ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on Digital Twin Construction and Immersive Teaching Application of Luoyang Cultural and Tourism Heritage

Xuebo Li[1,*]

[1] Department of General Education, Luoyang Vocational College of Culture and Tourism, Luoyang, 471000, Henan, China

SUMMARY: *Under the background of parallel promotion of digital teaching and cultural heritage activation, although Luoyang cultural tourism heritage has a good digital acquisition foundation, there are still some problems such as scene reconstruction emphasizing display, insufficient semantic organization, and weak teaching adaptation. Therefore, this paper constructs a digital twin scene model of Luoyang cultural tourism heritage, and proposes a virtual-real fusion interactive application method for immersive teaching. In the scene construction task, the Longmen Grottoes, the White Horse Temple and the ruins of Luoyang City in the Sui and Tang Dynasties are selected as the objects, and the laser point cloud, oblique photography images, panoramic sequences and historical text data are fused. In the teaching application task, the classroom interaction data and training process data are introduced to comprehensively evaluate the system effect. The results show that the geometric construction accuracy of the proposed method in the three types of heritage scenes reaches 89.67%, 93.21%and 91.34%respectively, and the comprehensive geometric construction accuracy and semantic anchor matching rate reach 95.84%and 94.36%respectively under mixed data sources. In the immersive teaching application, the knowledge test accuracy and task completion rate reached 91.48%and 89.36%respectively, and the average interactive response time was 96.14 ms. The experimental results show that the proposed method improves the construction quality and teaching application effect of Luoyang cultural tourism heritage digital scene through the collaborative design of multi-source data fusion, semantic scene modeling and dynamic interactive scheduling.*

KEYWORDS: *Digital twin; Cultural tourism heritage; Immersive teaching; Virtual-real fusion interaction*

1 Introduction

With the continuous development of digital perception, 3D reconstruction, virtual-real interaction and intelligent rendering technology, the intervention mode of digital technology in the dissemination of cultural heritage has been significantly deepened. It not only changes the display form of cultural resources, but also plays a more and more direct role in heritage protection, knowledge dissemination and educational application, making the cultural content originally limited by space, time and site conditions obtain a visual, interactive and continuously updated expression channel [1]. Digital twin of cultural tourism heritage refers to the construction of virtual mirrors corresponding to real heritage objects in the digital space with the help of 3D modeling, semantic mapping, real-time data organization and visual

*lywlxylxb@126.com
https://doi.org/10.65102/is2026308
Published: 30 April, 2026

computing, and enables them to have state expression, scene interaction and knowledge carrying capabilities. On this basis, immersive teaching introduces virtual reality, augmented reality and multimodal interaction technology to place learners in a digital environment with a sense of context and task, so as to improve their understanding depth, participation intensity and cognitive continuity [2, 3].

Gabellone pointed out that digital twin provides a new observation perspective for the management and utilization of cultural heritage, and its value is not limited to three-dimensional representation, but to the transformation of heritage objects into a sustainable and updated digital system [1]. Boboc et al. reviewed the application progress of augmented reality in the field of cultural heritage in the past decade, and believed that AR technology has shown strong potential in site guide, cultural relic interpretation and public participation, but there is still the problem of uneven interaction depth between different scenes [2]. Aguerre discussed the open sharing and cross-cultural communication mechanism from the perspective of international dialogue on digital heritage, and pointed out that the digitization of cultural resources was not only a technical issue, but also related to the way of knowledge organization and the way of audience understanding [3]. Zhou et al. conducted a meta-analysis on the application of virtual reality and augmented reality in museum learning, and the results showed that immersive media could improve learning interest, knowledge retention and experience engagement to a certain extent [4]. Barbara has constructed an immersive virtual reality learning environment around the prehistoric intangible cultural heritage. Research results show that learners' sense of historical substitution and content memory are improved after the combination of scene reappearance and interactive tasks [5]. Innocente et al. further proposed an immersive XR technical framework for the field of cultural heritage, emphasizing that the system design needs to consider content authenticity, interaction logic and user cognitive load at the same time [6]. Alatrash et al. constructed a VR interpretation framework for engineering heritage communication, and pointed out that virtual immersive environment can strengthen users' understanding process of cultural meaning [7]. Silva et al. reviewed the application of AR in cultural heritage from the perspective of participatory activities, and believed that the interactive mechanism design directly affects the public communication effect of heritage content [8]. Paolanti et al. conducted a study on the evaluation of learning outcomes for virtual reality applications in digital cultural heritage, and found that the way of task organization in immersive environments would significantly affect learning outcomes [9]. Ch 'ng et al. proposed the communication path of cultural heritage from the perspective of social augmented reality, and pointed out that heritage communication in digital environment is changing from one-way display to social collaborative understanding [10]. A summary of the above related studies is shown in Table 1.

Table 1: Summary Table of relevant studies

Study	Research object	Index	Key findings	Limitation
Gabellone [1]	Cultural heritage digital twin	Conceptual value and application orientation	Proposed digital twin as a new perspective for heritage management and utilization	More conceptual than teaching-oriented
Zhou et al. [4]	VR/AR museum learning	Learning effect, engagement, knowledge retention	Meta-analysis showed immersive technology can improve museum learning outcomes	Focused mainly on museum context
Innocente et al. [6]	Immersive XR in heritage	Framework completeness, interaction design	Built a framework for XR use in cultural heritage domain	Lack of localized teaching validation
Paolanti et al. [9]	VR learning for digital heritage	Learning outcomes	Verified that VR applications can support heritage learning assessment	Weak linkage with digital twin modeling
Vuoto et al. [12]	Built heritage digital twin	Concept definition, conservation logic	Systematically clarified digital twin shaping path in built heritage conservation	Emphasis on conservation rather than instruction
Guo et al. [16]	XR-digital twin in heritage risk management	Situation awareness, comparative performance	Demonstrated the extension path from XR to digital twin in heritage management	Application focus is risk management, not classroom teaching

On this basis, Xu et al. proposed CubeMuseum AR interactive interface to improve the interactive experience of cultural heritage learning and museum gifts with tactile augmented reality. The results show that the combination of physical operation and digital narrative is helpful to improve user engagement [11]. Through a systematic review, Vuoto et al. sorted out the concept shaping process of digital twin of architectural cultural heritage in the field of protection, and pointed out that data integration, state mapping and multi-scale expression are the current research focuses [12]. Buragohain et al. discussed the opportunities and challenges of metaverse applications to promote the digitization of cultural heritage, and argued that immersive cultural space is moving from a display platform to a compound interactive ecology [13]. Taking the Mona Lisa as an object, Amelio and Zarri discussed the problem of visual narrative modeling in the digital twin of cultural heritage, indicating that the digital model not only needs geometric accuracy, but also needs narrative structure support [14]. Niccolucci and Felicetti studied the problem of digital twin sensors in the application of cultural heritage ontology, revealing the importance of linkage between semantic layer and perception layer [15]. Guo et al. compared the extension path of X-reality technology to the risk management of cultural heritage digital twin from the perspective of situation awareness, and believed that digital twin could enhance the information integration ability in complex environments [16]. Wang et al. conducted research on gamified virtual exhibition of cultural heritage, and pointed out that immersive mechanism can effectively stimulate users' exploration intention and continuous interactive behavior [17]. Chen et al. analyzed the reasons for users to use the augmented reality system of heritage museums from the socio-technical perspective, and the results showed that technology usability, cultural identity and contextual value jointly affected the use intention [18]. Corrales-Serrano et al. verified the usability, learning and emotion measurement questionnaire of virtual reality in the legacy teaching of higher education, indicating that immersive teaching evaluation tools are becoming standardized [19]. Fontal et al. constructed a measurement scale of heritage learning in the digital environment, emphasizing that the evaluation of learning effects should cover multiple dimensions of cognition, experience and value identification [20]. Wang and Meng established an explanatory model among visitors' cognitive identity, technical media and cultural symbols around public participation in the digital context of museums [21]. Cecotti et al. used fully immersive virtual reality to carry out art history course evaluation, and the results showed that the virtual environment could support more task-oriented learning judgment and process recording [22].

Based on the existing research, it can be found that the research on digital twin of cultural heritage, virtual reality display and immersive learning evaluation has made solid progress, but there is still room for further improvement. Most of the existing achievements focus on single heritage display, digital protection or museum interaction, and the research on system modeling of regional cultural tourism heritage is still insufficient. Many studies can prove that immersive technology can promote learning interest and quality of experience, but there is not enough discussion on "how digital twin scene directly serves curriculum organization, teaching interaction and knowledge construction". At the same time, the mapping relationship between real heritage space, digital scene semantics and classroom learning tasks has not been fully developed in the teaching application research of local cultural resources. For Luoyang cultural tourism heritage, this problem is particularly prominent. Luoyang heritage resources have rich types, complex historical levels and wide spatial distribution. If static display or single-point virtual browsing is still used, it is often difficult to present the era correlation, spatial logic and cultural evolution sequence between heritage nodes, and it is also difficult to meet the requirements of immersive teaching for interactive continuity and task guidance.

Based on the above problems, this paper proposes the construction method of digital twin scene of Luoyang cultural tourism heritage, and the interactive application method of virtual-real fusion for immersive teaching. This paper aims to realize the digital twin expression of Luoyang typical cultural tourism heritage resources by integrating 3D modeling, scene semantic organization, interactive trigger mechanism and teaching task design, and on this basis, improve the knowledge presentation effect, learning participation degree and application feasibility in immersive teaching.

The innovation points of this paper are mainly reflected in two aspects: (1) Combining the digital twin modeling of cultural heritage with the teaching needs of Luoyang local culture and tourism, constructing a digital scene with spatial expression, cultural semantics and teaching organization capabilities; (2) Around the immersive teaching process, a virtual-real integration interaction mechanism is designed to enable learners to form a more coherent cognitive path in scene roaming, node identification, knowledge triggering and feedback evaluation.

The main contributions of this paper include: (1) provide a set of computational, interactive and scalable scene construction ideas for the digital teaching transformation of Luoyang cultural tourism heritage; (2) Enhance the perceptibility and understandability of cultural heritage content in the classroom through the linkage design of digital twin and immersive teaching; (3) Provide a new technical path and practical reference for the integration and application of local cultural resources in smart cultural tourism education and digital heritage communication.

2 Methods and materials

In order to improve the spatial reduction accuracy, semantic expression ability and teaching adaptation level of Luoyang cultural tourism heritage digital resources, this paper constructs the digital twin scene model of Luoyang cultural tourism heritage, and uses it as the basic support for the subsequent immersive teaching application. In this method, digital twin is not understood as a simple 3D modeling result, but the legacy entity, historical semantics, scene state and teaching content are unified into the same computing framework, so that the digital scene can not only correspond to the real heritage space, but also assume the functions of knowledge organization, node interaction and teaching call.

2.1 Construction method of digital twin scene of Luoyang cultural tourism heritage

The cultural tourism heritage of Luoyang has various types, including the physical heritage such as Longmen Grottos, White Horse Temple, and the ruins of Luoyang city in the Sui and Tang Dynasties, as well as cultural information such as the system of rites and music, the pattern of the capital, the art of stone carving and local folk customs. There are two prominent problems in the digitization process of such resources. One is that the data sources are complex, and the laser point cloud, oblique photography image, panoramic video, historical image and interpretation text have obvious differences in scale, format and semantic granularity. The other is that the existing digital display mostly stays at the static browsing level. Although the model is visual, it is difficult to support the route organization, knowledge trigger and scene explanation in teaching. Based on this, this paper proposes a digital twin scene construction method for Luoyang cultural tourism heritage, whose core goal is to complete multi-source data registration, scene semantic fusion and teaching node mapping under a unified coordinate framework. The overall structure is shown in Figure 1.

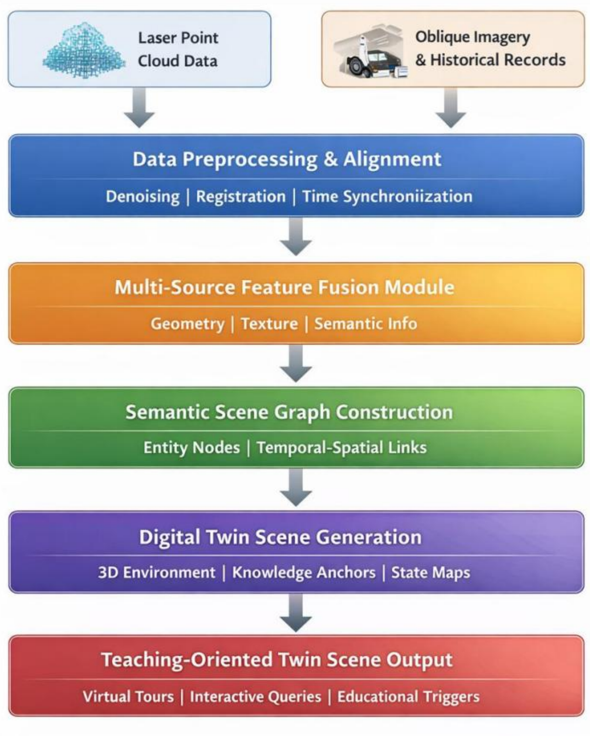

Figure 1: Construction framework of digital twin scene of Luoyang cultural tourism heritage

In Figure 1, the underlying input is composed of heritage geometry data, image data and historical document data. Firstly, the system de-noised the data from different sources, unified the coordinates and synchronized the time, and then formed the unified representation of the heritage object through the multi-source feature fusion module. Then, the scene semantic graph is constructed according to the entity attributes and spatial semantic relationships. On this basis, the digital twin scene that can be used for teaching is generated, and the teaching scene unit that can be roamed, searchable and triggered is output.

In the stage of geometric layer construction, the joint registration method of point cloud and image is used to complete the reconstruction of heritage space. Suppose the point cloud set obtained by laser scanning is $\mathcal{P} = \{p_i\}_{i=1}^{N}$, and the feature point set recovered by oblique photography is $\mathcal{Q} = \{q_i\}_{i=1}^{N}$. In order to make the two types of data into a unified space, the rigid transformation matrix T=[R | t] is defined, where R is the rotation matrix and t is the translation vector, then the registration target can be expressed as:

$$T^* = \arg\min_{R,t} \sum_{i=1}^{N} \|Rp_i + t - q_i\|_2^2 \tag{1}$$

Equation (1) is used to obtain the optimal spatial alignment result. After the registration is completed, the system projects the multi-view image texture onto the point cloud surface to form an initial scene mesh with both geometric details and appearance information. The completion strategy based on neighborhood consistency is used to correct the local sparse regions caused by occlusion and the change of acquisition distance in the point cloud. Let the neighborhood of node p_i be $\mathcal{N}(p_i)$, then its smooth update result is:

$$\hat{p}_i = \frac{1}{|\mathcal{N}(p_i)|} \sum_{p_j \in \mathcal{N}(p_i)} p_j \tag{2}$$

Equation (2) can weaken local discrete noise while preserving the overall contour, and provide a stable geometric basis for subsequent semantic attachment. Spatial morphology alone is still not enough to constitute a digital twin scene. For Luoyang cultural tourism heritage, the digital scene must also express the historical attributes, cultural functions and teaching value of heritage objects. Therefore, this paper represents each heritage unit as a joint vector of geometric features, texture features and semantic features. Let the three types of features of the ith heritage unit be g_i、t_i and s_i, respectively, then its fusion representation h_i is defined as:

$$h_i = \alpha_i^g g_i + \alpha_i^t t_i + \alpha_i^s s_i \tag{3}$$

Here, α_i^g, α_i^t and α_i^s are the weight coefficients of different modes, which are satisfied:

$$\alpha_i^m = \frac{\exp(w_m^T f_i^m)}{\sum_{n \in \{g,t,s\}} \exp(w_n^T f_i^n)}, \quad m \in \{g, t, s\} \tag{4}$$

Equation (3) and Equation (4) jointly complete the adaptive fusion of multi-source heterogeneous information. Compared with simple concatenation, this method can dynamically adjust the importance of different information sources according to the specific characteristics of legacy objects. For the grotto objects with rich details, the geometric features and texture features have higher weights. For site nodes with dense historical narratives, the proportion of semantic features will be increased accordingly. Its calculation process is shown in Figure 2.

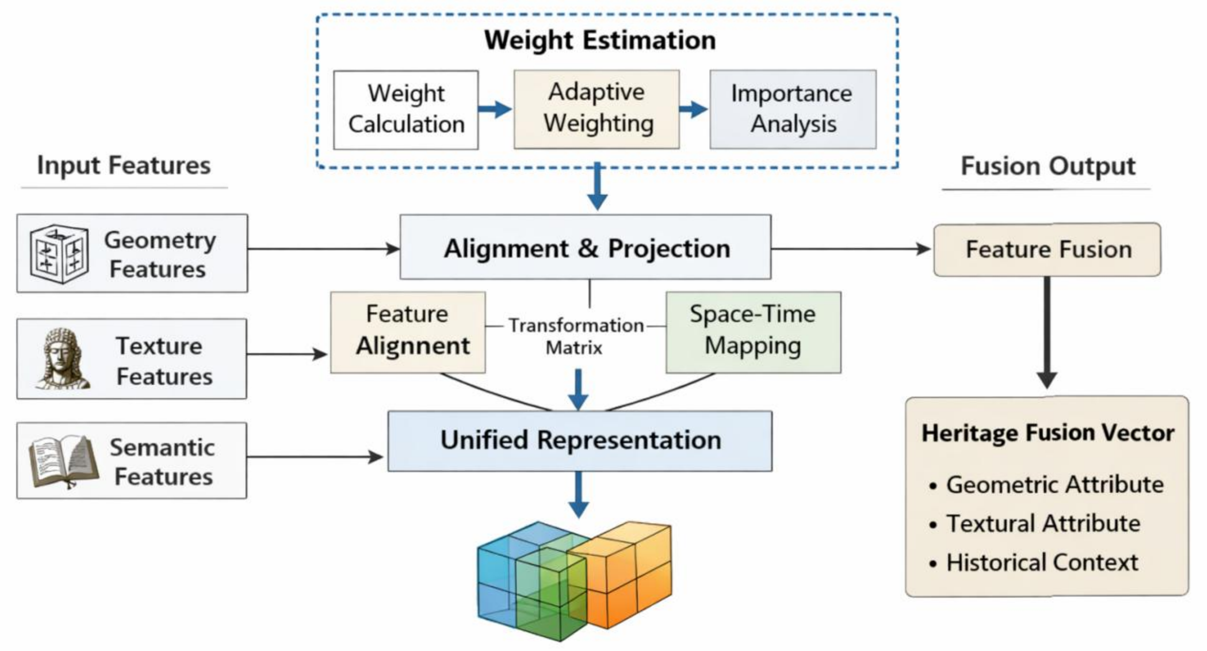

Figure 2:Multi-source feature fusion process of legacy unit

After obtaining the fusion representation, the system does not directly organize the scene into several isolated models, but further constructs a semantic scene graph for cultural interpretation and teaching invocation. Let the digital twin scene graph be:

$$G = (V, E, A) \tag{5}$$

Here, V represents the set of legacy entity nodes, E represents the set of relationship edges between nodes, and A represents the set of node attributes. For any two nodes v_i and v_j, we calculate their relationship strength from three dimensions: spatial proximity, historical relevance and functional similarity:

$$r_{ij} = \beta_1 \, Sim_{spa}(v_i, v_j) + \beta_2 \, Sim_{his}(v_i, v_j) + \beta_3 \, Sim_{fun}(v_i, v_j) \tag{6}$$

Here, β1+β2+β3=1. Equation (6) makes the digital twin scene not only a geometric aggregation result, but a knowledge-based space with traceable semantic relations. Taking the scene of Longmen Grottoes as an example, caves, statues, inscriptions, dynasty information, and artistic styles can be encoded as different types of nodes and edges, so that learners can not only "see" the object when entering the scene, but also understand its historical evolution and cultural function along the relationship path established by the system.

Considering that the digital twin scene is to serve the subsequent immersive teaching, this paper further establishes the mapping mechanism between the legacy entity node and the teaching knowledge unit. Let the set of teaching knowledge units be $\mathcal{C} = \{c_k\}_{k=1}^{M}$, then the attention weight of the KTH knowledge unit to the scene node vi is defined as:

$$a_{ik} = \frac{\exp(\, q_k^T W h_i)}{\sum_{j=1}^{N} \exp(\, q_k^T W h_j)} \tag{7}$$

Here, q_k is the semantic query vector of the KTH teaching unit, and W is the learnable mapping matrix. Then the representation result of this teaching unit in the scene is:

$$c_k{}' = \sum_{i=1}^{N} a_{ik} h_i \quad (8)$$

Equations (7) and (8) complete the directed association between "legacy object - knowledge content - teaching scene". In this way, teachers do not need to re-split scattered resources when organizing courses, but can directly call the set of scene nodes highly related to a teaching topic. For example, when the course topic is "Localized Expression in the Eastern Transmission of Buddhist Art", the system can prioritize the activation of nodes related to the morphology of cave statues, inscription information, and time background, and automatically generate the corresponding explanation anchor and walkthrough sequence.

In order to ensure that the digital twin scene can be continuously updated with the change of data supplement and teaching needs, this paper introduces consistency constraints in the scene layer. Let the geometric reconstruction error, semantic mapping error and teaching matching error be L_{geo}, L_{sem} and L_{tea}, respectively, then the overall optimization objective is:

$$L = \lambda_1 L_{geo} + \lambda_2 L_{sem} + \lambda_3 L_{tea} \quad (9)$$

Here, λ_1, λ_2 and λ_3 are the weight coefficients. Equation (9) implies that the optimization of digital twin scenarios does not only pursue geometric accuracy, but also requires that semantic organization be coordinated with pedagogical use. This is obviously different from the general 3D display model of cultural heritage: the latter emphasizes visual fidelity, while the scene constructed in this paper needs to have spatial authenticity, semantic interpretability and teaching callability at the same time. In summary, Figure 3 shows the structure of the proposed digital twin scene construction method for Luoyang cultural tourism heritage.

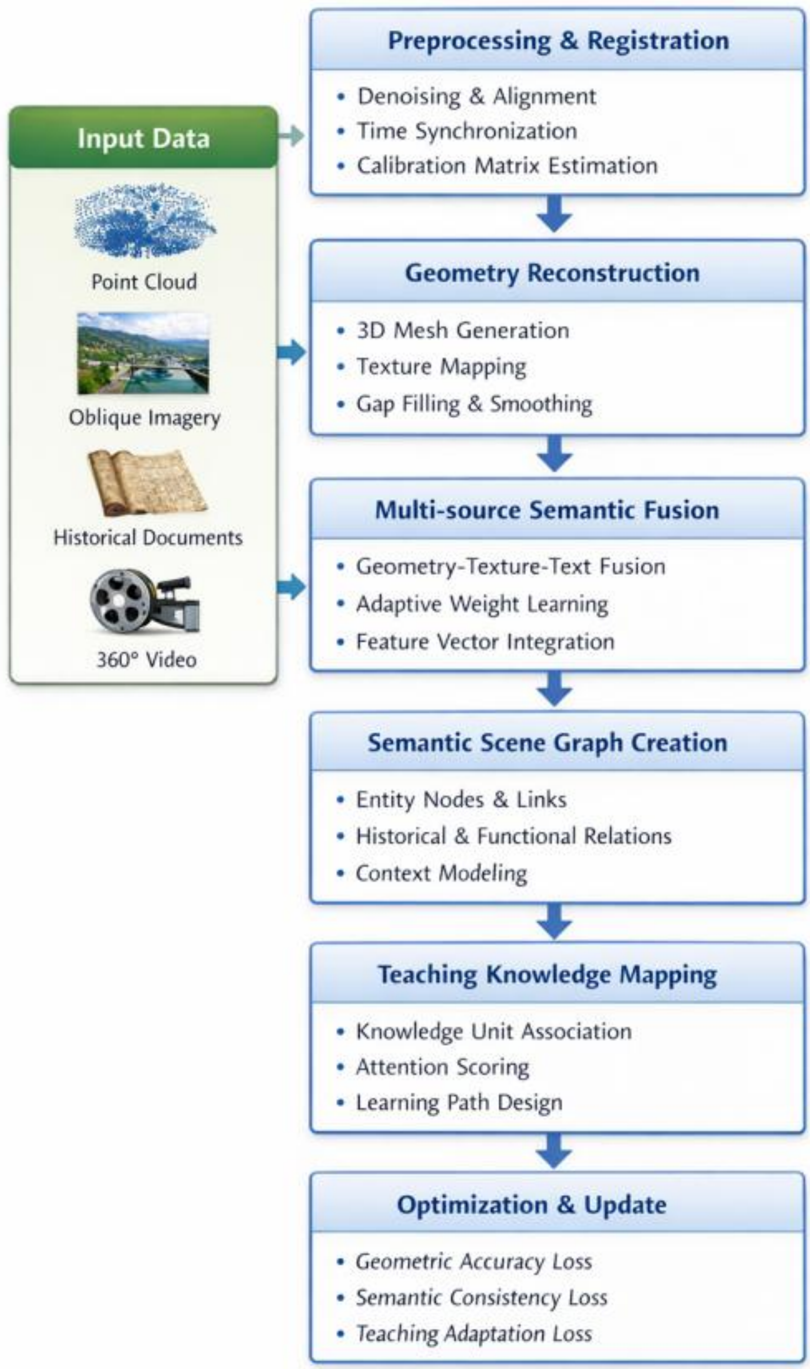

Figure 3: Structure diagram of digital twin scene construction method of Luoyang cultural tourism heritage

Figure 3 shows that the proposed method takes multi-source heritage data as input and completes geometric reconstruction after preprocessing and unified registration. Then, texture, historical text and heritage attributes are introduced into the fusion module to form a heritage node representation with semantic discrimination ability. Then, through scene graph modeling and teaching unit mapping, the static digital model was transformed into a digital twin scene that could serve the teaching organization. The resulting scene can not only restore the spatial structure of Luoyang cultural tourism heritage more completely, but also provide a computable basis for path design, node interaction and knowledge triggering in subsequent immersive teaching.

2.2 The Interactive application Method of Virtual-Real Integration for Immersive Teaching

The digital twin scene of Luoyang cultural tourism heritage constructed in the previous section can complete the unified expression of heritage entity, spatial structure and historical semantics, so as to provide a more stable scene basis for the digital display of cultural resources. However, this scenario model mainly solves the problem of "how to accurately construct and organize legacy resources", and cannot directly answer "how to enter the teaching process and respond to learning behaviors". If the immersive teaching still stays at the level of fixed script playing, preset route wandering and one-way information pop-up, although the digital twin scene has a high visualization level, it is difficult to truly form an interactive guidance oriented to learning goals, and it is difficult to dynamically adjust the content presentation method according to students 'knowledge base, behavior state and classroom rhythm. Based on this, this paper further proposes a virtual-real fusion interactive application method for immersive teaching on the basis of digital twin scene. By jointly perceiving learners' position, sight, operation and feedback information in the real teaching environment, and mapping them to knowledge nodes, explanation hotspots and task units in the digital twin scene, a teaching interaction model with state perception, node matching, path scheduling and feedback update capabilities was constructed.

In the immersive teaching of cultural heritage, the interactive objects are not only the virtual scene itself, but also the teachers, students, teaching tasks, classroom equipment and real space environment. Therefore, this paper defines the interaction process as a composite system composed of "real teaching side-digital twin side-fusion scheduling side". The real teaching side is responsible for collecting learners 'operation behavior, gaze stay, answer results and teacher control instructions in the classroom. The digital twin side is responsible for providing heritage scene models, knowledge anchors, interactive hotspots and explanation resources. The fusion scheduling side completes content matching, path triggering and feedback updating according to the current learning state. Its structure is shown in Figure 4.

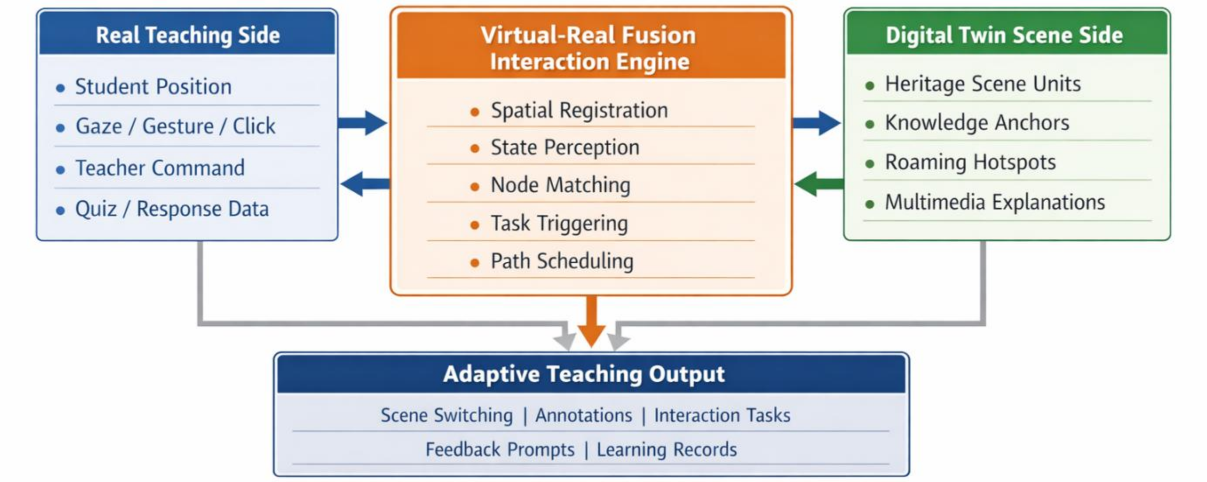

Figure 4: Schematic diagram of the interactive structure of virtual-real fusion immersive teaching

In Figure 4, the system does not treat the real classroom and the virtual scene separately, but treats them as a two-way channel for sustainable information exchange. Learners' head rotation, eye focus, handle click, dwell time and question answering performance in the real environment will be transformed into computable features through the state perception module. At the same time, cave units, building nodes, historical event cards and multimedia explanation resources in the digital twin scene enter the interaction engine in the form of semantic nodes. After the fusion and scheduling of the two, the system outputs the corresponding explanation stacking, path switching, task triggering or prompt feedback. The purpose of this process is to avoid immersive teaching to be reduced to a simple "scene watching", and transform it into a computational process that can dynamically organize the content according to the learning process.

In order to make the interaction in the real space consistent with the cultural nodes in the virtual scene, we first establish the coordinate mapping relationship between the virtual space and the real space. Let the learner's interaction position in the real teaching space be p_t, and its corresponding position in the digital twin scene be $\hat{p}_t$, then the mapping relationship between them is denoted by:

$$\hat{p}_t = R_{vr} p_t + t_{vr} \tag{10}$$

Here, R_{vr} represents the rotation transformation matrix between virtual and real space, and t_{vr} represents the translation vector. Equation (10) is used to complete the spatial alignment between the real interaction points and the anchor points of the digital scene, so that the observation direction and operation area of students in the classroom can be consistent with the explanation nodes in the virtual scene. This consistency is particularly important for teaching tasks such as reading the details of the Longmen Grottoes statues, discriminating the space of the White Horse Temple courtyard, or comparing the pattern of the Sui and Tang capitals. If the mapping relationship is not stable, the hotspots triggered by learners in the scene may deviate from the preset content of teachers, thereby weakening the continuity of immersive teaching.

After the spatial alignment is completed, the system needs to further identify the current learning state of the learner. Since immersive teaching is not a process driven by a single click behavior, this paper represents the student's interaction state as a state vector composed of gaze, stay, answer performance, operation behavior and teacher's instructions. Let the learner state at time t be u_t, then:

$$u_t = [\phi_g(g_t), \phi_d(d_t), \phi_q(q_t), \phi_a(a_t), \phi_c(c_t)] \tag{11}$$

g_t represents the sight focus feature, d_t represents the stay time on the scene node, q_t represents the instant test or question answering results, a_t represents the operation behaviors such as click, drag, zoom and roam, and c_t represents the phased instructions issued by the teacher console. Let $\phi(\cdot)$ denote the corresponding feature encoding function. The function of Equation (11) is to unify the scattered teaching behaviors into the state representation of the input interaction model. Different from the traditional digital courseware that only records "whether to click", this paper emphasizes the multi-dimensional perception of the learning process, because the understanding difficulties in cultural heritage teaching are often not only manifested by answer errors, but also may be manifested by frequent eye switching, insufficient stay at key nodes, or lack of continuous attention to important image details.

After obtaining the learner state, the system needs to decide which scenario node or teaching task should be activated preferentially at the present time. To this end, this paper establishes a matching score function between learning states and scene nodes. Let the representation of the JTH digital twin scene node be v_j, its teaching priority weight be w_j, and its current accessibility be o_{tj}, then the matching score between learner and node j at time t is defined as:

$$m_{tj} = \alpha \cdot \cos(W_u u_t, W_v v_j) + \beta w_j + \gamma o_{tj} \tag{12}$$

Here, W_u and W_v are the mapping matrices between state vector and node vector, respectively, and α, β, γ are the weight coefficients. Equation (12) shows that whether a node is activated depends not only on where the student is currently browsing, but also on the teaching importance of the node in this section of the course and whether it is suitable to be triggered at the current stage. For example, under the theme of "modeling features of Northern Wei Dynasty stone carvings", the system will increase the weights of nodes related to representative caves, clothing pattern depictions and image proportions. Under the theme of "Capital Space and Ritual Order", the system will give priority to matching scene units such as Miyagi, city gates, and central axis layout. In this way, the digital twin scenario is no longer a static repository of resources, but a knowledge space that can be adaptively organized around teaching topics.

On the basis of node matching, the system also needs to select the next interaction path from multiple candidates. If the current score is only used to jump point by point, the teaching process is easy to be fragmented, and students may also lose knowledge continuity in high-frequency switching. Therefore, this paper further introduces path scheduling mechanism after node screening. Suppose the set of candidate nodes at the current time is $\mathcal{C}_t$, the expected knowledge gain brought by the JTH candidate node is g_j, and the interaction cost of switching from the current node to the node is c_{tj}, then the optimal target node j^* is selected by:

$$j^* = \arg\max_{j \in \mathcal{C}_t} \left(m_{tj} + \lambda g_j - \mu c_{tj} \right) \tag{13}$$

where λ and μ are the tuning parameters. Equation (13) illustrates that the system does not simply pursue the "most relevant" node, but seeks a balance between relevance, knowledge gain, and interaction burden. In this way, students can avoid frequently entering multiple scene blocks in a short period of time, which causes cognitive interruption. For the teaching of cultural heritage, this is very critical, because the heritage content often depends on the intersection of time line, space line and cultural line. Without the rhythm control of path arrangement, it is difficult for learners to establish a complete understanding framework.

In order to enable the system to continuously revise the interaction strategy according to the actual performance of students, this paper further sets up the learning feedback update mechanism. Feedback in immersive teaching includes both explicit results such as answer scores and task completion, as well as teacher observation evaluation and interaction quality recorded by the system. Therefore, let the completion result of the student for the current node task be y_t, the teacher's feedback be f_t, and the knowledge mastery degree at the previous moment be k_t, then the updated mastery degree is:

$$k_{t+1} = \eta k_t + (1 - \eta)(\delta_1 y_t + \delta_2 f_t) \tag{14}$$

where η is the historical state retention coefficient, δ_1 and δ_2 are the weights of explicit grades and teacher evaluations. Equation (14) is used to re-incorporate the instant classroom performance into the learner state estimation, so that the system can automatically adjust the content difficulty, hint density and explanation depth in the subsequent node matching. If the student continued to make mistakes in the recognition task of a node, the system could reduce the frequency of path switching and increase the local detail amplification, voice explanation or comparison diagram. If students show a high degree of mastery, they can shorten the basic explanation session and move to more comprehensive scene comparison or inquiry tasks. Based on the above mechanism, the virtual-real fusion interactive application framework for immersive teaching constructed in this paper is shown in Figure 5.

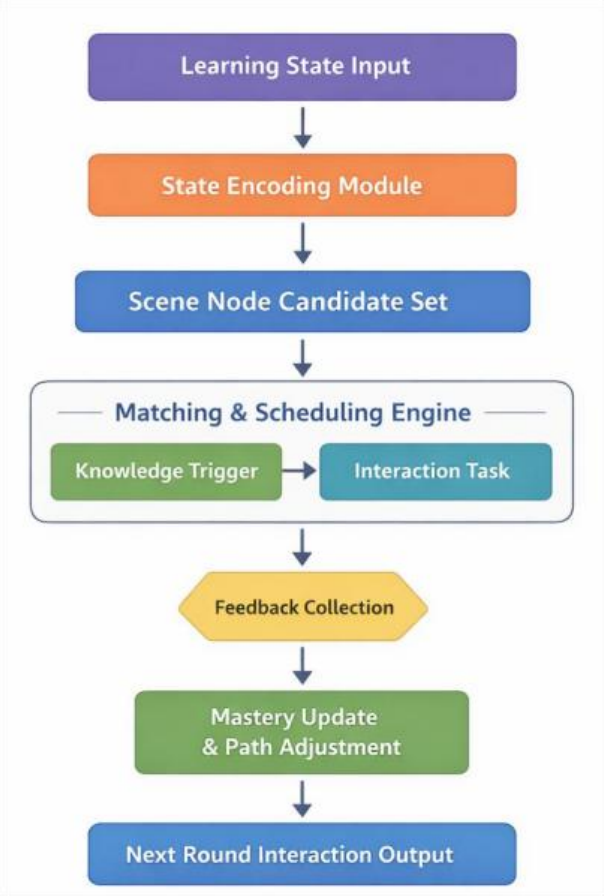

Figure 5: Framework diagram of interactive application method of virtual-real fusion for immersive teaching

Figure 5 shows that the proposed method takes the learner state input as the starting point, and matches with the set of candidate nodes in the digital twin scene after state encoding. Then, the scheduling engine decides the current knowledge content and interactive tasks that should be triggered. The feedback information is collected after the task is completed, which is used to correct the grasp estimation and the subsequent path arrangement. The key difference between the proposed approach and general virtual cultural display systems is that it does not understand interaction as isolated action triggering, but as a dynamic scheduling problem in a continuous learning process. The role of digital twin scene here is not only a visual carrier, but also a running platform for the coupling of teaching content, spatial structure and interaction logic.

3 Results

In this paper, the digital twin scene model of Luoyang cultural tourism heritage is constructed, and the virtual-real integration interactive application method for immersive teaching is designed, but its actual effect still needs to be further verified by experiments. Based on this, this paper first analyzes the construction effect of digital twin scene, and then investigates its application feasibility in immersive teaching.

3.1 Effect analysis of digital twin scene construction

In order to test the applicability of the proposed digital twin scene construction method in Luoyang cultural tourism heritage, this paper completes the system implementation in the environment of Python 3.11, PyTorch 2.1, Open3D and Unity 2022. The experimental platform is Intel Core i7-12700, 32 GB RAM, NVIDIA RTX 3080, and the operating system is Ubuntu 22.04. Three representative scenes including Longmen Grottoes, White Horse Temple and Luoyang City ruins in the Sui and Tang Dynasties were selected as experimental objects, corresponding to high-density stone detail scenes, courtyard architecture scenes and large-scale site scenes. The data sources include 18 groups of laser point clouds, 6420 oblique photographic images, 126 groups of panoramic sequences, and 312 structured historical text records. After the acquisition, the system removes the duplicate segments, performs statistical filtering on the abnormal points, and unifies the multi-source data into the same local coordinate frame. At the same time, the brightness correction of the image texture is carried out, and the entity extraction and attribute standardization of the historical text are performed. In the experiment, the training set, validation set and test set were divided by 8:1:1, the number of iterations was set to 220, the feature dimension was set to 128, the Dropout was set to 0.2, and Adam was used as the optimizer. The evaluation metrics include Geometric Construction Accuracy (GCA) and semantic anchor matching rate (SAA), the former is used to measure the consistency between the scene geometry recovery and the reference model, and the latter is used to measure the correspondence accuracy between explanation hotspots, historical nodes and scene units.

To analyze the influence of different parameter Settings on the model performance, the learning rate is set to 0.01, 0.001 and 0.0001, and the BatchSize is set to 4, 8, 16 and 32, and the results are shown in Table 2. It can be seen that when the learning rate is too high, the geometric reconstruction process has obvious oscillation, and the semantic mapping results are more susceptible to local noise interference. If the learning rate is too low, the model will update slowly and it will be difficult to fully capture the fine-grained differences between different legacy units. When the learning rate is 0.001 and BatchSize is 8, the model performance is the best, GCA reaches 95.84% and SAA reaches 94.36%. This result shows that the parameter combination can better coordinate the geometric recovery stability and semantic fusion ability, and provide suitable training conditions for subsequent experiments.

Table 2: Results of hyperparameter sensitivity analysis

Learning rate	BatchSize	GCA / %	SAA / %
0.01	8	90.26	88.41
0.001	8	95.84	94.36
0.0001	8	92.17	90.22
0.001	4	95.21	93.74
0.001	16	94.67	92.85
0.001	32	93.48	91.63

Under the condition of single scene data, the proposed method is compared with three common scene reconstruction methods, Photogrammetry, Point-cloud Meshing and NeRF, and the results are shown in Figure 6. Figure 6 shows the GCA performance of the four methods in three types of scenes: Longmen Grottots, White Horse Temple, and Luoyang City ruins in the Sui and Tang Dynasties. It can be seen that the GCA of the proposed method in the three types of scenes reaches 89.67%, 93.21% and 91.34%, respectively, which are higher than the other three methods. Among them, in the scene of Longmen Grotto, the proposed method is 9.32 percentage points higher than Photogrammetry. In the White Horse Temple scene, it is 4.47 percentage points higher than Point-cloud Meshing. In the scene of Luoyang city ruins in the Sui and Tang Dynasties, it still has an advantage of 4.06 percentage points compared with NeRF. The results show that the proposed method can maintain a relatively stable geometric construction ability when facing different objects with dense stone details, continuous courtyard space or large site scale. This is mainly due to the unified registration and semantic constraint fusion mechanism of multi-source data, which makes the geometric structure recovery no longer rely on a single image texture or local mesh fitting, thus reducing the influence of edge breakage, texture drift and large-scale regional homogenization.

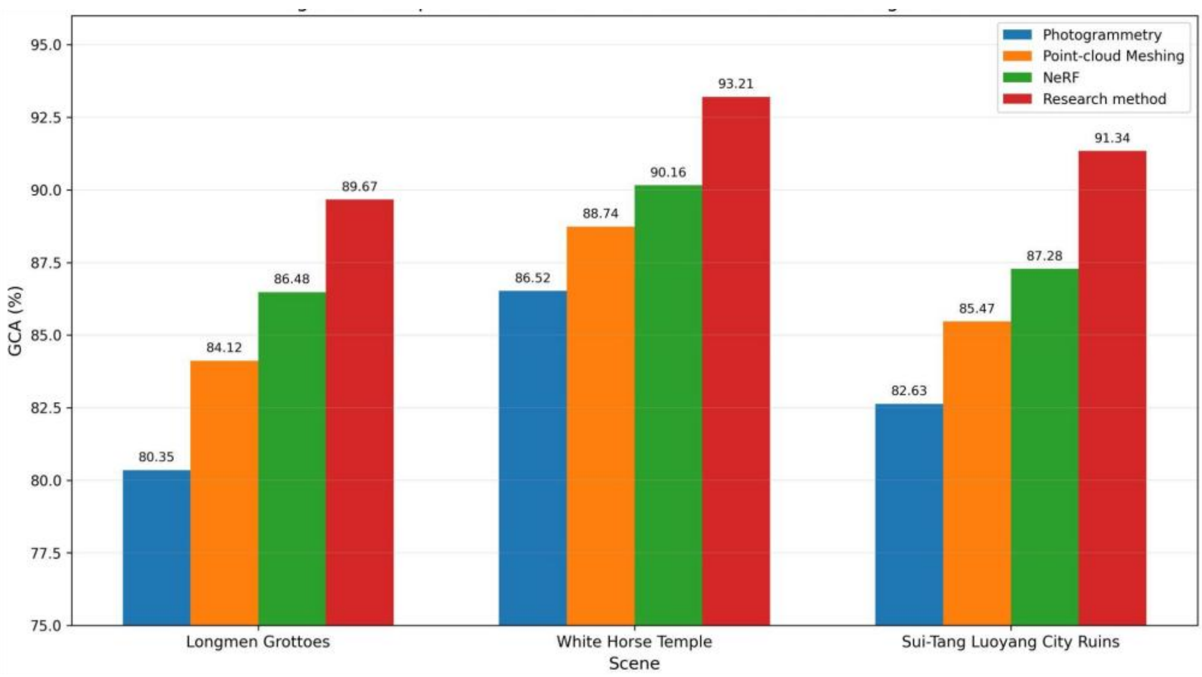

Figure 6: Comparison of geometric construction accuracy of four methods in three types of heritage scenes

After fusing all data sources, this paper further compares the comprehensive construction performance of the four methods, and the results are shown in Figure 7. The experimental results show that the GCA of the proposed method is 95.84%, which is significantly higher than 90.27% of NeRF, 86.74% of Point-cloud Meshing and 83.18% of Photogrammetry. In terms of SAA index, the proposed method reaches 94.36%, which is also significantly ahead of the other three methods. This result shows that the performance of different methods is improved after joint input of multi-source data, but the improvement of the proposed method is larger. The reason is that the scene of Luoyang cultural tourism heritage requires the model not only to recover the "form", but also to accurately organize the "meaning". For general 3D reconstruction methods, the improvement of geometric accuracy does not necessarily bring about the stable attachment of semantic nodes. However, the proposed method introduces the joint modeling among historical texts, spatial units and teaching knowledge points in the scene reconstruction stage, so it performs better in the matching rate of semantic anchors. This means that the constructed digital twin scene is not only suitable for display, but also more suitable for subsequent teaching calls.

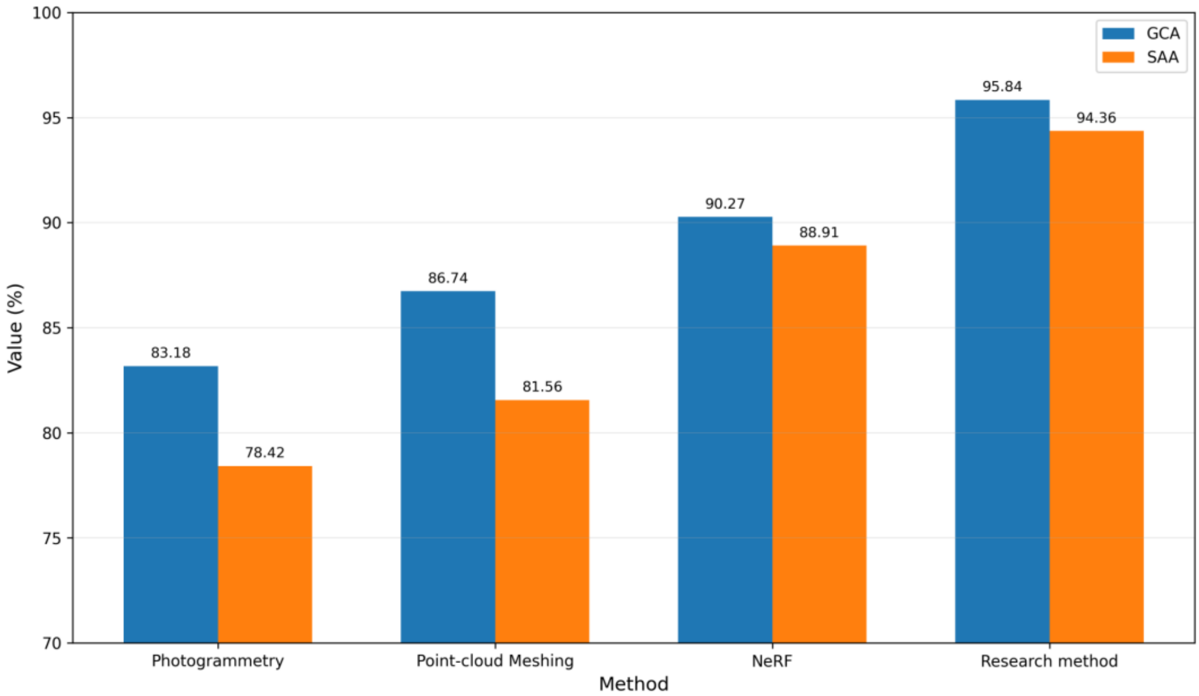

Figure 7: Comparison of the comprehensive geometric construction accuracy and semantic anchor matching rate of the four methods under mixed data sources

To further verify the stability of the proposed method in different scenarios, the semantic anchor matching rates of three types of scenarios are counted. The results show that the SAA of the Longmen Grottoes scene is 92.48%, the White Horse Temple scene is 95.17%, and the Luoyang city site scene of the Sui and Tang Dynasties is 93.86%. The matching rate of the White Horse Temple scene is slightly higher, indicating that the courtyard-style building is easier to form a stable mapping in terms of boundary identification, functional partition and knowledge node attachment. Although the scene of Longmen Grottoes has more complex local carving details, the system can still maintain a matching level of more than 92% due to the highly dense relationship between statues, inscriptions and caves. The spatial span of the site scene is large, the local remains are more incomplete, and the semantic organization is more difficult, but the overall result is still relatively stable. It can be seen that the proposed method does not show obvious performance imbalance due to the difference of scene types, and has good cross-scene adaptability.

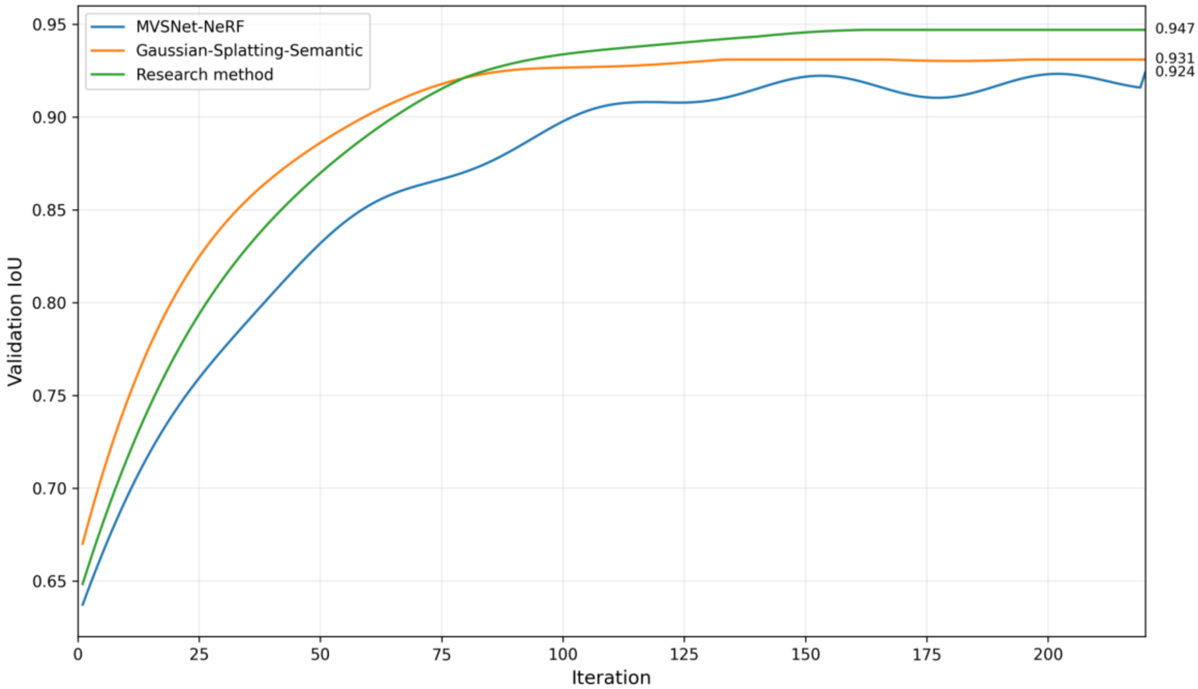

Figure 8: Comparison of convergence performance of the three methods on the validation set IoU

In order to test the convergence and generalization ability of the model, this paper compares it with MVSNet-NeRF and Gaussian-Splatting-Semantic two strong baseline models, and the results are shown in Figure 8. Figure 8 uses a single line chart to show the IoU convergence process of the three methods on the validation set. The results show that the proposed method tends to be stable after about 140 iterations, and the IoU of the validation set finally reaches 0.947. The final IoU of Gaussian-Splatting-Semantic is 0.931. Although the convergence is fast in the early stage, the improvement is slow in the late stage. The final IoU of MVSNet-NeRF is 0.924, and there are still obvious fluctuations in the middle and late stages. This phenomenon shows that the proposed method is not only better than the comparison model in the final construction effect, but also has better performance in training stability. This is particularly important for the digital twin scene of cultural heritage, because the lack of stable convergence process in scene construction often leads to inconsistent attachment positions of teaching nodes, which affects the triggering accuracy of subsequent immersive interaction.

3.2 Application effect and feasibility analysis of immersive teaching

In order to verify the feasibility of the proposed virtual-real fusion interactive application method in the immersive teaching of Luoyang cultural tourism heritage, this paper selects the "Luoyang cultural heritage digital teaching classroom dataset" and the "cultural tourism training extended dataset" to carry out tests. The former is derived from the 4-week classroom teaching experiment, including 128 students' gaze stops, node clicks, task submission, test scores and teacher intervention records in three theme modules: Longmen Grottoes, White Horse Temple and Luoyang City ruins in the Sui and Tang Dynasties. The latter came from the teaching activities that combined the on-campus training environment with the off-campus display environment, covering 96 students and 42 groups of complete interaction processes. Both types of datasets are divided into training, test and validation sets according to 8:1:1. In the experiment, BatchSize is set to 64, learning rate is set to 0.001, state vector dimension is set to 128, interaction threshold window is set to 12 s, path scheduling depth is set to 3, and AdamW is used as the optimizer. In order to reflect the advantages of the method, this paper takes traditional multimedia teaching, panoramic VR display, common 3D walkthrough teaching and interactive teaching based on rule trigger as the comparison methods. The evaluation metrics included knowledge test accuracy, task completion rate, and average interaction response time, where higher values for the first two items were better, and lower values for the latter were better.

The overall results of the five methods are shown in Table 3. It can be seen that the proposed method achieves the best results in the two indicators of knowledge test accuracy and task completion rate, reaching 91.48%and 89.36%respectively, which are significantly higher than the other four control methods. Among them, rule-based interactive teaching achieves 84.62%and 82.75%respectively in two indicators, which is better than ordinary 3D walkthrough teaching, but still lags behind the method proposed in this paper. This shows that although only relying on fixed rules and preset nodes can improve the problem of insufficient interaction in traditional demonstration teaching, it is difficult to adjust the content push in time according to the change of students 'status. Through the linkage of learning state perception, node matching and path scheduling, the method in this paper establishes a closer correspondence between scene call and teaching demand. It should be mentioned that the average interaction response time of the method in this paper is 96.14 ms, which is slightly higher than that of the other methods, but it is still within the acceptable range of immersive classroom interaction, indicating that the method still maintains good real-time performance while improving the teaching effect.

Table 3: Comparison of immersive teaching application effects of five methods

Method	Knowledge Quiz Accuracy / %	Task Completion Rate / %	Average Interaction Response Time / ms
Traditional Multimedia Teaching	71.36	68.42	41.27
Panoramic VR Presentation	78.54	74.63	58.31
Conventional 3D Roaming Teaching	82.17	79.84	76.25
Rule-triggered Interactive Teaching	84.62	82.75	88.49
Proposed Method	91.48	89.36	96.14

The correct rate of knowledge test can directly reflect students' understanding of cultural heritage content, so this paper further takes this index as the core to compare the performance of different methods in three types of teaching scenarios, and the results are shown in Figure 9. In the teaching scene of Longmen Grotto, the knowledge test accuracy of the method in this paper reaches 92.35%. In the teaching scene of White Horse Temple, the value is 90.84%. In the teaching scene of Luoyang city ruins in the Sui and Tang Dynasties, the value is 91.26%. All three types of scenarios are higher than the other methods. In contrast, the improvement in the scene of Longmen Grottoes is more obvious. The reason is that the details of the grottoes are dense and the historical information is layered, and the traditional display method is easy to cause students to only see the part and not know the context. In the White Horse Temple scene, the courtyard relationship and building function partition are relatively clear, and the gap between different methods is slightly smaller, but the method in this paper still keeps ahead. The performance in the site scene is also stable, indicating that the method does not rely on a single cultural space type, and has good cross-scene adaptability.

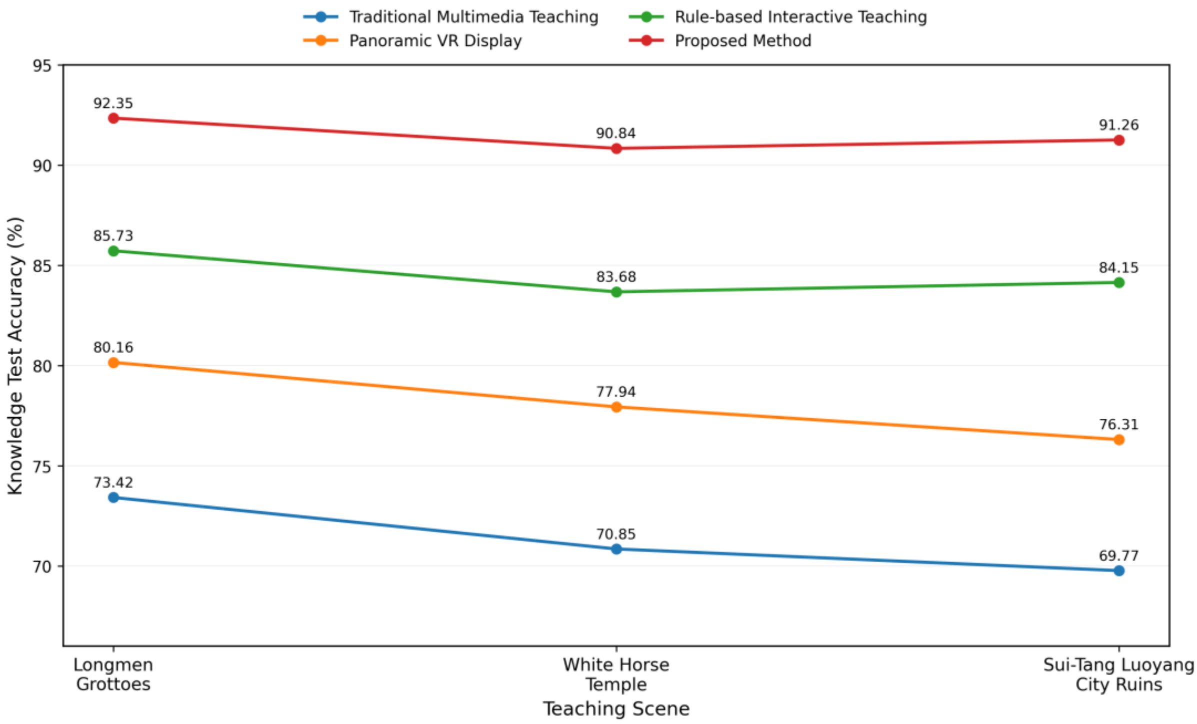

Figure 9: Comparison of correct rate of knowledge test

In addition to the correct rate of knowledge test, this paper also calculates the coverage of knowledge nodes to evaluate the richness of content access in the teaching process. Under the condition of the same 20 effective interactions, the knowledge node coverage of traditional multimedia teaching is only 0.46, panoramic VR display is 0.58, ordinary 3D walkthrough teaching is 0.67, rule-based trigger interactive teaching is 0.74, and the method in this paper reaches 0.83. The results show that the proposed method does not concentrate the interactive resources on a small number of high-frequency nodes, but can balance the scheduling among multiple knowledge units according to the learning state, so that students can access to a more complete legacy information structure in limited class time. This point is particularly important for the teaching of Luoyang cultural tourism heritage, because its teaching value is often not reflected in the knowledge memory of a single scenic spot, but reflected in the pattern of the capital, religious communication, stone carving art and local culture.

Table 4: Results of ablation experiments

Model	Knowledge Quiz Accuracy / %	Task Completion Rate / %
A Full Model	91.48	89.36
B Without State Perception Module	87.12	84.28
C Without Path Scheduling Module	85.94	82.67
D Without Feedback Update Module	86.31	83.15

In order to verify the contribution of each strategy module to the overall teaching effect, this paper conducted ablation experiments. The complete model is denoted as A, the model that removes the state sensing module based on this is denoted as B, the model that removes the path scheduling module is denoted as C, and the model that removes the feedback update module is denoted as D. The evaluation results are shown in Table 4. It can be seen that the complete model achieves the highest value in knowledge test accuracy and task completion rate, which are 91.48% and 89.36%, respectively. After removing the state-aware module, the two indicators dropped to 87.12% and 84.28%, respectively. After removing the path scheduling module, the results further decreased to 85.94% and 82.67%, indicating that the path scheduling module had the most obvious effect on maintaining teaching continuity and task cohesion. After removing the feedback update module, the knowledge test accuracy and task completion rate are 86.31% and 83.15%, respectively, which are also lower than the full model. The experimental results show that the three key strategies in the proposed method have practical contributions to the improvement of teaching effect, and the path scheduling module has the greatest impact.

4 Discussion

In order to improve the collaborative level of digital expression and teaching application of Luoyang cultural tourism heritage, this paper constructs a digital twin model for heritage scenes, and designs a virtual-real integration interaction method for immersive teaching. The results show that the proposed method achieves high geometric construction accuracy and semantic anchor matching rate in three types of scenes: Longmen Grottoes, White Horse Temple and Sui and Tang Luoyang City ruins, and the comprehensive GCA and SAA reach 95.84%and 94.36%respectively. In the teaching application level, the knowledge test accuracy rate and task completion rate reached 91.48%and 89.36%respectively, which were better than the traditional multimedia teaching, panoramic VR display, ordinary 3D walkthrough teaching and interactive teaching based on rule trigger. This shows that the method in this paper not only alleviates the problem of "visual but difficult to effectively teach and call" of cultural heritage digital scenes, but also enhances the content organization ability and interaction pertinence in the teaching process. Compared with general 3D reconstruction methods, this paper does not limit digital twin to geometric model recovery, but incorporates spatial structure, historical semantics and teaching nodes into a unified modeling framework, so that the scene construction results can directly serve subsequent teaching calls. Compared with the static display immersion teaching, the method in this paper weakens the content rigidity problem caused by fixed scripts through the linkage of learning state perception, node matching and path scheduling, so that the performance in knowledge understanding and task advancement is more stable. Its practical significance is that it provides a computable, interactive and scalable digital teaching path for the course of Luoyang cultural tourism heritage, and also provides a more operational technical scheme for the transformation of local cultural resources into smart teaching scenarios. However, the method in this paper still has some limitations. First, the current experiments are mainly based on limited scenarios and specific device environments, and their cross-terminal adaptation capabilities still need to be further verified in a wider range. Second, although deep feature fusion and dynamic scheduling improve the teaching effect, the model complexity and computational overhead also increase accordingly. Thirdly, some interactive decisions still have the problem of insufficient interpretability. Future research can focus on lightweight deployment, cross-platform collaboration and knowledge graph enhancement to further improve the real-time performance, transparency and promotion ability of the system.

5 Conclusion

In summary, this paper focuses on the construction of digital twin and immersive teaching application of Luoyang cultural tourism heritage, puts forward the construction method of digital twin scene for multi-source heterogeneous data, and the interactive application method of virtual-real fusion for teaching process. The experimental results show that the constructed scene has good performance in geometry recovery and semantic organization, and the comprehensive geometry construction accuracy and semantic anchor matching rate reach 95.84%and 94.36%respectively. In the teaching application level, the knowledge test accuracy rate and task completion rate reached 91.48%and 89.36%respectively, indicating that the method in this paper can effectively improve the teaching adaptation ability and interactive application effect of Luoyang cultural tourism heritage digital scene. However, while improving the accuracy of scene expression and the quality of teaching interaction, the proposed method also increases the complexity of model structure and computing resource consumption to a certain extent, which may bring constraints to real-time deployment in resource-constrained environments. Subsequent research can continue to explore more lightweight implementation paths, such as reducing redundant connections through model pruning, building smaller-scale student models through knowledge distillation, and combining cross-terminal adaptation and incremental update mechanism to further improve the operation efficiency and deployment flexibility of the system while maintaining the quality of scene construction and teaching effect as much as possible.

References

[1] Gabellone F. Digital Twin: a new perspective for cultural heritage management and utilization[J]. Acta Imeko, 2022, 11(1): 7 pp.-7 pp.

[2] Boboc R G, Băutu E, Gîrbacia F, et al. Augmented reality in cultural heritage: an overview of the last decade of applications[J]. Applied Sciences, 2022, 12(19): 9859.

[3] Aguerre A. Open World, Open Minds: Keeping a Global Dialogue. Reflections on the ReACH Initiative[M]//Visual Heritage: Digital Approaches in Heritage Science. Cham: Springer International Publishing, 2022: 15-26.

[4] Zhou Y, Chen J, Wang M. A meta-analytic review on incorporating virtual and augmented reality in museum learning[J]. Educational Research Review, 2022, 36: 100454.

[5] Barbara J. Re-live history: An immersive virtual reality learning experience of prehistoric intangible cultural heritage[C]//Frontiers in Education. Frontiers Media SA, 2022, 7: 1032108.

[6] Innocente C, Ulrich L, Moos S, et al. A framework study on the use of immersive XR technologies in the cultural heritage domain[J]. Journal of Cultural Heritage, 2023, 62: 268-283.

[7] Alatrash S, Arnab S, Antlej K. Communicating engineering heritage through immersive technology: A VR framework for enhancing users' interpretation process in virtual immersive environments[J]. Computers & Education: X Reality, 2023, 3: 100040.

[8] Silva C, Zagalo N, Vairinhos M. Towards participatory activities with augmented reality for cultural heritage: A literature review[J]. Computers & Education: X Reality, 2023, 3: 100044.

[9] Paolanti M, Puggioni M, Frontoni E, et al. Evaluating learning outcomes of virtual reality applications in education: a proposal for digital cultural heritage[J]. ACM Journal on Computing and Cultural Heritage, 2023, 16(2): 1-25.

[10] Ch’ng E, Cai S, Feng P, et al. Social augmented reality: communicating via cultural heritage[J]. ACM Journal on Computing and Cultural Heritage, 2023, 16(2): 1-26.

[11] Xu N, Li Y, Wei X, et al. CubeMuseum AR: A tangible augmented reality interface for cultural heritage learning and museum gifting[J]. International Journal of Human–Computer Interaction, 2024, 40(6): 1409-1437.

[12] Vuoto A, Funari M F, Lourenço P B. Shaping digital twin concept for built cultural heritage conservation: a systematic literature review[J]. International journal of architectural heritage, 2024, 18(11): 1762-1795.

[13] Buragohain D, Meng Y, Deng C, et al. Digitalizing cultural heritage through metaverse applications: challenges, opportunities, and strategies[J]. Heritage Science, 2024, 12(1): 1-16.

[14] Amelio A, Zarri G P. Cultural heritage digital twin: modeling and representing the visual narrative in Leonardo Da Vinci’s Mona Lisa[J]. Neural Computing and Applications, 2024, 36(20): 11859-11876.

[15] Niccolucci F, Felicetti A. Digital twin sensors in cultural heritage ontology applications[J]. Sensors, 2024, 24(12): 3978.

[16] Guo Y, Liu L, Huang W, et al. Extending X-reality technologies to digital twin in cultural heritage risk management: a comparative evaluation from the perspective of situation awareness[J]. Heritage Science, 2024, 12(1): 1-22.

[17] Wang H, Gao Z, Zhang X, et al. Gamifying cultural heritage: Exploring the potential of immersive virtual exhibitions[J]. Telematics and Informatics Reports, 2024, 15: 100150.

[18] Chen Y, Wang X, Le B, et al. Why people use augmented reality in heritage museums: a socio-technical perspective[J]. Heritage Science, 2024, 12(1): 1-19.

[19] Corrales-Serrano M, Merchán P, Merchán M J, et al. Virtual reality applied to heritage in higher education—validation of a questionnaire to evaluate usability, learning, and emotions[J]. Heritage, 2024, 7(6): 2792-2810.

[20] Fontal O, Ibañez-Etxeberria A, Arias V B, et al. A calibrated scale to measure heritage learning in digital environments. A network analysis approach[J]. Heliyon, 2024, 10(21).

[21] Wang Z, Meng J. Dialogues with cultural heritage via museum digitalisation: developing a model of visitors’ cognitive identity, technological agent, cultural symbolism, and public engagement[J]. Museum Management and Curatorship, 2024, 39(6): 810-833.

[22] Cecotti H, Huisinga L, Peláez L G. Fully immersive learning with virtual reality for assessing students in art history[J]. Virtual Reality, 2024, 28(1): 33.