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

The Application of digital Technology in the protection and renewal of traditional Village landscape in the implementation of rural revitalization strategy

Yang Huang[1] *and Yankui Chu*[2,*]

[1] Art School of Zhengzhou Technology and Business University, Zhengzhou 450000, Henan, China

[2] Engineering School of Zhengzhou Technology and Business University, Zhengzhou 450000, Henan, China

SUMMARY: *In the process of the implementation of rural revitalization strategy, the protection and renewal of traditional village landscape style are gradually facing the problems of spatial form distortion, historical information dispersion and insufficient basis for renewal. In response to this reality, this paper constructs a digital technical framework that fuses UAV images, ground photos, point cloud data and text data, and integrates 3D reconstruction, style recognition, attention guided multi-source fusion and update decision mapping into the same computing link, so as to improve the fine expression of traditional village landscape features and the ability to judge the protection and update. The experimental results show that the SSIM of the proposed method in the digital reconstruction task reaches 0.913, the PSNR reaches 34.72 dB, and the edge preservation rate is 91.6%. In the protection update recognition task, the accuracy rate is 94.3%, the recall rate is 92.8%, the F1 value is 93.5%, and the correlation coefficient with the expert score reaches 0.887. In the subjective evaluation, the effective restoration rate was 86.7%. The research shows that this method can better realize the linkage of digital reconstruction, identification and renewal decision-making of traditional village landscape style, and provide a targeted computer technology path for the protection and renewal of traditional villages under the background of rural revitalization.*

KEYWORDS: T*raditional villages; Landscape style; Digital protection; Multi-source data fusion*

1 Introduction

After rural revitalization has entered the stage of transforming from factor input to quality improvement, traditional villages are no longer just passively preserved historical relics, but become an important carrier connecting regional context, spatial production and rural governance. The texture of streets and lanes, the layout of courtyards, the scale of buildings, the color of materials and the landscape formed between landscapes and idyll not only bear the local memory, but also determine the recognition and sustainable development potential of rural space. However, under the superposition effect of construction renewal, population flow and tourism development, some traditional villages have problems such as broken style, distortion of pattern and weakening of cultural representation. In some areas, there is even a

*chuyankui0403@163.com
https://doi.org/10.65102/is2026214 Published: 30 April, 2026

tendency of "more emphasis on construction than recognition, more emphasis on repair than integrity", which makes the original landscape characteristics be continuously diluted in the rapid transformation.

From the existing practice, the protection of traditional villages has long relied on mapping documentation, image retention and expert experience judgment. Such methods are necessary at the level of basic investigation, but they are difficult to support continuous monitoring and fine updating in complex scenes. In recent years, with the introduction of UAV images, 3D laser scanning, GIS, HBIM and deep learning technologies into the research field of cultural heritage and rural space, the digital protection of traditional villages has begun to shift from static records to dynamic perception [1-3]. Some studies use panoramic images and deep models to evaluate the quality of village public space, some studies identify the distribution characteristics of traditional villages based on multi-source remote sensing and geographic detection methods, and some studies improve the heritage display ability with the help of 3D visualization platforms [4-6]. These results provide a methodological basis for the digital expression of traditional village landscape features, but there are still several shortcomings: first, the data sources are scattered, and there is no unified correlation between building facades, street and lane interfaces, terrain environment and activity scenes. Second, the digital achievements mostly stay in the "visible" level, and the connection with the protection and renewal decision-making is not strong; Third, some studies emphasize the accuracy of single modeling, but ignore the timing of landscape evolution and the difference of renewal demand, which makes it difficult for digital technology to truly intervene in spatial governance in the context of rural revitalization.

Based on this, this paper regards the protection and renewal of traditional village landscape as a continuous calculation process consisting of "collection-identification-assessment-decision", and tries to construct the application framework of digital technology for the implementation scene of rural revitalization. This study integrates UAV images, point cloud data, historical maps, field photos and text data into the unified processing link, and improves the integrity of traditional village spatial information extraction and the pertinence of update judgment through modules such as landscape feature collection and reconstruction, landscape recognition and update decision, and multi-source data fusion guided by attention mechanism. The core issue of this paper was not only how to "restore" villages more clearly, but also how to use computer technology to identify which features should be strictly preserved, which Spaces can be orderly updated, and how the digital model can serve the subsequent repair design, style control and display communication. Research on these issues can help to promote the protection of traditional villages from experience-led to data-supported, and also provide an operational technical path for the refined renewal of landscape features under the background of rural revitalization.

2 Literature Review

2.1 Research status of digital protection of traditional village landscape style

Under the background of continuous rural revitalization, the protection and renewal of traditional village landscape has gradually shifted from empirical judgment to data support. The existing research generally follows the path of "spatial acquisition, information archive-feature identification-display and communication". In terms of data acquisition, UAV images, oblique photography, panoramic images, 3D laser scanning and mobile photogrammetry have become important means of digital recording of traditional villages. Yu et al. carried out village landscape analysis with the help of virtual earth and 3D participatory geographic information system, and showed that 3D spatial expression could enhance the readability of historical environment [1]. Li et al. combined UAV images with deep learning for the investigation, evaluation and dynamic monitoring of traditional village buildings, showing that automatic recognition has high efficiency in large-scale sample processing [2]. This kind of research has improved the limitations of only relying on paper mapping and static photos in the past, so that the village texture, architectural form and environmental interface can be more completely digitally preserved.

However, according to the existing results, there are still obvious cracks in digital protection. One kind of research focuses on spatial geometric reconstruction, which can better recover building contours, roof slopes and street boundaries, but lacks the expression of material texture, structural details, color style and village life scenes [3]. The other type of research emphasizes landscape evaluation and public space perception, and identifies spatial quality features with street view images, panoramic photos or convolutional neural networks [4]. However, the output results mostly stay at the scoring level, and the connection with subsequent update design is limited. Some other studies focus on the distribution pattern, influence factors and regional differences of traditional villages [5], which provide a geographical analysis basis for macro cognition, but it is difficult to directly answer the micro decision-making problem of a single village in landscape protection and renewal.

2.2 Application analysis of digital model and intelligent recognition technology in related fields

With the development of computer vision, deep learning and cultural heritage information modeling, the research of landscape has begun to move from single modeling to multi-model collaboration. Wang et al. summarized that deep learning is changing the method basis of traditional village landscape evaluation, especially showing strong adaptability in image classification, style recognition, and spatial semantic extraction [6]. In the cultural heritage scene, HBIM has been used in the data organization, hierarchical modeling and disease information linking of historical buildings [7-9]. Its value is not only in 3D expression, but also in the integration of geometry, age, material, component and repair information into a unified database. At the same time, the development of 3D display platforms and multimodal visualization systems on the Web has changed cultural heritage from "isolated modeling achievements" to "interactive shared resources" [10-12]. The introduction of digital twin further promotes the extension of heritage information from static documentation to real-time perception and dynamic evolution simulation [13].

These technologies provide a reference path for the protection and renewal of traditional village landscape style. On the one hand, space-ground LiDAR, multi-source image stitching and point cloud reconstruction help to synchronously obtain composite information such as building entities, terrain boundaries, road organization and plant cover [14, 15]. On the other hand, attention mechanism, multi-source feature fusion and semantic segmentation model can improve the recognition accuracy of style elements in complex scenes, so that objects such as walls, roofs, pavement, water bodies, greening and public nodes can be distinguished more stably. The problem is that when the existing technology is transferred to traditional villages, the phenomenon of "strong at both ends and weak in the middle" often occurs: the data collection is more and more refined, and the result display is more and more intuitive, but there is a lack of close enough computational mapping between the recognition results and the update strategy. Especially in the protection and renewal of villages, the decision object is not only a single building, but an overall style composed of building groups, street interfaces, landscape nodes, historical layered relationships and life activities, which requires the model to be able to process multi-scale, multi-temporal and multi-semantic level data, rather than only completing a single recognition task.

2.3 Existing research deficiencies and future development directions

On the whole, the main shortcomings of the current research are reflected in three levels. First, the data level emphasizes the acquisition accuracy, but the unified organization of heterogeneous data is insufficient. Aerial images, point clouds, historical documents, oral documents and field photos are often stored separately, and there is no association mechanism for updating tasks. Secondly, the method level focuses on the identification or modeling itself, and the calculation relationship between "landscape protection object, update intervention intensity, and implementation priority" is not discussed enough, which makes the digital achievements difficult to truly formulate service plans. Thirdly, there is a tendency of "emphasizing display and ignoring governance" at the application level. Although some platforms can realize three-dimensional browsing and results publishing, they lack support in style early warning, transformation comparison and selection, repair tracking and dynamic evaluation.

Therefore, the application of digital technology of traditional village landscape should not stay in the superposition of single tools, but should turn to the construction of an integrated framework: multi-source data collection, spatial reconstruction, semantic recognition, landscape assessment and update decision-making into the same processing link, and the attention mechanism is used to guide multi-source information fusion to improve the extraction ability of key features. With the cooperation of HBIM, GIS and Web platform, results management, sharing and visual deduction were realized. Furthermore, digital twin and temporal monitoring were combined to transform traditional village protection and renewal from one-time modeling to continuous governance. Table 1 summarizes the relevant research paths, and it can be seen that the direction that really meets the needs of rural revitalization is not to simply improve a certain technical index, but to establish a digital technology system that takes into account recognition accuracy, spatial interpretation and update availability.

Table 1: Comparison of technological paths related to the digitalization of traditional village landscape features

Research Path	Main Data Sources	Typical Technical Methods	Advantages	Main Limitations
Image Surveying and 3D Reconstruction	UAV imagery, oblique photography, point clouds	Photogrammetry, 3D modeling, LiDAR fusion	Can restore building outlines and spatial forms relatively completely	Insufficient in representing material details, daily-life scenes, and historical semantics
Landscape Evaluation and Style Recognition	Panoramic images, street-view photos, field-sampled images	Convolutional neural networks, semantic segmentation, feature extraction	High recognition efficiency and suitable for large-scale analysis	Evaluation results are mostly static scores and are weakly connected with renewal decision-making
Heritage Information Modeling	Architectural surveying data, component archives, restoration records	HBIM, hierarchical modeling, attribute association	Facilitates component-level management and integration of historical information	Insufficient coverage of the overall village landscape pattern and open environment
Spatial Distribution and Pattern Analysis	GIS data, remote sensing images, geostatistical data	GIS analysis, geographic detectors, spatial statistics	Suitable for studying macroscopic distribution patterns	Difficult to directly support renewal design at the single-village level
Platform-Based Display and Digital Twin	Multimodal heritage data, monitoring data	Web3D, visualization platforms, digital twins	Strong interactivity and convenient for sharing, dissemination, and dynamic display	High modeling cost, and the embedding of decision-making algorithms remains insufficient
Multi-Source Fusion and Intelligent Decision-Making	Images, point clouds, text, historical archives, monitoring data	Attention mechanisms, feature fusion, decision models	Can simultaneously support recognition, evaluation, and renewal judgment	Still at the development stage and highly dependent on high-quality annotated data

3 Research Methods

3.1 Framework design of digital technology for protection and renewal of traditional village landscape style

The protection and renewal of traditional village landscape is not a single architectural object, but a composite system composed of architectural entities, street space, terrain and landform, vegetation and water system and local life scenes. If only relying on static mapping or local image recognition, only discrete results can be obtained, which is difficult to support the overall update judgment under the background of rural revitalization. Based on this, this paper constructs a digital technology framework for the protection and renewal of traditional village landscapes, which organizes multi-source data acquisition, 3D reconstruction, landscape recognition, fusion assessment and update decision-making into a unified computing link, so that digital technology not only serves for "record", but also directly enters the "identification-diagnosis-intervention" link.

The input terminal of the framework is composed of UAV tilt images, ground truth photos, laser point clouds, historical maps, village survey texts and basic geographic data, which are denoted as

$$X = \{X_u, X_p, X_l, X_h, X_t, X_g\} \tag{1}$$

Here, X_u represents aerial photography, X_p represents ground photos, X_l represents point cloud data, X_h represents historical data, X_t represents text description, and X_g represents GIS spatial base map. Different types of data have obvious differences in resolution, time scale and semantic granularity. Therefore, the framework sets a unified preprocessing layer at the front end to complete coordinate registration, noise removal, time alignment and semantic annotation of the data, and forms a standardized feature set $F = \{f_1, f_2, \cdots, f_n\}$.

In the core modeling layer, the digital representation of landscape is divided into two coupled sub-processes: "spatial reconstruction" and "landscape discrimination". For the continuous form of village material space, 3D reconstruction function is used

$$M = \Phi(F_s; \theta_s) \tag{2}$$

Here, F_s is the spatial geometric feature, M is the reconstructed village landscape model, and θ_s is the model parameter. The process focuses on recovering building contours, roof slopes, street-lane interfaces, and elevation relationships to provide a stable geometric basis for subsequent feature recognition. For style attributes and update requirements, a recognition decision function is established

$$Y = \Psi(F_v, F_c, F_e; \theta_d) \tag{3}$$

where F_v represents visual features, F_c represents cultural semantic features, F_e represents environmental association features, and Y is the output of style classification and update suggestions. The purpose of such processing is to avoid simplifying traditional village protection into a pure image classification problem, but to synchronously bring "visible forms" and "interpretable cultural information" into the judgment.

Considering that there are significant primary and secondary differences between village features, the framework further introduces an attention-guided fusion mechanism to express the key landscape information in a weighted manner. It is calculated as

$$\alpha_i = \frac{\exp(w^\top f_i)}{\sum_{j=1}^{n} \exp(w^\top f_j)}, \quad F^* = \sum_{i=1}^{n} \alpha_i f_i \tag{4}$$

Here, α_i is the weight of the ith feature and F^* is the global representation after fusion. Through this mechanism, elements with strong landscape recognition, such as traditional roofing forms, wall materials, street and lane scales, and nodal landscape interfaces, will obtain higher contribution values in the model, thereby weakening the influence of cluttering background and accidental interference on the judgment results.

At the output end, the system not only gives the digital reconstruction results of traditional village landscapes, but also synchronously generates the protection level, the identification results of updated sensitive areas and the intervention priority matrix, which provides a basis for subsequent repair design, style improvement and display dissemination. In general, the framework links multi-source acquisition, intelligent identification and renewal decision-making, making the protection of traditional village landscape style from experience leading to model-assisted, and laying a technical foundation for fine spatial governance in the implementation of rural revitalization.

3.2 Detailed description of digital processing and model operation process

In the traditional village landscape protection and update scenario, the model operation is not a simple image input and classification output, but a continuous processing process consisting of data cleaning, spatial alignment, feature extraction, fusion and discrimination, and result writing back. In order to ensure that data from different sources can enter the same computing link, this paper first uniformly preprocesses UAV images, point clouds, ground photos, historical drawings and text records. Let the original data set be

$$D = \{(X_i, y_i)\}_{i=1}^{N} \tag{5}$$

Here, X_i represents the multi-source input of the ith village sample, and y_i is the manually labeled landscape category or updated label. After coordinate registration, noise removal, resolution normalization and timestamp correction, standardized samples are obtained

$$\widetilde{X}_i = T(X_i) \tag{6}$$

where, $T(\cdot)$ represents the preprocessed mapping function. The role of this step is to reduce the bias introduced by different devices, different time phases, and different viewpoints, so that building boundaries, road contours, and environmental backgrounds are comparable in subsequent training.

In the feature encoding stage, the system inputs standardized samples into a shared encoder to extract geometric, texture and semantic features:

$$H_i = E(\widetilde{X}_i; \theta_e) \tag{7}$$

where $E(\cdot)$ is the coding network and θ_e is its parameter. Considering that the information of traditional village style is hierarchical, and the contribution of roof form, wall material, street scale, landscape node and surrounding terrain to the final judgment is not the same, this paper uses the attention weighting mechanism to reallocate the local features. Let the KTH class of local features be h_k, and its weight is defined as

$$\beta_k = \frac{\exp(q^{\top} h_k)}{\sum_{j=1}^{m} \exp(q^{\top} h_j)} \tag{8}$$

Then the fusion features are obtained

$$H_i^* = \sum_{k=1}^{m} \beta_k h_k \tag{9}$$

This processing can highlight key information that can characterize the continuity of traditional style and suppress interference factors such as billboards, temporary structures and random occlusion. In the output stage, the fused features enter the reconstruction branch and the decision branch at the same time. The former generates a digital model of village landscape

$$M_i = R(H_i^*; \theta_r) \tag{10}$$

The latter outputs the results of style recognition and update suggestions

$$\hat{y}_i = C(H_i^*; \theta_c) \tag{11}$$

Here, $R(\cdot)$ represents the reconstruction module, $C(\cdot)$ represents the classification decision module, and θ_r and θ_c are the corresponding parameters, respectively. In order to make the model take into account the quality of spatial restoration and the accuracy of update judgment, this paper constructs a joint loss function:

$$L = \lambda_1 L_{rec} + \lambda_2 L_{cls} + \lambda_3 L_{con} \tag{12}$$

In the formula, L_{rec} is used to constrain the error between the reconstruction result and the real spatial shape, L_{cls} is used to measure the deviation between the feature recognition category and the manual annotation, and L_{con} is used to maintain the consistency between the historical feature information and the current update state. The $\lambda_1, \lambda_2, \lambda_3$ are the weight coefficients. Model training uses mini-batch iteration to jointly update the parameters of each module, and its basic update form is as follows

$$\theta \leftarrow \theta - \eta \nabla_\theta L \tag{13}$$

where η is the learning rate. As the iteration progresses, the system gradually forms a stable recognition ability for traditional village building forms, landscape interfaces and updated sensitive areas. After the model runs, the platform will output three types of results: the first is the visual 3D reconstruction results, the second is the style protection level map, and the third is the update priority matrix. In this way, the digital process goes beyond archiving information and is further transformed into a decision-making basis that can be used for renovation control and updating design.

3.3 System model composition and key module design

3.3.1 Design of landscape feature acquisition and reconstruction module

The landscape feature acquisition and reconstruction module assumes the starting task of digital modeling of traditional villages. Its core is not simply to obtain images, but to transform information such as building facades, street and lane interfaces, roof forms, courtyard boundaries, pavement textures and surrounding terrain into a unified spatial expression. Considering the problems of large scale differences, complex occlusion relationships, and obvious material aging within traditional villages, this paper uses the composite collection method of "aerial survey + ground photography + point cloud complementary survey" to construct the original data set. Uav oblique photography is used to obtain the relationship between the overall pattern and elevation, ground multi-view images are used to supplement the texture and node details of doors, Windows, walls, and laser point clouds are used to correct the geometric deviation of deep lanes, coreaths, and corner areas. After the coordinate unification of multi-source data is completed, it is input into the reconstruction network to form the landscape feature tensor F_r.

In the feature extraction stage, the module uses the convolutional encoder to jointly represent the image texture and geometric edge. Let the input image block be I and the convolution kernel be K, then the LTH layer feature map can be expressed as

$$F^{(l)} = \sigma(K^{(l)} * F^{(l-1)} + b^{(l)}) \tag{14}$$

Here, $*$ denotes the convolution operation, $b^{(l)}$ is the bias term, and σ is the nonlinear activation function. The process can extract key features such as ridge line, wall boundary, window rhythm and material texture layer by layer. In order to reduce the reconstruction fracture caused by different views, this paper further introduces the geometric consistency constraint to control the distance error between the reconstructed point set P_g and the measured point set P_t within an acceptable range, and its loss function is written as

$$L_{geo} = \frac{1}{|P_g|} \sum_{p \in P_g} \min_{q \in P_t} \| p - q \|_2^2 \tag{15}$$

This formula is used to constrain the model to keep the geometric shape consistent with the real scene during the training process, avoiding contour drift or local stretch.

At the output end, the module generates two types of results: one is the 3D village landscape model for overall expression, and the other is the local landscape feature fragments that can be called by the subsequent recognition module. The significance of this process is that the reconstruction results not only serve for visual display, but also provide underlying data support for style recognition, judgment of update sensitive areas, and comparison and selection of repair schemes. Table 2 presents the main structural Settings of this module.

Table 2: Landscape feature acquisition and reconstruction module structure Settings

Hierarchy / Module	Input Content	Output Content	Main Parameter Settings	Functional Description
Data Acquisition Layer	UAV imagery, ground photographs, and LiDAR point clouds	Raw multi-source dataset	Aerial survey resolution: 2–5 cm; point cloud density: 200 pts/m²	Acquires the overall spatial pattern and local detailed information
Preprocessing Layer	Raw multi-source data	Registered standard samples	Coordinate unification, noise filtering, and distortion correction	Eliminates differences caused by viewpoints and devices
Feature Encoding Layer	Standard sample image patches	Multi-scale feature maps	Convolution kernel (3×3), stride = 1	Extracts texture, edge, and structural information
Geometric Constraint Layer	Feature maps and point cloud data	Geometrically consistent features	Euclidean distance constraint, error threshold = 0.05 m	Improves reconstruction accuracy and continuity
3D Reconstruction Layer	Fused feature tensors	Village 3D model and local feature fragments	Mesh resolution: 256×256	Outputs a computable landscape-style model

3.3.2 Design of style recognition and update decision module

The feature recognition and update decision module is located in the middle of the system. Its role is not to simply classify the image, but to transform the spatial shape characteristics of traditional villages into judgment results that can be used for protection intervention. Whether the landscape style should be completely preserved often depends on the comprehensive state of multiple types of information, including architectural style continuity, interface integrity, material coordination, historical identification and matching degree of the surrounding environment. Based on this characteristic, this paper designs the module as a three-level structure of "feature discrimination-level output-strategy mapping", so that the recognition results can directly correspond to the subsequent update scheme.

The module input is the local style feature block and the global scene feature vector output in the previous stage, denoted as X_f. In the encoding process, the network gradually compresses the spatial size and enhances the high-order semantic expression through multi-layer convolution and pooling. The LTH layer feature response can be expressed as

$$Z^{(l)} = \phi\left(W^{(l)} \odot Z^{(l-1)} + b^{(l)}\right) \tag{16}$$

Here, $\odot$ represents the convolution operation, $W^{(l)}$ and $b^{(l)}$ are the weights and biases, and $\phi(\cdot)$ is the activation function. After continuous down-sampling, the system can stably extract structural information sensitive to style judgment, such as roof contour, wall texture, window opening rhythm, street and lane enclosure degree.

In order to avoid the recognition results staying at the abstract label level, this paper adds an update decision mapping unit after the classification head. Let the set of style categories be $C = \{c_1, c_2, \cdots, c_m\}$, then the probability that the sample belongs to the JTH class is

$$P(c_j|z) = \frac{\exp(\omega_j^T z)}{\sum_{k=1}^{m}\exp(\omega_k^T z)} \tag{17}$$

Here, z is the discriminant vector output by the fully connected layer and ω_j^T is the corresponding category parameter. After the category results are generated, the system further combines the building completion degree d, the historical value weight h and the environmental coordination coefficient e to calculate the update priority index:

$$U = \lambda_a d + \lambda_b h + \lambda_c(1 - e) \tag{18}$$

where $\lambda_a, \lambda_b, \lambda_c$ are the weight coefficients. The larger the value of U, the more unsuitable the object is for general remediation, and it should enter the focus protection or fine repair sequence. In this way, the model output is no longer just "what look?" but further answers "how to update?"

From the perspective of implementation, this module takes into account both recognition accuracy and decision interpretability. On the one hand, the convolutional discriminant structure can deal with the spatial texture differences in complex backgrounds. On the other hand, the renewal index relates the identification results to the protection intensity, so that the digital model can participate in the repair ranking, style improvement and node optimization. Table 3 lists the main structural Settings of this module.

Table 3: Structure Settings of feature recognition and update decision module

Hierarchy / Unit	Input	Output	Main Parameters	Functional Description
Convolutional Discrimination Layer 1	Local landscape image patches	Low-level texture features	Kernel size (3×3), stride = 1	Extracts edge, material, and contour information
Convolutional Discrimination Layer 2	Low-level texture features	Mid-level structural features	128 channels, max pooling (2×2)	Enhances relationships among roofs, walls, and interface structures
Convolutional Discrimination Layer 3	Mid-level structural features	High-level semantic features	256 channels, global average pooling	Forms an overall landscape-style representation
Classification Output Layer	High-level semantic features	Landscape category probabilities	Softmax, multiclass output	Identifies categories such as protected type, coordinated renewal type, and renovation-optimization type
Decision Mapping Layer	Category results and attribute parameters	Renewal priority index (U)	Adaptive learning of weight coefficients	Outputs protection level and renewal recommendations

3.3.3 Design of multi-source data fusion module guided by attention mechanism

The digital expression of traditional village landscape features has obvious multi-source heterogeneous characteristics. Aerial photography is good at showing the overall pattern and spatial hierarchy, ground photos are more conducive to capturing the details of facades and material textures, point cloud data can stably describe the elevation difference, rotation Angle, and interface boundaries, and historical drawings and survey texts preserve the chronological information and cultural semantics that are difficult to extract directly from images. If these data are simply spliced, it is not only difficult to reflect the contribution differences of different modalities, but also easy to cause the problem of local details being submerged and semantic pointing unclear. Based on this, this paper sets up a multi-source data fusion module guided by the attention mechanism in the system, and constructs the comprehensive representation of village features through the way of "unified coding, correlation calculation, weight allocation, gated aggregation". As shown in Figure 1, the module takes UAV images, ground photos, point cloud data and historical text as input, generates query, key and value vectors after unified feature projection, and then realizes information reweighting by cross-modal attention calculation. Finally, the fusion feature for style recognition and update decision is output.

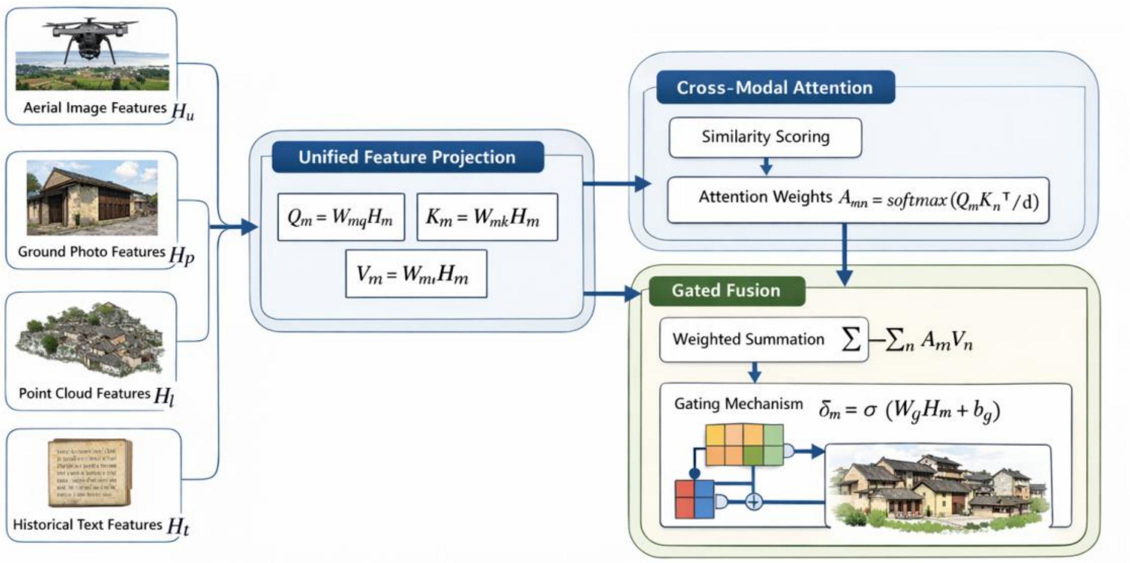

Figure 1: Schematic diagram of the multi-source data fusion module structure guided by the attention mechanism

Suppose that the encoded features of aerial images, ground photos, point clouds and text are H_u, H_p, H_l, H_t respectively. The module first maps them into the same latent space to obtain the query, key and value vectors:

$$Q_m = W_m^q H_m, \quad K_m = W_m^k H_m, \quad V_m = W_m^v H_m \tag{19}$$

Here, $m \in u,p,l,t$, W_m^q, W_m^k, W_m^v are the learnable parameter matrices. Then, the inter-modal correlation weights are calculated to measure the contribution of each type of data to the current feature discrimination task:

$$A_{mn} = \mathrm{softmax}\left(\frac{Q_m K_n^T}{\sqrt{d}}\right) \tag{20}$$

where, d is the feature dimension. The weight matrix does not change the original data source, but highlights the information that is more explanatory in the face recognition. For example, when there are complex ridges and facade decorations in village nodes, the weights of ground photos and local point clouds will rise accordingly. The contribution of aerial photography and GIS base map will be more obvious when the research object emphasizes more on pattern continuity and street scale.

In order to avoid a single modality occupying a high proportion in the fusion, this paper further sets up a gated update mechanism to adaptively filter the information of each modality. The fusion result is written as

$$F_{fusion} = \sum_{m} \delta_m \odot \left(\sum_{n} A_{mn} V_n\right), \delta_m = \sigma(W_g H_m + b_g) \tag{21}$$

Here, δ_m is the gating coefficient, $\sigma(\cdot)$ is the Sigmoid function, and $\odot$ represents element-wise multiplication. After this process, the system can retain the overall style and structure, while enhancing the sensitivity to the key parts, so that the cornice line foot, wall material, courtyard enclosing interface, laneway closure relationship and other characteristics can be more fully expressed in the subsequent identification and decision-making process.

3.4 Comparative analysis of different technical architectures

To more clearly illustrate the pertinency of the technical route in this paper, Table 4 compares the conventional digital protection architecture with the proposed multi-source fusion architecture. Traditional methods are usually based on single image modeling or local recognition. Although the implementation path is relatively direct, in scenes with complex spatial levels and interwoven style information such as traditional villages, only geometric restoration or static classification can be completed, and it is difficult to take into account style recognition, update judgment and result linkage. In contrast, on the basis of retaining the main link of 3D reconstruction, the architecture of this paper introduces the attention mechanism, multi-source data fusion and update decision mapping module, so that the model no longer stays at "building villages", but can further answer the question "which features should be protected, which Spaces can be updated, and how to control the update intensity". From the system performance, this architecture is more suitable for serving the traditional village protection and renewal task under the background of rural revitalization, especially has stronger advantages in detail maintenance, semantic integrity and decision-making availability.

Table 4: Comparison of different digital technology architectures

Comparison Item	Conventional Digital Architecture	Proposed Architecture in This Study
Data sources	Single-source imagery or partial surveying/mapping data	Multi-source data, including aerial images, ground photographs, point clouds, and text
Core functions	3D modeling or landscape-feature classification	Integrated reconstruction, recognition, fusion, and decision-making
Feature processing method	Conventional convolution-based extraction	Attention-guided enhancement of key features
Data organization method	Decentralized processing with weak interconnections	Unified encoding and cross-modal fusion
Update support capability	Mainly for visualization and archiving	Able to output protection levels and renewal recommendations
Application effectiveness	More intuitive results, but limited decision support	Higher detail restoration and stronger usability for renewal decisions

As shown in Table 4, the improvement of the method in this paper does not lie in simply increasing the number of model layers, but in incorporating the spatial information, visual information and historical semantics in traditional village landscape into the same computing framework, so as to enhance the practical support ability of digital technology for protection and renewal practice.

4 Experimental Evaluation

4.1 Experimental scheme design

In order to verify the practical applicability of the digital technical framework constructed in this paper in the protection and renewal of traditional village landscape style, the experiment is carried out from three levels of "spatial reconstruction effect - style recognition ability - update decision stability", focusing on the comprehensive performance of the model on form reduction, feature extraction and result output in complex village scenes. The experimental samples are collected from representative traditional villages in Henan, Guizhou and Fujian, and the dataset of traditional village landscape is constructed by combining field collection and publicly available data. The data content included UAV oblique images, ground multi-view photos, local laser point clouds, historical maps and manually labeled texts. A total of 1260 groups of village building samples and 840 groups of street and node landscape samples were collected. The data is divided into training set, validation set and test set according to the proportion of 70%, 15% and 15% to ensure the relative independence between model training and result testing.

Three common methods are selected as baselines for comparative experiments: single photogrammetric reconstruction method, common convolutional recognition method, and multi-source input model without attention fusion mechanism. The former is used to test the advantages of the proposed method in geometric recovery and detail preservation, and the latter two are used to compare the stability differences between the output of style recognition and update proposals. In the reconstruction task, the structural similarity index SSIM and peak signal-to-noise ratio PSNR were used to measure the restoration quality of the model to the building and landscape form. In the recognition task, the accuracy rate, recall rate and F1 value were used to evaluate the classification results of style classification. In the update decision part, the interpretability and practical reference value of the model output are tested by the consistency coefficient and the correlation degree of expert ratings.

The experimental environment is deployed on Ubuntu 22.04 platform with Intel Xeon 2.8 GHz processor, 64 GB memory and NVIDIA RTX 4090 graphics card. The model is implemented based on PyTorch. In the training phase, Adam optimizer is used, the initial learning rate is set to 0.0002, the batch size is 16, and the iteration rounds are 200. Considering the problems of large scale differences and many partial occlutions in traditional village data, preprocessing steps such as image cropping, brightness normalization, point cloud filtering and text label cleaning are added before training to reduce noise interference and improve sample consistency. Through this experimental scheme, the effectiveness of the model in the digital protection and renewal of traditional village landscape can be systematically tested.

4.2 Analysis of experimental results

4.2.1 Analysis of digital reconstruction results of traditional village architecture and landscape style

In order to test the performance of the model in the digital reconstruction of traditional village buildings and landscape features, this paper selects five high-frequency scenes such as building roof turning, wooden structure node, wall texture, street and lane interface and courtyard boundary for comparison experiments, and compares the proposed method with single photogrammetric method and common multi-source reconstruction method. As shown in Figure 2, the proposed model performs better in both detail restoration and spatial continuity. Especially in the reconstruction of cornice lines, wall material boundaries and corner areas of laneways, the contour is more complete and the boundary is clearer, and the deformation caused by partial occlusion is significantly reduced. The reason is that multi-source data realizes differentiated weighting under the guidance of the attention mechanism, aerial photography is responsible for the overall pattern recovery, ground photos supplement the facade details, and point cloud data constrain the geometric boundaries. The three jointly improve the expression ability of the model for complex landscape features.

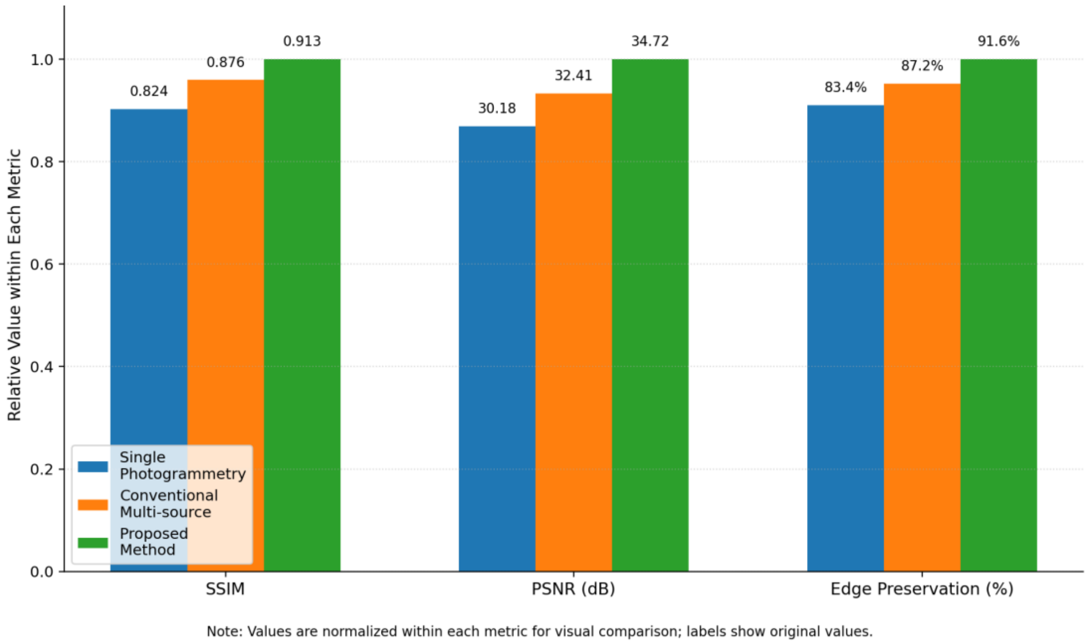

Figure 2: Comparison of digital reconstruction results of traditional village buildings and landscape features under different methods

From the perspective of quantitative indicators, the SSIM of the proposed method on the test set reaches 0.913, the PSNR is 34.72 dB, and the edge preservation rate is 91.6%, which are higher than 0.824, 30.18 dB and 83.4% of the single photogrammetric method. It is also better than the common multi-source reconstruction method of 0.876, 32.41 dB and 87.2%. If we further focus on the detail areas that are more sensitive to the protection of traditional villages, the accuracy of the model in restoring the texture of wood components is 89.7%, and the recognition and retention rate of the cracks and repair boundaries of masonry walls is 88.4%. In the continuity evaluation of the street and lane interface composed of expert scoring and interface integrity index, the comprehensive score of the method in this paper reaches 90.1. In contrast, the single photogrammetric method only scored 78.9%, 75.6% and 80.7 in the above three indicators, respectively. It shows that although the traditional method can restore the basic contour, it is easy to lose details when it comes to material texture, age traces and local structural relationships, while the proposed model can better retain the real level of traditional village features in the reconstruction process.

To verify whether the difference is statistically significant, the SSIM results of the three methods are tested with independent samples. The results show that the difference between the proposed method and the single photogrammetric method is significant ($p<0.001$), and the difference between the proposed method and the common multi-source reconstruction method is also significant ($p=0.003$). This indicates that the digital reconstruction framework constructed in this paper does not only perform well on local samples, but also has a relatively stable advantage on the overall test data. It can be seen that the proposed method has good accuracy and practicability in the digital reconstruction task of traditional village buildings and landscape features, and can provide reliable data basis for subsequent feature recognition and protection and update judgment.

4.2.2 Analysis of identification results of landscape style protection and renewal

After the digital reconstruction of traditional village architecture and landscape style, whether the model can accurately identify the types of protection objects, judge and update sensitive areas, and output intervention suggestions with reference value is directly related to whether the digital results can enter the actual governance link. To this end, this paper divides the village samples in the test set into three categories: "key protection type, coordinated repair type, and remediation and optimization type", and compares them with the ordinary convolutional recognition model and the multi-source recognition model without introducing the attention fusion mechanism. As shown in Figure 3, the recognition results of the proposed method on the three types of tasks are more focused, and the misjudgment samples are significantly reduced, indicating that the attention mechanism and the multi-source fusion strategy can better distinguish the subtle differences between traditional roofing, street and lane interface, material coordination and environmental damage traces.

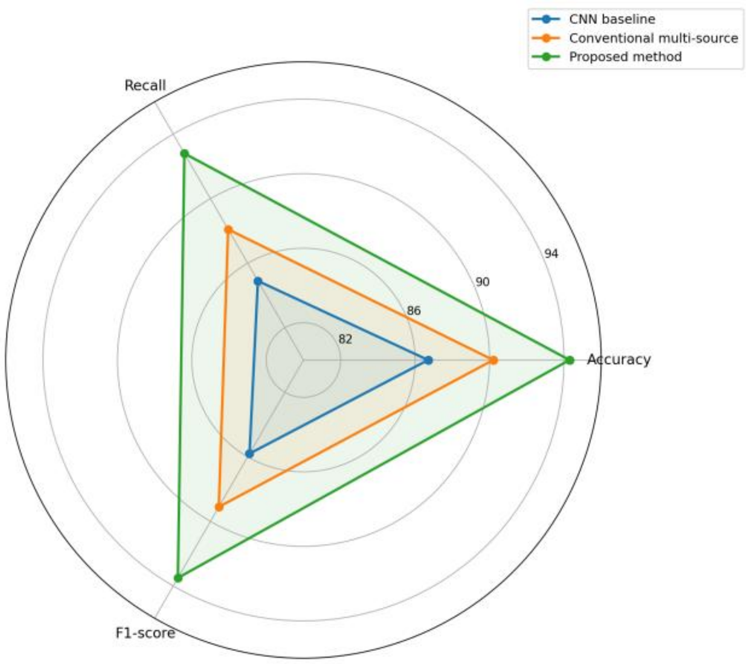

Figure 3: Comparison of results of different models in the recognition task of landscape style conservation and renewal

From the quantitative results, the overall accuracy of the proposed model on the test set reaches 94.3%, the recall rate is 92.8%, and the F1 value is 93.5%. 86.7%, 84.9% and 85.8% for the common convolutional recognition model, and 90.2%, 88.1% and 89.1% for the common multi-source recognition model. If we further observe different categories, the recognition accuracy of the proposed model for the "key protection type" samples is 95.6%, the "coordination repair type" is 93.4%, and the "remediation and optimization type" is 94.0%. The fluctuation among the three types of results is small, indicating that the model still maintains good stability under the condition of incomplete balanced class distribution. At the same time, the Pearson correlation coefficient between the update priority judgment and the expert score reaches 0.887, which is higher than 0.742 of the ordinary convolutional recognition model and 0.811 of the ordinary multi-source recognition model, which means that the update suggestions output by the proposed method have stronger consistency with human judgments.

From the perspective of recognition details, the ordinary convolution model is easy to misjudge temporary structures, advertisement occlusion and local repair as overall style damage on the street renovation samples, resulting in a high protection level. Although the ordinary multi-source model has been improved, the boundary ambiguity and type confusion still occur in the scenes where historic buildings and later renovated buildings coexist. By introducing the attention weighting mechanism, the model can more fully express the roof shape, wall material, courtyard boundary and surrounding environment, so that it can better maintain the rationality of recognition results in complex scenes. In general, this method not only improves the accuracy of landscape recognition, but also enhances the interpretability of conservation and renewal decisions, which provides a more reliable calculation basis for subsequent repair sequencing and partition control.

4.2.3 Stability analysis of model training

The stability of model training is directly related to the reliability of digital reconstruction results and feature recognition output. Since the proposed model simultaneously processes aerial images, ground photos, point clouds and text annotations, there are resolution differences and semantic noise between multi-source inputs. If the training process fluctuates too much, it is easy to lead to drifting reconstruction boundaries, repeated category judgments and unstable update suggestions. Based on this, this paper records the total loss value, the F1 value of the validation set and the change of the reconstruction similarity in 200 training cycles. As shown in Figure 4, the total loss decreases rapidly in the first 40 rounds, from 1.86 to 0.74. After the 120th round, the curve tends to be smooth and stable around 0.31 around the 160th round, and the subsequent fluctuation range is controlled within ±0.03, indicating that the model has formed a relatively stable convergence state.

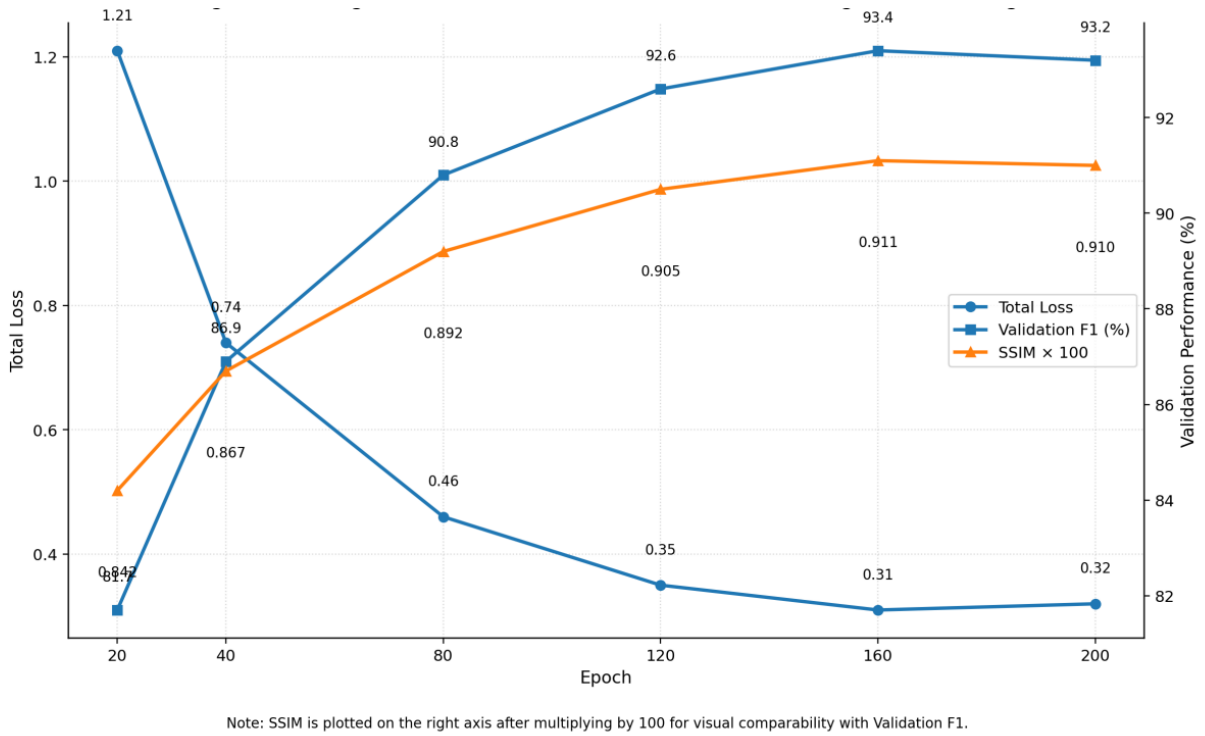

Figure 4: Loss values versus validation set performance changes during model training

Corresponding to the change of loss, the F1 value of the validation set shows a continuous upward trend, increasing from 81.7% in the 20th round to 91.8% in the 100th round, reaching 93.4% in the 160th round, and remaining around 93.2% at the end of the training without a significant decline. At the same time, the SSIM of the reconstruction task steadily increases from 0.842 to 0.911, and the standard deviation is only 0.007 in the later stage, indicating that the model does not sacrifice the quality of spatial reconstruction while maintaining the improvement of recognition performance. Compared with the ordinary multi-source model, it can be found that the latter still has obvious oscillation after 140 rounds, and its loss fluctuation range is 0.09, which is significantly higher than 0.03 of the proposed method.

4.2.4 Subjective evaluation and analysis of style restoration effect

In addition to objective indicators such as SSIM, PSNR and F1 value, whether the digital results of traditional village landscape are really close to the real scene needs to be verified by artificial perception evaluation. To this end, this paper organizes 15 evaluators from the research fields of architectural heritage protection, urban and rural planning, digital media and rural construction to participate in the subjective test in a double-blind manner. There are 45 groups of test samples, including 15 groups of reconstruction results generated or output by the proposed model, the common convolutional recognition model, and the common multi-source model. The evaluation dimensions were set as visual realism, local style restoration and rationality of update suggestions, all of which were scored on a 5-point scale. As shown in Figure 5, the proposed model achieves higher scores on the three indicators, with an average of 4.63 points, 4.71 points and 4.52 points respectively, while the corresponding scores of the ordinary convolution model are 3.88 points, 3.76 points and 3.69 points, and the ordinary multi-source model are 4.21 points, 4.08 points and 4.03 points, respectively.

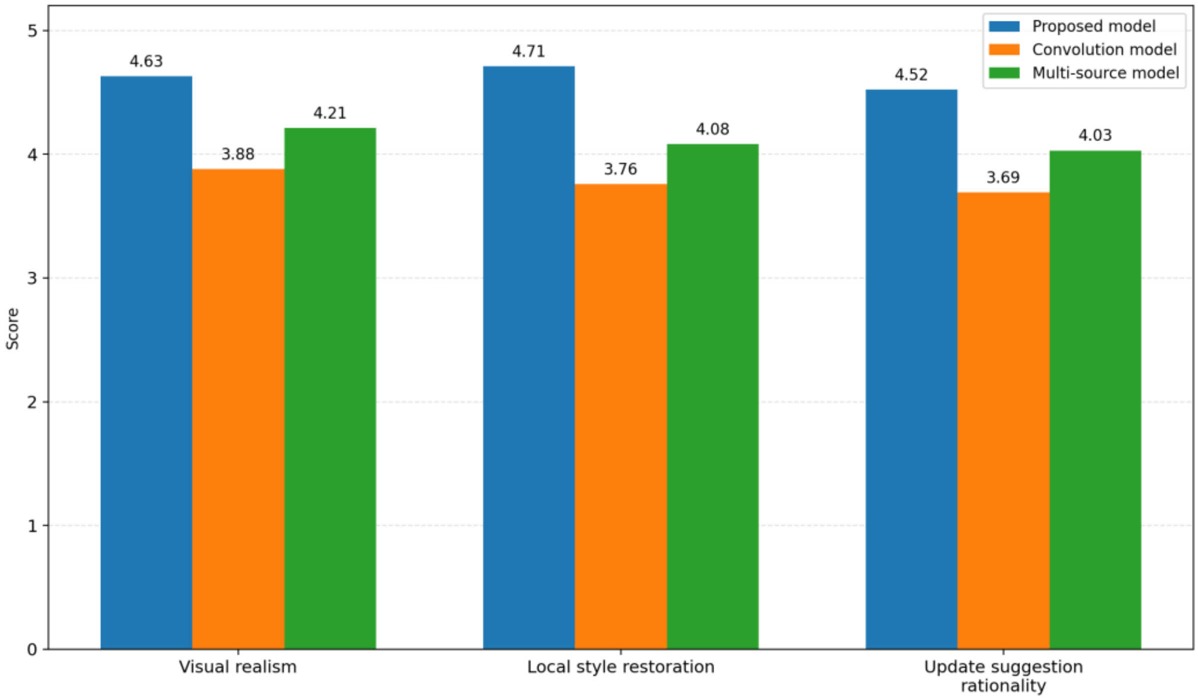

Figure 5: Comparison of subjective evaluation scores of different models

In this paper, 'at least 10 out of 15 raters give no less than 4 scores for the same sample' is defined as effective style restoration. Under this criterion, the overall restoration achievement rate of the proposed model is 86.7%, which is significantly higher than 61.1% of the ordinary convolution model and 73.3% of the ordinary multi-source model. Among them, the advantages of the proposed method are more obvious in the samples with more complex identification of wood details, street and lane enclose relationships and courtyard boundaries. The evaluators generally believe that the output results of the model are closer to the real face of traditional villages in terms of building proportion, material texture and spatial level, and the relationship between local repair traces and the overall environment is also handled more naturally. In order to test the consistency of the scores, this paper further calculated the Fleiss 'Kappa coefficient, and the result was 0.74, indicating that there was high consistency between different raters and the subjective test results were credible.

4.3 Experimental Discussion

Experimental results show that the proposed method has a stable comprehensive performance in multi-source scenes, and its advantages are mainly reflected in the improvement of reconstruction accuracy and the stability of recognition output. In the reconstruction experiment, the SSIM of the model reaches 0.913, the PSNR reaches 34.72 dB, and the edge preservation rate is 91.6%, which indicates that the building contour, street and lane interface and landscape details can be recovered completely after multi-source data collaboration. In the recognition experiment, the accuracy, recall and F1 value reach 94.3%, 92.8% and 93.5% respectively, and the correlation coefficient with the expert score reaches 0.887, which indicates that the output of the model is not only the judgment of the image level, but also can stably enter the protection level division and update priority analysis. Subjective evaluation results also supported this conclusion, with an effective restoration rate of 86.7% and a Fleiss 'Kappa coefficient of 0.74, indicating that the digital results also have strong credibility at the level of human perception. This performance is closely related to the model structure design.

The attention mechanism strengthens the key features such as roof shape, wall material, and street and lane closure relationship, and multi-source fusion alleviates the shortcomings of single image in occlusion, scale and semantic expression. However, the experimental samples are still mainly typical villages in the central and eastern parts of the country, with insufficient coverage of extreme terrain, high-density reconstruction areas and a few special material scenes. Therefore, the generalization ability of the model in more complex areas needs to be further verified. The follow-up research can introduce more fine-grained time series monitoring data on the basis of expanding cross-regional samples, so as to enhance the continuity and adaptability of digital protection renewal.

4.4 Discussion of Results

Compared with the comparison methods, the improvement of the proposed model is not a local improvement of the single performance, but reflects in a more complete linkage relationship between the reconstruction results, the recognition output and the update decision. In the reconstruction stage, the SSIM of the model reaches 0.913 and the PSNR reaches 34.72 dB, which indicates that the expression of building contour, material texture and street lane interface after multi-source fusion is closer to the real scene. In the recognition stage, the F1 value reaches 93.5%, and the correlation coefficient with the expert score is 0.887, indicating that the model output has the actual value to enter the protection renewal judgment. In the subjective evaluation, the effective restoration rate is 86.7%, which further indicates that the results are not only "computable", but also "credible". However, this performance improvement comes at a computational cost. Due to the introduction of the attention mechanism and the multi-source feature fusion module, the single-round training time of the model in this paper is about 128 s, which is significantly higher than the 74 s of the ordinary convolution model, and the peak video memory is also increased from 7.6 GB to 11.4 GB. This means that the method is more suitable for scenarios with high accuracy requirements such as county landscape census, key village documentation and repair scheme calculation, while further lightweight is still needed in grassroots units with weak equipment conditions. At the same time, the samples in this paper are still mainly traditional villages in the Central Plains and southeast regions, and the coverage of extreme terrain, high-density reconstruction areas and ethnic minority settlements is insufficient, indicating that there is still room for improvement in the cross-regional adaptability of the model. In terms of method significance, the value of this paper does not lie in simply stacking algorithm modules, but in putting spatial form recognition, cultural semantic extraction and updating decision mapping into the same computing framework, which provides a more operable technical pivot for the digital protection of traditional villages from static display to dynamic governance.

5 Conclusion

Aiming at the protection and renewal requirements of traditional village landscape in the implementation of rural revitalization strategy, this paper constructs a digital technical framework that integrates multi-source acquisition, 3D reconstruction, landscape recognition and renewal decision-making, and introduces the attention mechanism into the fusion processing of aerial photography, ground photos, point clouds and text data. So that the landscape shape information and historical semantic information can be co-expressed in the same computing link. The experimental results show that the proposed method achieves good results in the digital reconstruction task of traditional village buildings and landscape features, with SSIM of 0.913, PSNR of 34.72 dB, and edge preservation rate of 91.6%. In the protection update recognition task, the accuracy, recall and F1 value reach 94.3%, 92.8% and 93.5%, respectively, and the correlation coefficient with the expert score reaches 0.887. In the subjective evaluation, the effective restoration rate of style is 86.7%, which indicates that the model output not only has high calculation accuracy, but also has strong actual explanatory power. The research shows that the key to the digital protection and renewal of traditional villages is not to improve the accuracy of single modeling in isolation, but to connect the spatial reconstruction, style discrimination and renewal proposal generation, so that the digital results can truly enter the process of repair design, zoning control and dynamic governance. Of course, the sample of this paper is still dominated by typical villages, and the coverage of extreme terrain, complex reconstruction areas and special regional materials is still insufficient. In the future, cross-regional data samples can be expanded, and cloud storage, time series monitoring and lightweight deployment technologies can be combined to improve the applicability of the model in a wider range of rural scenes.

Funding

Project number: 262400410212; Project name: 2026 Henan Soft Science Project "Study on the Optimization Path of Rural Tourism Driving Henan Rural Comprehensive Revitalization from the Perspective of" Three-life Space "System Reconstruction"

References

[1] Yu L, Zhang X, He F, et al. Participatory Historical Village Landscape Analysis Using a Virtual Globe-Based 3D PGIS: Guizhou, China[J]. Sustainability, 2022, 14(21): 14022.

[2] Li X, Yang Y, Sun C, et al. Investigation, evaluation, and dynamic monitoring of traditional Chinese village buildings based on unmanned aerial vehicle images and deep learning methods[J]. Sustainability, 2024, 16(20): 8954.

[3] Wang Z, Zhu J, Wu Z. Study on the spatial distribution characteristics of traditional villages and their response to the water network system in the lower yangtze river basin[J]. Scientific Reports, 2024, 14(1): 22586.

[4] Meng S, Liu C, Zeng Y, et al. Quality evaluation of public spaces in traditional villages: A study using deep learning and panoramic images[J]. Land, 2024, 13(10): 1584.

[5] Chen W, Yang L, Wu J, et al. Spatio-temporal characteristics and influencing factors of traditional villages in the Yangtze River Basin: A Geodetector model[J]. Heritage Science, 2023, 11(1): 1-15.

[6] Wang T, Chen J, Liu L, et al. A review: How deep learning technology impacts the evaluation of traditional village landscapes[J]. Buildings, 2023, 13(2): 525.

[7] Lovell L J, Davies R J, Hunt D V L. The application of historic building information modelling (HBIM) to cultural heritage: a review[J]. Heritage, 2023, 6(10): 6691-6717.

[8] Sampaio A Z, Tomé J, Gomes A M. Heritage building information modelling implementation first steps applied in a castle building: historic evolution identity, data collection and stratigraphic modelling[J]. Heritage, 2023, 6(10): 6472-6493.

[9] Penjor T, Banihashemi S, Hajirasouli A, et al. Heritage building information modeling (HBIM) for heritage conservation: Framework of challenges, gaps, and existing limitations of HBIM[J]. Digital Applications in Archaeology and Cultural Heritage, 2024, 35: e00366.

[10] Pham Huu T, Van Westen A C M, Zoomers A. Compensation and resettlement policies after compulsory land acquisition for hydropower development in Vietnam: policy and practice[J]. Land, 2013, 2(4): 678-704.

[11] He C, Liang Y W, Zhang S Y. A study on the spatial structures and mechanisms of intangible cultural heritage and traditional villages in the Dongting Lake Basin[J]. Buildings, 2024, 14(6): 1736.

[12] Zhang M, Shen C, Gu W, et al. Identification of traditional village aggregation areas from the perspective of historic layering: evidence from hilly regions in Zhejiang Province, China[J]. Land, 2023, 12(12): 2088.

[13] Guo L, Ma W, Gong X, et al. Digital preservation of classical gardens at the San Su Shrine[J]. Heritage Science, 2024, 12(1): 1-22.

[14] Yamamuro A M, Rosi-Marshall E J, Lamberti G A, et al. Quantity, controls and functions of large woody debris in midwestern USA streams[J]. 2022.

[15] Li X, Yang Q, Lyu X, et al. Multidimensional framework for analyzing the distribution patterns of traditional villages in the karst landscape regions of China[J]. Ecological Informatics, 2023, 77: 102184.

[16] Gaspari F, Barbieri F, Fascia R, et al. An open-source web platform for 3D documentation and storytelling of hidden cultural heritage[J]. Heritage, 2024, 7(2): 517-536.

[17] Angheluță L M, Popovici A I, Ratoiu L C. A Web-based platform for 3d visualization of multimodal imaging data in cultural heritage asset documentation[J]. Heritage, 2023, 6(12): 7381-7399.

[18] Dipasquale L, Ammendola J, Montoni L, et al. Harnessing vernacular knowledge for contemporary sustainable design through a collaborative digital platform[J]. Heritage, 2024, 7(9): 5251-5267.

[19] Vital R, Sylaiou S. Digital survey: How it can change the way we perceive and understand heritage sites[J]. Digital Applications in Archaeology and Cultural Heritage, 2022, 24: e00212.

[20] Moreno M, Ortiz R, Cagigas-Muñiz D, et al. ART-RISK 3.0 a fuzzy—based platform that combine GIS and expert assessments for conservation strategies in cultural heritage[J]. Journal of Cultural Heritage, 2022, 55: 263-276.

[21] Wang W, Hei M, Peng F, et al. Development of “air-ground data fusion” based LiDAR method: Towards sustainable preservation and utilization of multiple-scaled historical blocks and buildings[J]. Sustainable Cities and Society, 2023, 91: 104414.

[22] Dang X, Liu W, Hong Q, et al. Digital twin applications on cultural world heritage sites in China: A state-of-the-art overview[J]. Journal of Cultural Heritage, 2023, 64: 228-243.