete these basic hands-on operations; with YOLOv8 and Vosk tools as the data-layer processing means, the cleaning of various types of raw collected data is completed and data processing is performed in the local environment; with on-site interference conditions such as illumination changes and environmental noise as references, the decision layer can adjust the data fusion schemes, continuously iterate the system, and realize the autonomous judgment of devices. The practical training classroom divides students into groups of 3-4, replicating the real scenarios of time-limited development and on-site troubleshooting in academic competitions, and combines hands-on experiments, algorithm debugging, system integration,The layered construction of model is shown in Figure 2.

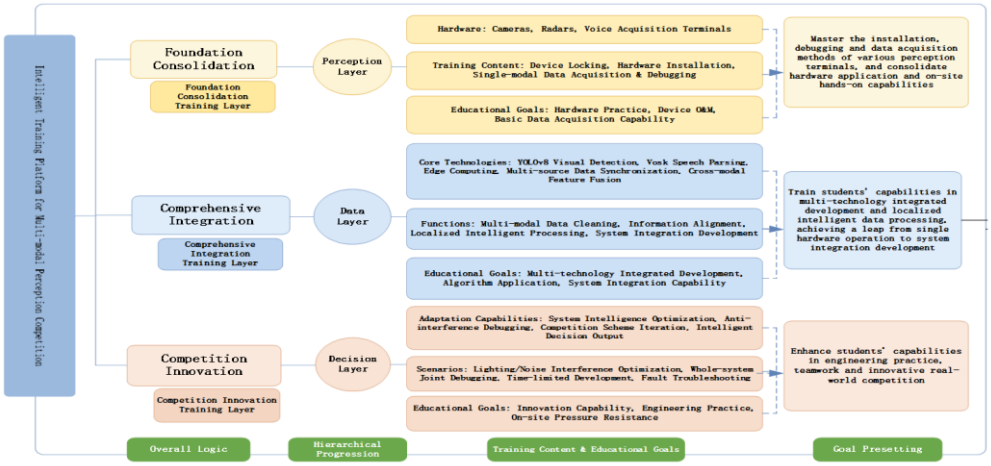

Figure 2. Three-stage Layered Practical Training Model of Multimodal Perception Competitions

This practical training builds a stepped layered learning structure. At the outset, students mainly work on their own, completing entry-level operations such as sensor installation and calibration and basic data acquisition to build up hands-on hardware experience first. Once past this step, the course moves on to comprehensive practical training centered on the YOLOv10-MCS multimodal collaborative perception technology.

2.3. Design of the Implementation Process of Competition Project Practical Training

Daily practical training teaching is divided into four different execution stages, including preliminary theoretical lecturing, layered hands-on exercises, repeated optimization of competition projects, and post-competition problem summarization and sorting, For detailed implementation steps, refer to Figure 3.

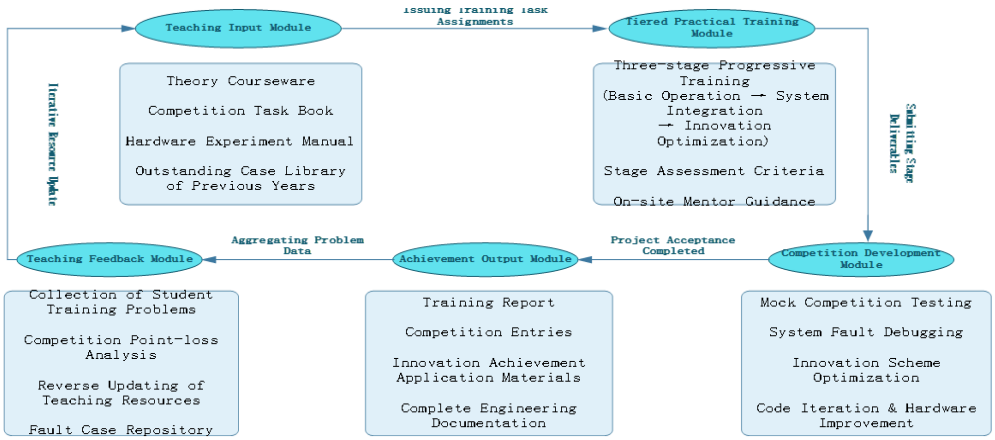

Figure 3. Functional Composition of the Whole Implementation Process of Multimodal Competition Project Practical Training

Figure 3 illustrates the closed-loop implementation process of the multimodal perception competition-based practical training, which comprises five key stages: instructional input, tiered practical training, competition development, outcome delivery, and instructional feedback. These stages are tightly integrated and advanced iteratively in a cyclical manner.3. Testing and Analysis of the Teaching Effect of the Practical Training Education Model

2.4. Design of the Teaching Test Scheme

To test the teaching effect, two classes with 50 students each were selected to carry out a one-academic-year controlled experiment. The two classes used the same textbooks, the same teaching progress, and the same final assessment standards; the only difference was that the experimental class added competition-driven layered practical training in addition to regular teaching. For specific comparison data, see Table 1.

Table 1. Comparison of Teaching Effects between Traditional Practical Training and Competition-Driven Practical Training

Experimental group (both the control group and the reform group have 50 students)	Practical operation pass rate Traditional engineering practical training mode	Multimodal perception competition project-driven practical training mode
Practical operation pass rate	60%	97.7%
Provincial-level and above competition award rate	12.2%	54.9%
Overall assessment score	Above 90 points accounts for 2%	Above 90 points accounts for 10%
Below 60 points is 0% for both	60-90 points accounts for 98%	60-90 points accounts for 90%

Training data collected in this study shows that the practical assessment pass rate rose from 60% to 97.7%, an increase of 37.7 percentage points. The share of students winning awards at provincial level or higher jumped from 12.2% to 54.9%, representing a 42.7-percentage-point gain.Looking at score distributions, only 2% of students in the control class scored above 90, while 98% fell between 60 and 90 points. In the experimental class, 10% achieved scores over 90 and 90% landed in the 60–90 range. Neither group recorded failing grades.This set of data reveals that competition-oriented stratified training abandons isolated teaching of scattered knowledge points. It guides students through full-project training centered on hardware equipment and related algorithms. This teaching plan can be implemented in practice and holds value for wider promotion.

2.5. Simulation Test of Multimodal System Performance

Ten groups of student competition works were selected from the experimental class for simulation tests, simulating high-frequency competition working conditions such as illumination changes, environmental noise, and long-distance recognition, to test the system recognition accuracy the test results are shown in Figure 4.

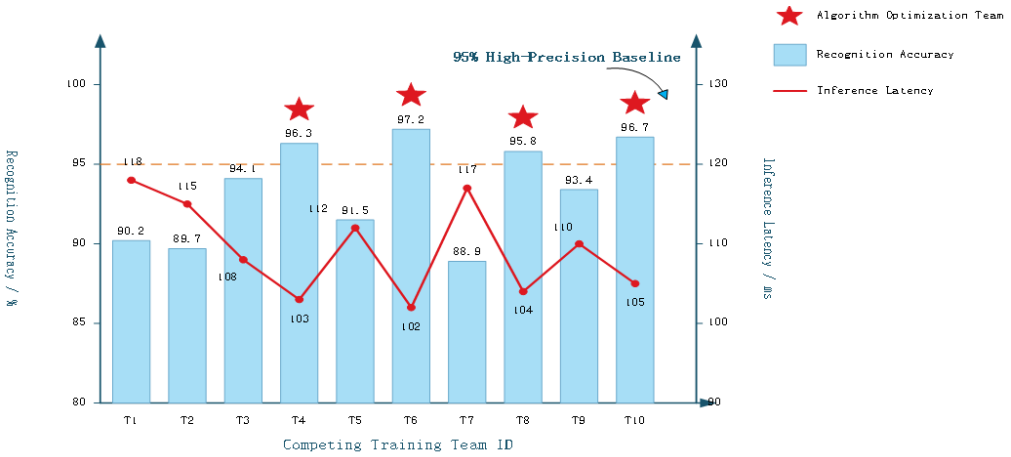

Figure 4. Simulation Test Results of the Multimodal Perception System of Competition Practical Training Works

After implementing the three-stage stratified practical training, students achieved an average score of 81.86 in the Arduino Hardware Development course. The average time for individual algorithm debugging was reduced to 3.5 hours. Compared with the control class, the group development efficiency increased by 16.7%, the pass rate of practical assessment reached 97.7%, and the proportion of students winning provincial-level or higher competitions stood at 54.9%. The measured data fully verifies that this practical training system delivers remarkable effects on competency cultivation. The competition-education integration scheme proposed in this paper is feasible, and it can offer practical references for practical training of IoT-related majors.

3. Conclusion

At the present stage, practical training teaching in intelligence-related majors of higher vocational colleges generally has problems such as fragmented training content and poor alignment among job demands, curriculum teaching, and academic competitions. rely on multimodal perception academic competitions, build an overall four-layer teaching framework, design a three-stage layered practical training model, and establish a competition-teaching integration practical training system adapted to professional teaching, so as to cultivate high-caliber technical to serve regional industrial development.

Funding: National University Innovation and Entrepreneurship Training Program: Research on Full-Dimensional Perception Fusion Optimization of Intelligent Connected Electric Vehicles for ASEAN Tropical Scenarios (Project No.: 202614684012); 2026 Guangxi Undergraduate Teaching Reform Project "Innovation and Practice of Teaching Mode for Electronic Hardware Cross-Modal Fusion Experimental Cabin Based on Artificial Intelligence" (Grant No. 2026JGB598)

References

1. C. Xu, T. Zheng, X. Xie, et al., "Multimodal sensing-computing devices: Toward a new paradigm for embodied intelligence," *Materials Today*, 2026, Art. no. 103223, doi: 10.1016/j.mattod.2026.103223.
2. J. He, "Application of multimodal learning in robotic perception: an intelligent perception framework integrating vision, sound, and touch," *Intelligent Service Robotics*, vol. 19, no. 2, p. 31, 2026, doi: 10.1007/s11370-025-00686-y.
3. Y. Q. Xie, F. Y. Y., "Learn by Competing, Empower by Competing, Teach by Competing, Refine by Competing" Research on the integration of teaching and discipline competition in Events Management major based on OBE concept," *Pacific International Journal*, vol. 8, no. 1, 2025, doi: 10.55014/pij.v8i1.777.

4. J. Zhang, Q. Lu, and Y. Wu, "Advancing Sustainable Practical Engineering Education: A Closed-Loop VR-Blended Learning System for Diesel Engine Assembly Training and Assessment," *Sustainability*, vol. 18, no. 17, p. 8959, 2026, doi: 10.3390/su18178959.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.