A Computer Vision-Enabled Closed-Loop Framework for Adaptive Practical-Skill Instruction in Higher Vocational Education: A Rail Transit Maintenance Case

Yingge Li*, Ruihao Guo, Hongmei Liu, Weixun Chen
Department of Guangzhou Railway Polytechnic, Guangzhou, China
**Corresponding Author*

Abstract: Practical training is central to talent cultivation in vocational education, yet adaptive instruction for practical skills remains difficult to implement at scale. Taking rail transit programs in Chinese higher vocational colleges as an example, this framework study identifies three major challenges: large individual differences in skill acquisition, insufficient fine-grained perception of operational processes, and limited capacity for targeted remediation. To address the data gaps inherent in traditional learner profiling, which primarily relies on online logs and test results, this paper proposes an adaptive practical skills teaching application model based on computer vision. This model is constructed around a closed-loop structure of collection-diagnosis-remediation-assessment, encompassing four levels: data foundation, skill diagnosis, targeted remediation, and outcome assessment, while also incorporating a risk management mechanism. The model focuses on five core dimensions of practical-skill diagnosis and assessment: operational compliance, safety compliance, action quality, job competence, and skill internalization. By extending learner-state data from online behavior to visual and multimodal evidence generated during practical operations, the model makes the operational process easier to observe and diagnose. This paper uses a rail transit maintenance training scenario to illustrate the implementation path of the model. This model provides a practical framework for data-driven, personalized skill instruction in higher vocational education.

Keywords: Computer Vision; Practical Skills; Learning Diagnosis; Digital Transformation of Vocational Education

1. Introduction

Digital and intelligent technologies are reshaping education. Promoting large-scale adaptive instruction through artificial intelligence (AI) has become an important direction for high-quality educational development. Guangdong Province's "1333" initiative also identifies AI-enabled educational applications and large-scale adaptive instruction as key tasks. For vocational education, this transformation is especially significant. Because vocational education aims to cultivate technical and skilled talent, practical training is not only an instructional component but also the core arena in which learners develop occupational competence.

Compared with knowledge-based instruction, however, adaptive instruction for practical skills is more difficult to deliver. Taking rail transit programs as an example, the accelerated development of the South China urban agglomeration and the transformation of rail transit toward intelligent operation and maintenance have increased the demand for skilled front-line workers, including rolling-stock maintenance technicians, signaling maintenance technicians, and train drivers [1-3]. At the same time, practical training still faces three persistent problems. Firstly, there are significant differences among students in terms of starting point, learning speed, and skill acquisition trajectory, making it difficult for standardized training programs to meet their needs. Secondly, teachers often rely on on-site observation and result-oriented scoring methods, which makes it difficult to capture subtle differences and maintain objectivity during the operation process. Thirdly, remedial measures for skill deficiencies still rely heavily on experience, and personalized corrections lack reusable, evaluable, and scalable implementation ways.

A key reason for these issues is the lack of data regarding practical skill learning. Existing learner analysis and precision teaching research primarily rely on online behavior logs, test scores, and structured learning records. While these data can describe knowledge acquisition, they struggle to capture the inherent practical skills involved in operational processes. For example, whether procedures were followed, whether key actions were performed correctly, whether tools were used properly, and whether safety requirements were adhered to. These elements are crucial for skill development but have long been excluded from traditional educational data collection. Meanwhile, computer vision and deep learning have matured in human action recognition, object detection, pose estimation, and action quality assessment, providing a technical basis for perceiving skills through visual evidence and diagnosing skill deficiencies through data.

Accordingly, as a framework study rather than an empirical evaluation, this paper proposes a computer vision-enabled application model for adaptive instruction of practical skills in the context of vocational practical training. The central contribution is the extension of learner-state data from online behavior logs to visual and multimodal evidence generated during practical operations. On this basis, the paper adapts the learning analytics logic of collection, diagnosis, intervention, and optimization to real practical-training environments, constructing a four-layer model consisting of data foundation, skill diagnosis, targeted remediation, and outcome assessment. The paper also illustrates the implementation path through a rail transit maintenance training scenario, providing a design reference for higher vocational colleges seeking to promote data-driven and personalized skill instruction.

2. Related Work

2.1 AI-Enabled Adaptive Instruction and Learner-State Modeling

The core of educational digital transformation lies in data-driven change, and data have become an important lever for large-scale adaptive instruction [4,5]. In vocational education, studies and practices related to student profiling, precision teaching, and data-driven adaptive instruction have continued to develop [6]. To identify learner differences and provide differentiated instruction, learning analytics and educational data mining have developed a range of learner-state modeling methods. Knowledge tracing, for example, models students' historical response sequences to characterize their knowledge states and predict future performance. It has become a key technique for personalized tutoring and adaptive learning, evolving from early probabilistic and Bayesian models to deep knowledge tracing models based on recurrent neural networks and attention mechanisms [7-9]. Researchers have also increasingly emphasized interpretability, arguing that diagnostic results should be not only accurate but also understandable to teachers and students so that they can support instructional decisions [10].

Nevertheless, existing studies have two limitations. Firstly, their data sources are dominated by online behavioral logs and test results, which mainly represent symbolic and structured learning behavior. Secondly, their research objects are often knowledge-based courses, where learner state is largely understood as knowledge mastery. However, the core operational skills of vocational education, including operational compliance, safety compliance, movement quality, work ability, and skill internalization, are difficult to capture solely through online behavioral data.

2.2 Computer Vision and Intelligent Assessment of Skill Actions

Computer vision combined with motion quality assessment technology can automatically assess the standardization and quality of motion in videos and has been widely applied in fields such as sports, rehabilitation and surgical skills [11]. Deep learning methods have driven the development of motion recognition from manual feature extraction to architectures such as CNN, two-stream networks, Transformer and pose estimation, and have played an important role in fields such as public safety, motion analysis and medical rehabilitation [12,13].

These works establish a methodological basis for perceiving practical skills from visual data. However, applications in education, especially vocational practical training, remain limited. On the one hand, many studies focus on recognition or scoring tasks themselves and seldom embed them into a complete educational loop connecting diagnosis, remediation, and assessment. On the other hand, existing applications often target relatively standardized actions in fields such as sports and medicine, while less attention has been paid to complex industrial training tasks involving operational procedures, tools, equipment, safety rules, and job competence, such as rail transit maintenance and signaling.

2.3 Digitalization and Intelligentization of Vocational Practical Training

At the practical level, vocational education is accelerating the digitalization of practical training. Virtual simulation training has been widely used to address the problems of high cost, high consumption, high risk, and limited observability in practical training. The Construction Guideline for Demonstration Virtual Simulation Training Bases in Vocational Education proposes using sensors and monitoring devices to collect process data from teaching, learning, examination, practice, and assessment, and applying AI to mine and analyze such data to improve training quality [14]. Existing studies also indicate that the circulation of training data and the effective application of AI can support dynamic monitoring of learning progress, diagnosis of learning difficulties, and personalized feedback, although problems such as data silos, lack of standards, and insufficient system integration remain [15].

It should be noted that virtual simulation training mainly collects in-simulation interaction data, such as the correctness of operation steps and simulation scores. In contrast, the visual perception and diagnosis of students' real operational processes in physical training environments remain underdeveloped. This is precisely the entry point of the present study.

2.4 Research Gap and Entry Point

In summary, existing research has four shortcomings. First, the data dimension is relatively narrow: learner-state modeling relies heavily on online behavior and test data, while visual and multimodal data generated during practical operations are underused. Second, research objects are biased toward knowledge-based courses, and dedicated studies on adaptive instruction for operational skills remain insufficient. Third, technical applications often remain at the level of action recognition, action scoring, or virtual-simulation process recording, without forming an educational closed loop that connects diagnosis, remediation, and assessment in real training environments. Fourth, technological empowerment is often discussed more fully than governance, while privacy protection, data security, algorithmic bias, and ethical risks related to video and biometric data require more systematic attention. To address these gaps, this paper constructs a computer vision-centered application model for adaptive instruction of practical skills in rail transit training, while embedding risk governance throughout the model.

3. Theoretical Foundations and Design Principles

3.1 Theoretical Foundations

The model construction draws on three bodies of theory. The first is skill formation theory. Unlike knowledge acquisition, skill acquisition emphasizes repeated practice, immediate feedback, and continuous correction of movements. Low-quality repetition without targeted feedback is unlikely to lead to effective skill internalization [16]. Therefore, practical skills training requires a detailed diagnosis of operational processes and personalized feedback.

The second aspect is the concept of adaptive instruction. Adaptive instruction can be understood as an identification-guidance-development process: first, identifying learners' differences, then providing differentiated instruction, and finally promoting the development of each learner [6]. In the context of practical skills, this means that instruction should not only identify what knowledge students have mastered, but also how they complete tasks.

The third aspect is learning analytics theory, which emphasizes a closed loop of data collection, analytical modeling, intervention feedback, and continuous optimization. These three perspectives all converge on a common proposition: Adaptive teaching of practical skills requires precise perception, diagnosis, and feedback of the operational process.

3.2 Design Principles

Based on these theoretical foundations, the proposed model adheres to four principles. Firstly, it is skill-centered and driven by process data, shifting focus from the final outcomes to operational procedures. Secondly, it establishes a diagnosis-intervention-evaluation loop for continuous teaching optimization. Thirdly, it balances accuracy with interpretability: diagnostic results must be both reliable and capable of identifying the location and cause of the problem. Fourthly, it gives equal emphasis to technology empowerment and governance, incorporating privacy protection, data security, algorithmic fairness, and ethical compliance into the model design.

4. Construction of the Collection-Diagnosis-Remediation-Assessment Application Model

Based on the above principles, this paper constructs a computer vision-enabled application model for adaptive instruction of practical skills. The model is organized around a collection-diagnosis-remediation-assessment closed loop and consists of four layers: the data foundation layer, the skill diagnosis layer, the targeted remediation layer, and the outcome assessment layer. These layers form a bottom-up flow of data and features and a top-down loop of feedback and instructional adjustment. The overall architecture, including the core functions, input data, methods, and outputs of each layer, is summarized in Table 1. Risk governance is embedded throughout the whole process. Policy orientation provides value guidance, while teachers, students, and managers constitute the main application endpoints of the model.

Table 1. Four-Layer Architecture of the Proposed Application Model

Layer	Core function	Main input data	Key CV/data methods	Output
Data foundation layer	Collect, structure, annotate, and integrate practical-training process data.	Operation video, pose data, tool/equipment data, sensor data, assessment records.	Video structuring, key-action segmentation, object/tool recognition, pose extraction, multimodal fusion.	Vision-centered multimodal database and practical-skill indicators.
Skill diagnosis layer	Interpret operational processes and identify skill deficiencies.	Structured video clips, pose trajectories, detected tools/ components, task rules, historical error records.	Object detection, pose estimation, action recognition, temporal modeling, compliance detection.	Learner-state profile, error type, evidence segment, confidence level, weakness diagnosis.
Targeted remediation layer	Transform diagnosis into differentiated instructional intervention.	Diagnostic results, weakness categories, resource repository, learner progress records.	Rule-based matching, resource recommendation, error-pattern analysis, intelligent prompting.	Personalized practice tasks, micro-lectures, corrective prompts, follow-up practice path.
Outcome assessment layer	Evaluate learning outcomes and support iterative optimization.	Process evidence, remediation records, task performance, teacher assessment, repeated-practice data.	Multidimensional scoring, progress analysis, formative and summative assessment, feedback modeling.	Assessment report, progress evidence, model/refinement feedback, instructional decision support.

4.1 Data Foundation Layer

The data foundation layer is responsible for collecting, annotating, structuring, and integrating training-process data. Unlike conventional learner profiling, which relies mainly on online behavior logs, this layer takes operation video as the core and supplements it with pose data, equipment data, sensor data, and training assessment records. Together, these data form a vision-centered multimodal database for practical training.

To make practical skills diagnosable, a learner-state indicator system for practical skills should first be established. This paper defines five core dimensions: operational compliance, safety compliance, movement quality, job competence, and skill internalization, as shown in Table 2. These core dimensions can be further specified through fine-grained observable indicators. For example, operational compliance may include step sequence, action timing, and tool usage sequence; work capability may include task completion quality, equipment status response, and troubleshooting; skill internalization may be reflected in the progress achieved through repeated practice and task transfer. In the data processing stage, operational videos undergo structural processing such as key action segmentation, object and tool recognition, pose extraction, time series alignment, and annotation. Then, data cleaning, multimodal fusion, and privacy anonymization are performed to provide reliable input for higher layers.

4.2 Skill Diagnosis Layer

The skills diagnostic layer is the core mechanism for interpreting the training process and identifying skills deficiencies. This layer employs computer vision and deep learning technologies (including target detection, pose estimation, action recognition, time series modeling, and behavior understanding) to analyze the operational process from four dimensions: (1) task and step identification, identifying the tasks and steps being performed; (2) compliance detection, comparing the trainee's actions with standard procedures and technical requirements; (3) safety compliance warnings, marking risks such as lack of protective equipment, unsafe postures, improper tool use, or violations of on-site operating procedures; and (4) weakness diagnosis, precisely locating specific steps, actions, or decision points where errors repeatedly occur.

Crucially, the diagnostic output must be interpretable: the system should provide evidence (keyframes, action segments, detected targets, motion trajectories, and deviation descriptions), not just scores. This ensures that the diagnosis is no longer limited to black-box scoring but provides a clear and understandable overview of the learner's state, thus facilitating targeted remedial interventions. Simultaneously, visual diagnosis in real training environments should be viewed as a constrained probabilistic process. Factors such as occlusion, camera angle, lighting conditions, equipment complexity, clothing variations, and non-standard operating habits can all affect recognition accuracy. Therefore, the diagnosis layer should integrate visual models with sensor data, task rules, teacher review, and confidence thresholds, rather than relying on fully automated visual judgments.

Table 2. Five-Core Dimensions for Practical-Skill Diagnosis and Assessment

Dimension	Observable evidence	Diagnostic or assessment output
Operational compliance	Step sequence, action timing, tool-use order	Procedure errors, missing steps, repeated weak links
Safety compliance	Protective equipment, unsafe posture, live-working rules	Safety warnings and risk behavior records
Movement quality	Pose trajectory, key movement amplitude, stability	Movement deviation and correction suggestions
Job competence	Task completion, equipment-state response, fault handling	Competence profile and task-readiness judgment
Skill internalization	Improvement across repeated practice and transfer tasks	Developmental progress and value-added evidence

4.3 Targeted Remediation Layer

The targeted remediation layer provides differentiated intervention based on diagnostic results. Its purpose is to transform diagnosis into instructional action. The layer may include three types of functions. The first is resource recommendation, in which personalized training tasks, micro-lectures, demonstration videos, and decomposed practice materials are pushed according to identified weak links. The second is movement correction, in which the system indicates the specific deviation between a student's operation and standard procedures, such as missing a step, using a tool in the wrong order, applying an incorrect posture, or violating a safety rule. The third is intelligent assistance, in which digital teachers or training assistants provide prompts, explanations, and formative feedback during or after practice.

Remediation should be layered, categorized, and iterative. Students with different types of weaknesses should receive different remediation paths. For example, students with procedural errors may need step-by-step rehearsal; students with unsafe behavior may need safety-rule reinforcement and warning-based practice; students with poor movement quality may need repeated movement correction with visual comparison. As new diagnostic data is generated, the remedial path should be dynamically updated to form a personalized training cycle and continuously improved through practice.

4.4 Outcome Assessment Layer

The outcome assessment layer aims to evaluate the effectiveness of adaptive instruction and support model iteration. Assessment should not be limited to a single final score but should focus on the five core dimensions used in diagnosis: operational compliance, safety compliance, action quality, work capacity, and skill internalization. Assessment should also combine formative and summative assessments, integrating process evidence with final performance. While knowledge mastery can serve as supplementary evidence when training tasks require explicit understanding of rules, it is not considered a core practical skill dimension in this model.

The assessment results serve a dual purpose: firstly, they support feedback between teachers and students to adjust teaching methods and promote personalized learning; secondly, they provide feedback to the diagnostic and remedial modules to optimize indicators, update models, and improve resource recommendations. This feedback loop enables continuous improvement of the diagnostic and instructional systems.

4.5 Risk Governance Throughout the Process

Therefore, technology empowerment must be complemented by robust governance mechanisms. At the data level, measures such as informed consent, data minimization, access control, encryption, anonymization, and privacy anonymization should be implemented. At the diagnostic level, algorithm bias and fairness should be monitored to avoid systematic misjudgments caused by imbalanced training data or differences in body shape, gender, clothing, equipment, or shooting angles. At the application level, the auxiliary role of artificial intelligence should be clarified: AI should assist rather than replace teachers' judgments, and respect students' autonomy. In addition, traceability, interpretability, manual review, and ethical review mechanisms should be established to ensure the compliance and credibility of applications.

5. Application Scenario and Implementation Path: Rail Transit Maintenance Training

Rail transit maintenance training is a typical application scenario for the proposed model. In tasks such as overhead line inspection or vehicle component maintenance, trainees need to strictly adhere to operating procedures, use tools correctly, maintain appropriate operational quality, respond promptly to equipment status, comply with safety rules, and gradually improve skills through deliberate practice. These requirements correspond to five core dimensions: operational compliance, safety compliance, operational quality, work competence, and skill internalization.

In a typical setup, multi-view cameras are deployed around the training station to record the operational process. The data foundation layer collects operation videos, tool usage records, equipment status data, and evaluation records. The skill diagnostic layer identifies operation steps, verifies the correctness of disassembly or assembly sequences, confirms correct tool usage, and monitors safe behaviors (e.g., use of protective equipment, adherence to procedures). If the trainee repeatedly skips inspection steps, misuses tools, or violates safety requirements, the system records the type of error, timestamp, visual evidence, and possible causes.

Based on these diagnostic outputs, the target remediation layer recommends fault-based practical tasks, micro-lectures, standard operating procedure videos, and corrective technique exercises. For example, trainees who repeatedly encounter torque control errors may receive demonstration videos, troubleshooting exercises, and follow-up assessments; trainees with safety compliance issues will receive intensive safety training and must complete a critical safety task to continue learning. Finally, the results evaluation layer assesses improvements across five dimensions and feeds the results back into the system for further model optimization.

At the implementation level, a phased approach is recommended. The first phase is to build the practical skills indicator system and collect high-quality training data. The second phase is to develop the diagnosis model and prototype system, beginning with a limited number of representative tasks. The third phase is to develop remediation and assessment functions and conduct pilot courses. The fourth phase is to expand the model to more training tasks, improve data standards, and refine governance mechanisms. Corresponding safeguards are also needed, including training-base infrastructure, teacher digital literacy, data-use standards, ethical review procedures, and cooperation between colleges and enterprises.

6. Conclusion and Outlook

This paper addresses the difficulty of implementing adaptive instruction for practical skills in higher vocational rail transit education. Its main contribution is to extend learner-state data from online behavior to visual and multimodal data generated during practical operations, thereby making operational processes more observable, diagnosable, and improvable. On this basis, the paper proposes a computer vision-enabled application model organized around a collection-diagnosis-remediation-assessment closed loop. The model supports fine-grained diagnosis of operational compliance, safety compliance, movement quality, job competence, and skill internalization, as well as targeted remediation and continuous optimization of practical training.

This study proposes a conceptual framework for integrating computer vision into vocational training, with the core innovation of expanding data dimensions through a closed-loop model. However, this framework is still in the design phase and requires empirical verification. Recognition accuracy faces challenges such as occlusion, viewpoint, lighting, and task diversity, while issues such as annotation cost, generalization ability, privacy, and fairness also urgently need to be addressed. Future research should develop prototypes, conduct curriculum-based intervention studies, explore multimodal fusion and large-scale modeling to improve diagnostic performance and teaching effectiveness.

Acknowledgments

This work was supported by the Project of Young Innovative Talents of General Colleges and Universities of Guangdong Province (No. 2024KQNCX212), the Guangzhou Railway Vocational and Technical College Newly Introduced Talents Scientific Research Start-up Project (No. GTXYRC250105), and the Guangzhou Municipal Higher Education Teaching Reform Project 2025 (2025YBJG074).

References

[1] Bao X. D. Urban Rail Transit Present Situation and Future Development Trends in China: Overall Analysis Based on National Policies and Strategic Plans in 2016-2020. Urban Rail Transit, 2018, 4: 1-12.

[2] Lu K., Zhang L., Li S., Huang Y. P. Ding X, Hao J N, Huang S Q, Li X J, Lu F, Zhang H W. Urban Rail Transit in China: Progress Report and Analysis (2015-2023). Urban Rail Transit, 2025, 11: 1-27.

[3] Sun B. F., Liu J. J., Hao J. Y., Shen X. X., Mao X. H., Song X. M. Maintenance Decision-Making of an Urban Rail Transit System in a Regionalized Network-Wide Perspective. Sustainability, 2020, 12(22): 9734.

[4] Yu S. Q. The Key Path of Educational Digital Transformation. Journal of East China Normal University (Educational Sciences), 2023, 41(3): 62-71.

[5] Zhu Z. T., Hu J. The Practical Logic and Development Opportunities of Educational Digital Transformation. e-Education Research, 2022, 43(1): 5-15.

[6] Yang X. M., Zhang Y. How to Resolve the Contradiction Between Scale and Personalization in Education: The Logical Framework and Practical Path of Data-Driven Large-Scale Adaptive Instruction. Distance Education in China, 2022(8): 42-52+79.

[7] Liu H. Y., Zhang T. C., Wu P. W., Yu G.. A Review of Knowledge Tracing. Journal of East China Normal University (Natural Science), 2019(5): 1-15.

[8] Liu T. Y., Chen W., Chang L., Gu T. L. Research Advances in Knowledge Tracing Based on Deep Learning. Journal of Computer Research and Development, 2022, 59(1): 81-104.

[9] Wang Y., Zhu M. X., Yang S. H., Lu X. S., Zhou A. Y. A Survey and Performance Comparison of Deep Knowledge Tracing Models. Journal of Software, 2023, 34(3): 1365-1395.

[10] Yang W. Y., Yang Y. H. Focusing on Interpretability: A Review and Outlook of Knowledge Tracing Models. Modern Educational Technology, 2024, 34(5): 53-63.

[11] Shen Y. Y., Zhang Y. M., Shen Y. F. A Survey of Vision-Based Human Action Quality Assessment. Acta Automatica Sinica, 2025, 51(2): 404-426.

[12] Wu J. C., Wang L. M., Wu G. S. A Survey of Video Group Activity Recognition. Journal of Software, 2023, 34(2): 964-984.

[13] Zhang Y., Wen G. Z., Mi S. Y., Zhang M. L., Geng X.. A Survey of 2D Human Pose Estimation Based on Deep Learning. Journal of Software, 2022, 33(11): 4173-4191.

[14] Science and Technology Development Center, Ministry of Education. Construction Guideline for Demonstration Virtual Simulation Training Bases in Vocational Education. 2021.

[15] Han X. B., Shi W. R. Analysis of the Dilemmas and Solutions of Virtual Simulation Training in Vocational Education. e-Education Research, 2025(9): 50-56.

[16] Ericsson K. A., Krampe R. T. Tesch-Romer C. The role of deliberate practice in the acquisition of expert performance. Psychological Review,1993,100(3):363-406.