Cultural Translation Methods of Multi-Ethnic Mural Heritage Based on Digital Twins and Knowledge Graphs

Fei Sun[1], **Li Wang**[2]
[1]Lu Xun School of Arts, Yan'an University, Yan'an 716000, Shaanxi, China
[2]College of Art and Design, Xi'an Innovation College of Yan'an University, Xi'an 710100, Shaanxi, China
Corresponding author: Li Wang (e-mail: proligent@163.com)

ABSTRACT The preservation and interpretation of multi-ethnic mural heritage are challenged by fragile material carriers, heterogeneous cultural semantics, and the lack of unified digital representation. To address these issues, this study proposes a cultural translation framework based on Digital Twins (DT) and Knowledge Graphs (KG), integrating geometric perception and semantic reasoning into a dual-domain alignment architecture. Multi-view three-dimensional point clouds are processed using PointNet++ to achieve semantic segmentation of ethnic styles, while Graph Attention Networks (GAT) and Cross-modal Contrastive Learning (CCL) establish visual-semantic correspondence within a unified embedding space. Dynamic physical parameters from the digital twin environment are further incorporated into a Relational Graph Convolutional Network (R-GCN) to generate structured cultural translation assertions with enhanced robustness. Experimental results demonstrate style recognition accuracies of 88.45%, 86.70%, and 94.35% for Tibetan, Mongolian, and Uyghur murals, respectively, outperforming representative baseline methods in semantic segmentation and cross-modal alignment tasks. The proposed framework enables closed-loop interaction between geometric reconstruction and semantic inference, providing an efficient engineering paradigm for intelligent cultural heritage preservation. Moreover, the integration of digital twin sensing, multimodal information fusion, and wireless data acquisition offers valuable references for electromagnetic sensing systems, antenna-enabled digital monitoring platforms, and next-generation intelligent perception networks in complex heritage environments.

INDEX TERMS digital twin, knowledge graph, cross-modal alignment, electromagnetic sensing, intelligent monitoring networks

I. INTRODUCTION

MULTI-ethnic mural heritage, as a visual archive of exchanges and integration among ethnic groups along the Silk Road and the Qinghai-Tibet Plateau, carries diverse cultural information, including clothing patterns, religious narratives, and architectural components. This information not only records the technological exchange between different ethnic groups in visual narratives but also reflects deep cultural identity and evolutionary processes [1-2]. However, fluctuations in the cave microenvironment, salt crystallization, and microbial activity cause irreversible degradation of the ground layer and pigment layer [3-4]. Traditional photogrammetry and 2D scanning can only record surface color information and cannot preserve the micro-undulations of cultural symbols and the spatial depth of multi-ethnic overlay painting [5-6]. Therefore, constructing a DT-based method, which generates the precise geometric shape of the mural material layer through multi-view 3D reconstruction and semantic segmentation, and decodes the deep cross-ethnic cultural semantics within it, has become a fundamental requirement in the intersection of heritage digital protection and cultural computing.

Current research faces three specific challenges. First, the ambiguity of geometric boundaries: the surfaces of murals from various ethnic groups often show signs of multiple repaintings and repairs, and the differences in the particle size of mineral pigments used in the brushstrokes of different ethnic groups and the polishing marks of the ground layer are slight [7-8], making it impossible for single-color imaging to distinguish the painting boundaries and the order of chronological superposition. Multi-view reconstruction generates a point cloud that records both spatial coordinates and color values, where micro-relief and stroke thickness variations induced by distinct mineral pigments are encoded as local neighborhood distribution patterns. Geometric processing of the point cloud resolves boundary ambiguity through these encoded features. Second, semantic heterogeneity: the patterns, shapes, and other cultural elements appearing in murals have their own nomenclature and symbolic meanings in the iconographic traditions of different ethnic groups [9-10]. Heterogeneous terms in archaeological documents have not been linked to digital images at the pixel level, making cultural origin tracing rely on expert experience. Third, the lack of dynamic correlation: the physical aging process of mural materials (such as pigment blackening and ground layer cracking) directly affects the identification characteristics of cultural symbols [11-12]. Existing systems separate environmental monitoring data from image semantic analysis, losing information on the dynamic impact of material state changes on the confidence level of cultural semantic discrimination [13-14]. To address the aforementioned challenges, some studies have attempted to segment the outlines of murals using deep learning networks [15-16]. However, this type of supervised learning method struggles to handle regions with unclear boundaries where styles vary across different ethnic groups [17]. Some scholars have constructed knowledge graphs (KGs) for murals of specific themes, associating image tags with knowledge entries [18-19]. However, the entity extraction process is independent of the DT model of the mural, and high-precision grid vertices cannot be automatically mapped to knowledge graph nodes [20]. Other studies have used DT technology to simulate the cave microenvironment [21-22]. However, the twin only transmits geometric coordinates and temperature and humidity data, without embedding a cultural semantic layer [23]. The above methods are limited by the separation of geometry and semantics: deep learning segmentation methods do not establish a connection between segmentation results and knowledge systems, KG construction does not utilize the geometric boundary constraints of the 3D model, and DT simulation does not introduce cultural semantic nodes. Each of the three paths lacks a bidirectional alignment and joint reasoning mechanism between the digital geometric model and the knowledge semantic model.

This work defines cultural translation as the mapping from digital twin geometric features and visual observations to structured semantic assertions within the knowledge graph that encode cross-ethnic cultural knowledge. To address the lack of cross-modal alignment between the DT geometric model and the KG semantic model, this paper proposes a multi-ethnic mural heritage cultural translation method based on DT and KG. First, PointNet++ is used to perform ethnic style semantic segmentation on the mural point cloud, outputting geometric features carrying cultural tags from multiple ethnic groups, including Tibetan, Mongolian, and Uyghur. Second, GAT and CCL are used to map visual features and KG text nodes to a unified embedding space. Finally, the physical state parameters of the DT volume, dynamically simulated based on historical meteorological data, are used as dynamic attributes and input together with the KG into the R-GCN, outputting structured cultural translation assertions. This paper constructs a DT-KG dual-domain alignment framework, realizing a closed-loop cultural translation from geometric perception to semantic reasoning. The dual-domain alignment framework can provide quantitative evidence for preventive conservation decisions of mural heritage and provide an engineering paradigm for the establishment of cross-modal digital archiving standards.

II. DUAL-DOMAIN JOINT TRANSLATION METHOD

A. SEMANTIC SEGMENTATION OF ETHNIC STYLE POINT CLOUDS BASED ON POINTNET++

The DT branch serves as the geometric basis for cultural translation, with data flow beginning with multi-view image acquisition. Image sequences are reconstructed into 3D point cloud data using SfM (Structure from Motion) technology [24-25]. The original point cloud undergoes statistical filtering to remove outliers, and point cloud density is unified through voxel grid downsampling. The processed point cloud data then enters the PointNet++ network for semantic segmentation of ethnic style point clouds. This segmentation yields geometric features tagged with ethnic labels, providing the foundational geometric basis for the subsequent knowledge graph alignment. This network utilizes a multi-layer ensemble abstraction structure to extract local neighborhood geometric features. The ensemble abstraction layer consists of sampling, grouping, and feature extraction steps. Max pooling aggregates feature information within local regions.

The feature aggregation process is expressed as:

$$F_k = \max_{i \in S_k} \{\mathrm{MLP}(p_i, f_i)\} \tag{1}$$

F_k represents the aggregated feature vector of the k-th local region; $\max$ represents element-wise max pooling; S_k represents the neighborhood point set based on the centroid; p_i is the spatial coordinate of the i-th point in the neighborhood; f_i corresponds to the initial feature vector of the point; MLP (Multi-Layer Perceptron) is the mapping function.

The encoder downsamples layer by layer and increases the feature abstraction level. The decoder recovers the point cloud density through interpolation and propagates high-level semantic features to the original resolution. The network training process uses the cross-entropy loss function to optimize the parameters. The loss is calculated as follows:

$$L_{seg} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \tag{2}$$

L_{seg} represents the total semantic segmentation loss; C is the total number of ethnic style categories; y_c represents the true label indicator variable for the c-th class; $\hat{y}_c$ corresponds to the probability value of the c-th class predicted by the network.

The network output layer calculates the probability distribution of each point belonging to the Tibetan, Mongolian, or Uyghur ethnic style, thereby generating a semantic segmentation result carrying an ethnic style label (each point obtains a category corresponding to the highest probability)

[26-27]. This result can be regarded as a pointwise mapping from the geometric-color space to the ethnic semantic space. This result can be regarded as a pointwise mapping from the geometric-color space to the ethnic semantic space. Local neighborhood distribution patterns in the point cloud encode micro-relief and stroke thickness variations induced by distinct mineral pigments and ground layer processing techniques of each ethnic group. These geometric features directly correspond to ethnic stylistic attributes. The PointNet++ local feature aggregation module learns statistical associations between these patterns and expert-annotated labels without relying on pre-labeled metadata. The mapping from low-level geometric measurements to high-level semantic labels thus emerges from data-driven feature extraction and classification.

The point cloud segmentation label is mapped and passed to the vertices of the mesh model generated by reconstructing the Poisson surface. Based on the highest probability category label of each vertex in the segmentation result, the DT volume is divided into independent ethnic style sub-mesh. The sub-mesh retains the original topological connectivity and texture mapping information. The segmented sub-mesh and its texture image patch form the visual input that bridges the digital twin geometry to the knowledge graph semantic nodes.

Figure 1 consists of a DT branch, a KG branch, a cross-modal alignment module, and a joint inference component. The DT branch reconstructs the point cloud and performs semantic segmentation using SfM technology. The KG branch stores entity relationships using RDF (Resource Description Framework). GAT and CCL complete the visual-semantic mapping. R-GCN combines dynamic physical parameters to generate cultural translation assertions. Data flow runs through the four modules, forming a closed-loop processing chain.

B. KG CONSTRUCTION FOR MULTI-ETHNIC MURALS

The KG branch is constructed independently of the DT branch. During the ontology design phase, based on the Complete Collection of Dunhuang Murals and archaeological documents on murals from multiple ethnic groups, three cultural concept systems—Tibetan, Mongolian, and Uyghur—are defined. There are 4 types of entity categories: ethnicity, motif, pattern, technique. The entity categories of motif are classified into 3 subcategories respectively: religious narrative, production and daily life, decorative borders. The entity categories of pattern are classified into corresponding type according to ethnic traditions, for example, Tibetan Eight Auspicious Symbols, Mongolian Cloud Patterns, Uyghur Plant Patterns, etc. Technique entities record mineral pigment types, brushstroke characteristics, and ground treatment techniques.

In the phase of the definition of relationship type, two kinds of semantic connection type are defined: inheritance and fusion. The inheritance relationship represents the historical transmission path of the cultural entities within the same ethnic group while the fusion relationship represents the historical borrowing of patterns and technique exchange between different ethnic groups. The instances of relationship between individuals are manually annotated from archaeological information and mural iconography. Class hierarchy and attribute constraint are finished in ontology editing by using Protégé tool.

The entity extraction and instantiation process follows these rules: Each entity node corresponds to an instance of a recognizable cultural element in a mural. Node attributes include ethnic affiliation, era, cave number, and spatial coordinate range.

The semantic organization of the knowledge graph is presented through a hierarchical structure of ontology and instance layers. Figure 2 shows the ontology structure and instance fragments of the multi-ethnic mural knowledge graph.

Figure 2 shows the ontology layer on the left and the instance layer on the right. The ontology layer uses a tree structure, containing four entity types: ethnicity, motif, pattern, and technique. The ethnicity type is further divided into three sub-categories: Tibetan, Mongolian, and Uyghur. The motif type is divided into three subcategories: religious narrative, production and daily life, and decorative borders. The pattern type displays the corresponding ethnic patterns. The technique type displays mineral pigments, brushstroke characteristics, and ground treatment. Inheritance relationships between types in the ontology layer are indicated by solid arrows, and fusion relationships by dashed arrows. The instance layer on the right presents specific examples of mural cultural elements in a diagram structure, with each instance node filled with a different color according to its entity category. Inheritance relationships within the same ethnic group are connected by solid blue lines, and fusion relationships across ethnic groups are connected by dashed red lines. The instance mapping is indicated by a gray dashed arrow between the ontology layer on the left and the instance layer on the right.

C. VISUAL-SEMANTIC ALIGNMENT BASED ON GAT AND CCL

The mapping process between ethnic-style sub-mesh image blocks and KG text nodes of multi-ethnic murals relies on GAT and CCL to construct a unified embedding space. KG is stored using RDF, with entity sets encompassing ethnicity, motifs, patterns, and techniques, and relation sets defining inheritance and fusion semantic connections. The graph structure is represented as $G = (V, E)$, where V represents the set of entity nodes and E represents the set of relation edges. The text encoder uses GAT to aggregate neighbor node information to generate text vectors, and node feature updates depend on attention coefficient calculation.

$$e_i' = \sigma\left(\sum_{j\in\Omega_i} \alpha_{ij} W e_j\right) \tag{3}$$

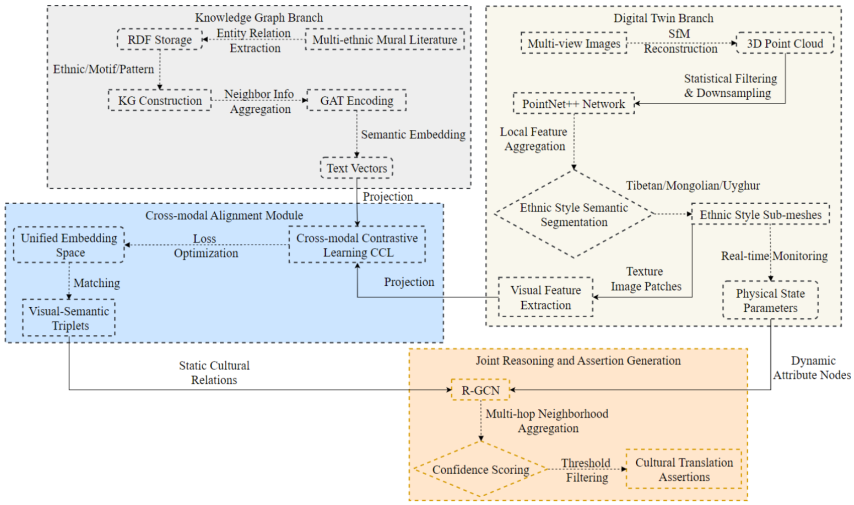

FIGURE 1. Mural culture translation based on DT and KG dual-domain alignment

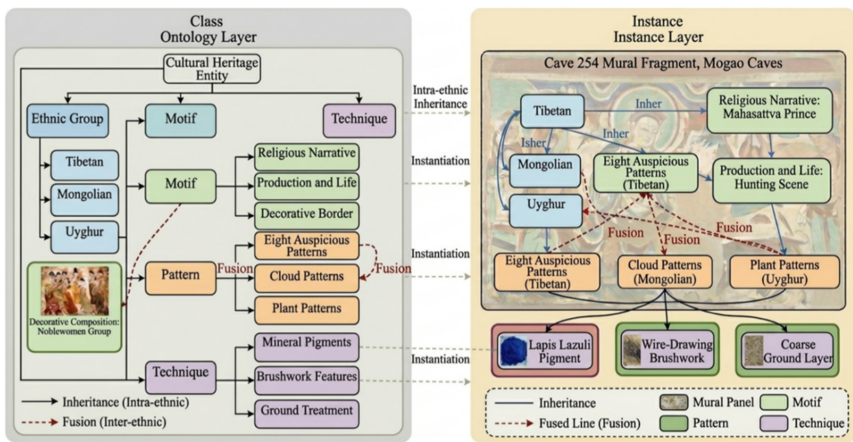

FIGURE 2. Ontology and instance structure of the knowledge graph of multi-ethnic murals

e_i' represents the updated node embedding vector; Ω_i represents the set of neighboring nodes of node i; α_{ij} represents the attention coefficient of node j on node i; α_{ij} is obtained by calculating the correlation between node i and neighbor j through a feedforward network and then normalizing it using softmax. W represents the trainable weight matrix; σ represents the nonlinear activation function; e_j represents the neighbor node embedding vector.

Image patch feature extraction uses a residual network to process the output ethnic-style sub-mesh texture image. Visual feature vectors and text vectors are mapped to a shared dimension space via independent projection layers. The projection operation is implemented through a linear transformation, and the projected vectors are L2 normalized to use cosine similarity as a distance metric. The parameters of the visual projection matrix and the text projection matrix are optimized independently.

The CCL framework calculates the similarity distance between matching image-text pairs. A normalized temperature-scale cross-entropy loss function is used to bring matching pairs closer and separate mismatched pairs in the embedding space.

$$L_{ccl} = -\frac{1}{B}\sum_{k=1}^{B}\log\frac{\exp(\mathrm{sim}(z_k^v, z_k^t)/\tau)}{\sum_{m=1}^{B}\exp(\mathrm{sim}(z_k^v, z_m^t)/\tau)} \tag{4}$$

L_{ccl} represents the contrastive loss value; B represents the batch size; sim represents the cosine similarity function; z_k^v represents the visual feature embedding vector; z_k^t represents the text feature embedding vector; τ represents the temperature scaling parameter.

Each aligned image patch is associated with a corresponding KG subgraph, and the subgraph nodes carry updated semantic embedding vectors. The KG semantic nodes are updated via GAT and fed back to the DT volume segmentation module. The semantic embedding vectors inversely constrain the point cloud segmentation decision boundary, forming a geometric-semantic bidirectional closed loop.

During training, gradients are backpropagated to update the projection matrix and encoder parameters until the loss function converges. The embedding space distribution reflects the correspondence between visual features and semantic concepts.

D. R-GCN-BASED CULTURAL TRANSLATION ASSERTION GENERATION

The aligned visual-semantic subgraphs constitute the static topology of the R-GCN. The physical state parameters of the DT volume, dynamically simulated based on historical meteorological data, are instantiated as dynamic attribute nodes in the graph structure. Time-series data on temperature gradient, relative humidity, and pigment aging constitute the set of physical state parameters. Each relation type learns an independent weight matrix in the neural network layers. A message-passing mechanism aggregates neighbor node information within a multi-hop neighborhood. The node hidden state update process is driven by a combination of linear transformations and nonlinear activation functions.

$$h_i^{(l+1)} = \sigma\left(\sum_{r\in R}\sum_{j\in N_i^r}\frac{1}{c_{i,r}}W_r^{(l)}h_j^{(l)} + W_0^{(l)}h_i^{(l)}\right) \tag{5}$$

$h_i^{(l)}$ represents the feature vector of node i in layer l; R represents the set of relation types; N_i^r is the set of neighboring nodes of node i under relation r; $c_{i,r} = |N_i^r|$ represents the number of neighboring nodes of node i under relation r, used to normalize the aggregation result. $W_r^{(l)}$ corresponds to the trainable weight matrix of relation r; $W_0^{(l)}$ is the self-looping weight matrix; σ represents the activation function.

The feature vectors of dynamic attribute nodes are updated with time steps and participate in the above aggregation operation. After multiple convolutional operations, the node embedding vector contains static cultural semantics and dynamic physical state information. The confidence scores of candidate cultural translation assertions are generated through mapping by fully connected layers. The output layer projects the node embedding vectors onto the assertion space.

$$s_k = \mathrm{softmax}\left(W_{out}h_k^{(T)} + b_{out}\right) \tag{6}$$

s_k represents the confidence score of the k-th candidate assertion; W_{out} is the output layer weight matrix; $h_k^{(T)}$ is the embedding vector of node k in the last layer; b_{out} is the bias term.

Assertions with confidence scores higher than a preset threshold are selected as the final cultural translation result. The structured assertion output consists of a triplet of subject entity, relational predicate, and object entity. The generated structured assertions are directly input into the subsequent experimental verification stage. The structured assertions are back-mapped to the vertices of the DT volume grid, updating the ethnic style label confidence distribution of that vertex, realizing the dynamic correction of the geometric model by the semantic reasoning results.

III. EXPERIMENT CONFIGURATION AND DATASET CONSTRUCTION

Experimental data are sourced from the International Dunhuang Project (IDP) and the Digital Dunhuang image database [28-29], Digital achievements of the murals in the tomb of Prince Zhanghuai of Tang Dynasty. Point cloud data underwent a preprocessing process, with the voxel grid downsampling side length set to 0.005 meters, resulting in a point cloud density of approximately 50,000 points per square meter.

The experimental data were selected from the digitization of Tang Dynasty tomb mural heritage. Original mural images were reconstructed into point cloud data through multi-view 3D reconstruction, and then processed using statistical filtering for noise reduction and voxel mesh downsampling. Some representative multi-ethnic mural experimental samples are shown in Figure 3.

Figure 3 shows three original mural images from the tomb of Prince Zhanghuai of the Tang Dynasty. The images depict a group of ladies in richly patterned clothing and a dynamic hunting activity. The style showcases equestrian figures in a line-drawing style. The images reveal the interwoven cultural information of the multi-ethnic groups during the Tang Dynasty. The original image data was transformed into point cloud models using 3D reconstruction technology. Visual information, including color, texture, and geometric details, was preserved. The diversity of samples supports subsequent semantic segmentation and knowledge graph construction tasks. Multi-source data fusion technology enables the association between geometric features and semantic information. Experimental samples involve different themes and styles. The research process verifies the applicability of digital twin and knowledge graph methods.

The semantic segmentation of the point cloud is independently completed by three mural research experts, with annotation categories including Tibetan, Mongolian, and Uyghur ethnic styles. Inter-expert annotation consistency is evaluated using the Kappa coefficient, which is 0.87. The dataset is divided into training, validation, and test sets in a 7:2:1 ratio.

KG construction is based on the publicly available literature, Complete Collection of Dunhuang Murals [30-31], and entity and relation definitions are completed using the Protégé ontology editing tool.

Dynamic environmental parameters are generated using publicly available historical meteorological data from National Oceanic and Atmospheric Administration (NOAA) passed through a cave microclimate transfer model [32-33]. The transfer model maps macro-scale weather variables to mural surface micro-environmental conditions using cave wall thermal properties and air exchange rates as physical constraints. Local temperature and humidity time series computed by the transfer model drive the pigment aging model to simulate blackening progression. Discrepancies persist between NOAA macro-data and true cave microclimate. The experiment is based on accelerated aging

FIGURE 3. Mural in the Tomb of Prince Zhanghuai of the Tang Dynasty

simulation of macro trends, and accurate twinning relies on on-site sensor data. Time-series data on temperature gradient, relative humidity, and pigment blackening degree are generated at a daily sampling frequency. The aging model parameters are set based on accelerated aging experimental data of mineral pigments.

The experimental environment consisted of an NVIDIA RTX 3090 graphics processor and the PyTorch framework. The mIoU (mean Intersection over Union) metric is used for point cloud semantic segmentation. The recall rate (Hit@K) is used for cross-modal alignment. The F1 score is used for cultural translation assertion generation. The comparison methods are shown in Table 1. All comparison methods used the same dataset and training/validation split ratio as the method presented in this paper.

Table 1 shows four comparison methods covering two task dimensions: point cloud semantic segmentation and knowledge reasoning. Point-MAE (Point-Masked Autoencoder) and CLIP (Contrastive Language-Image Pre-Training) are point cloud segmentation pre-training methods, used to verify the performance of the proposed PointNet++ in fine-tuning and zero-shot scenarios, respectively. SAM (Segment Anything Model) projects the segmentation results of the 2D base model to 3D, contrasting it with the direct point cloud segmentation. TGN (Temporal Graph Network) serves as the dynamic graph reasoning baseline, comparing its generation performance with the proposed R-GCN for cultural translation assertions. The operating environment and data partitioning of all comparison methods are consistent with the proposed method.

IV. RESULTS ANALYSIS

A. POINT CLOUD SEMANTIC SEGMENTATION ACCURACY ANALYSIS

The ambiguous geometric boundaries of multi-ethnic mural heritage pose challenges to style recognition. This study constructs a point cloud semantic segmentation experiment based on DT volume to quantitatively evaluate the classification accuracy of different ethnic style regions. The experiment chooses Tibetan, Mongolian and Uyghur styles as semantic class. Statistics based on confusion matrix can analyze the mapping relationship between true label and predicted label, and reflect the discriminative ability of model in style transition zone. mIoU index represents the overall performance, and confusion matrix can show the confusion situation between different categories, which reflects the historical fusion phenomenon of adjacent ethnic style and causes misclassification. The comparison of proposed method with mainstream baseline model on test set, can reflect the discriminative ability of local feature aggregation module on geometric difference between local feature.

Figure 4 shows the normalized confusion matrices of the four methods on the ethnic style point cloud segmentation task. The diagonal values represent the proportion of correct classifications, while the off-diagonal regions indicate the degree of style confusion.

Fig. 4(a), method accuracy for Tibetans is 88.45% and for Mongolians is 86.70%, and for Uyghurs is 94.35%. Note that, off-diagonal region value is about 8% for Tibetan and Mongolian style confusion, which is induced by the blending of historical styles and thus blur geometric boundaries.

Fig. 4(b) is the method accuracy of Point-MAE, method accuracy for Tibetans is 81.20% and confusion on Tibetan

TABLE 1. Summary of Comparison Methods

Comparison Methods	Core Principles	Evaluation Indicators	Features
Point-MAE [34]	Masked autoencoders are used for self-supervised pre-training of point clouds, followed by downstream fine-tuning of segmentation.	mIoU	Suitable for pre-training of large-scale point clouds.
CLIP2Point [35]	Transfer CLIP text-image alignment knowledge to 3D point clouds.	mIoU, Hit@K	Zero-sample transfer capability.
SAM [36]	Backprojection of 2D SAM segmentation results to 3D point cloud segmentation.	mIoU	Depends on 2D segmentation accuracy.
TGN [37]	Temporal graph neural networks process dynamic attribute nodes.	F1 value	Support for continuous-time dynamic graphs.

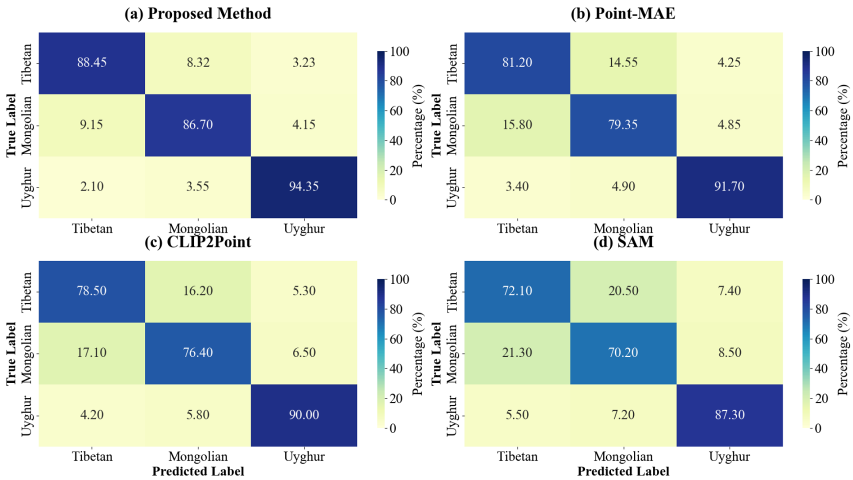

FIGURE 4. Comparison of semantic confusion matrices for point cloud semantic segmentation in multiple ethnic styles: (a) Proposed Method; (b) Point-MAE; (c) CLIP2Point; (d) SAM

and Mongolian style is 14.55%. Due to the poor matching between pretrained features of mask autoencoder and local geometry of the mural, the discriminative ability decreases. In Figure 4(c), the accuracy of CLIP2Point for Tibetans drops to 78.50%, and the confusion rate between Tibetans and Mongolians reaches 16.20%. The zero-shot transfer strategy struggles to adapt to the subtle geometric differences in specific ethnic patterns. Figure 4(d) shows that the accuracy of SAM for Tibetans is 72.10%, and the confusion rate between Tibetans and Mongolians is 20.50%. The loss of depth information during the backprojection of the 2D segmentation results to 3D causes the highest misclassification rate. The proposed method enhances the identification of subtle geometric differences through the PointNet++ local feature aggregation module, reducing the probability of misclassification in style transition zones. High-precision segmentation of the DT volume geometric basis provides a reliable basis for subsequent cross-modal cultural translation.

B. CROSS-MODAL VISUAL-SEMANTIC ALIGNMENT ANALYSIS

The cross-modal alignment results validate the mapping quality between DT volume visual features and KG semantic nodes. The DT geometric basis provides visual input. The experiment uses t-SNE (t-Distributed Stochastic Neighbor Embedding) to project high-dimensional embedding vectors onto a two-dimensional plane to observe the distribution of different cultural entity categories. The categories to be analyzed cover Tibetan clothing, Mongolian clothing, Uyghur patterns, and religious implements. These entities represent significantly different cultural symbols in murals from multiple ethnic groups. The cluster density and inter-class separation in the embedding space directly reflect the ability of the GAT and CCL modules to eliminate semantic heterogeneity. The coordinate axis dimensions represent the feature components after dimensionality reduction, and the scatter point positions indicate the relative distance of samples in the unified semantic space. The comparison results of the distribution patterns are shown in Figure 5.

Figure 5(a) shows the feature space state without the introduction of contrastive learning constraints, based on the baseline model embedding distribution. Tibetan clothing samples are distributed over a range of 10 units along the horizontal axis, while Mongolian clothing and Tibetan clothing regions show visual overlap, indicating that semantic heterogeneity leads to a large distance between the embedding vectors of similar entities. Figure 5(b) shows the embedding distribution of the proposed method, exhibiting a tight clustering pattern. On the horizontal axis, the coordinates of the clothing sample come from Tibetan, and the range of values converges to -10.45 to -9.89, decreasing the intra-class variance. The four domains occupy separate quadrants, and the religious artifacts and Uyghur patterns are well separated. GAT aggregation of neighbor node information enhances semantic representation consistency, CCL shortens the distance between matched image-text pairs. The temperature scaling parameter improves the accuracy

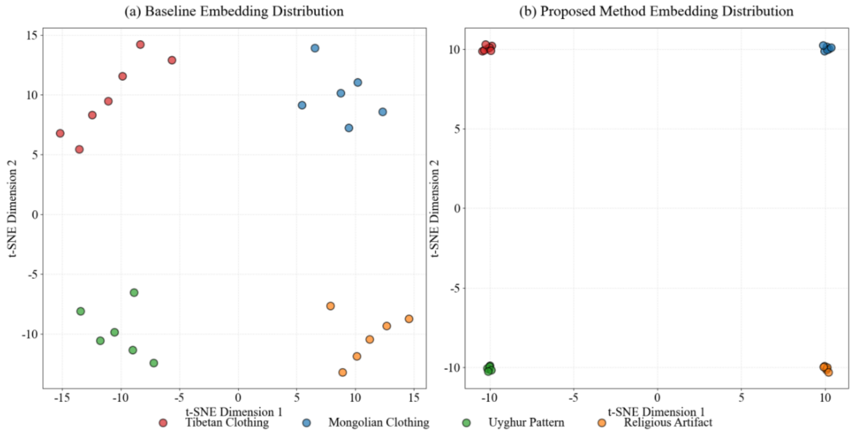

FIGURE 5. Visualization of the spatial distribution of cross-modal visual-semantic alignment embeddings: (a) Baseline Embedding Distribution; (b) Proposed Method Embedding Distribution

of the similarity measurement and promotes the separation of the mismatched pairs in the embedding space. Structured separation of feature space proves that the dual-domain alignment framework can effectively solve the problem of referential heterogeneity.

C. STABILITY ANALYSIS OF CULTURAL TRANSLATION ASSERTIONS

To study the influence of changes in the physical state of mural materials on semantic reasoning, this study performs an experiment to evaluate the sensitivity of cultural translation assertions. The experiment uses F1 score as the basic evaluation index to measure the accuracy and completeness of structured assertion generation. This index balances precision and recall and is applicable to the performance evaluation of data distribution with imbalance. The compared models are the method in this paper, which contains dynamic physical attribute nodes; a static baseline model without dynamic attributes; and a time-series graph neural network baseline. By simulating the accelerated aging process of murals and different levels of environmental stress, the fluctuations in model performance over time and stress are observed.

In Figure 6, the shaded area represents the standard deviation of multiple independent experiments. The line trend denotes the model performance change trend under different time spans and stress levels. Solid and dashed lines denote different methods and marked points denote the mean at certain time/stress level, respectively. The horizontal axis denotes the number of simulated aging days and environmental stress level, respectively. The vertical axis denotes the F1 score of assertion generation.

Figure 6(a) illustrates the necessity of dynamic parameters. The F1 score of the proposed method decreased from 0.912 to 0.865, and the static baseline decreased from 0.910 to 0.642. The static model lacks dynamic physical parameters, making it difficult to adapt to pigment blackening and leading to performance degradation. The static baseline decline accelerated after 200 days, due to the nonlinear characteristics of material aging. The proposed method incorporates dynamic properties to maintain stability. Figure 6(b) illustrates robustness. At stress level 5, the proposed method has an F1 score of 0.848, and the TGN baseline is 0.726. KG static semantics constrain dynamic noise. The shaded area is narrower in the proposed method, with the standard deviation remaining below 0.021. The TGN baseline standard deviation widened to 0.047, due to the sensitivity of the pure data-driven model to noise. The dual-domain alignment framework utilizes static cultural relationships to suppress environmental fluctuations, while dynamic physical nodes compensate for material state changes in real time. This mechanism ensures that cultural translation assertions maintain high confidence in long-term simulations. Introducing dynamic physical parameters is necessary to maintain the stability of semantic inference.

V. CONCLUSION

The dual-domain alignment model developed in this paper benefits from joint inference between DT geometric model and KG semantic model. The local feature aggregation module of PointNet++ can aggregate the subtle geometric difference and suppress the ambiguity of ethnic style boundary. GAT with CCL can eliminate the semantic heterogeneity and obtain correct mapping from visual feature to text node. R-GCN can integrate dynamic physical state parameter and improve the stability of cultural translation assertion in material aging process. Experimental data show that this method outperforms existing baseline models in point cloud segmentation accuracy and cross-modal alignment. The introduction of dynamic attribute nodes suppresses environmental fluctuation interference, maintaining high confidence in semantic inference during long-term simulations. This result supports preventative conservation decisions for mural heritage and provides an engineering paradigm for establishing cross-modal digital archiving standards. This technological approach promotes the development of the intersection of heritage digital conservation and cultural computing. The dual-domain joint mechanism compensates

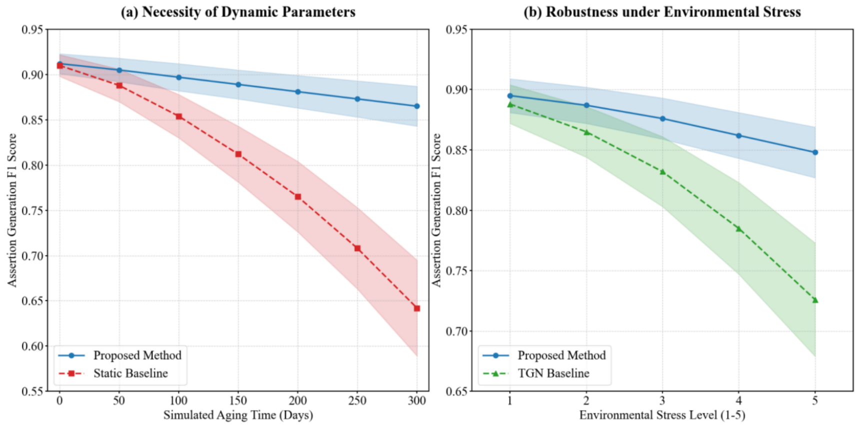

FIGURE 6. Stability of cultural translation assertions under dynamic environmental simulation: (a) Necessity of Dynamic Parameters; (b) Robustness under Environmental Stress

for the lack of bidirectional alignment between digital geometric models and knowledge semantic models, providing a new perspective for decoding deep information in complex cultural heritage and extending the technological boundaries of cultural heritage digital conservation.

AUTHOR CONTRIBUTIONS

Conceptualization –Fei Sun and Li Wang; methodology – Fei Sun and Li Wang; investigation – Fei Sun and Li Wang; writing-original draft preparation – Fei Sun and Li Wang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This work was supported by: 2025 Humanities and Social Sciences Research Project of the Ministry of

Education (Western and Border Regions): A Study on the Authentication and Multi-Ethnic Cultural Identity of Temple and Taoist Mural Remains from the Ming and Qing Dynasties in Northern Shaanxi* (Project No.: 25XJA760006) — Interim Results.

2025 Annual Project of Shaanxi Provincial Social Science Fund: Research on the Active Utilization of Shaanxi Cultural Heritage Resources under the National Cultural Digitization Strategy (Project No.: 2025J032) — Interim Results.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] J. Yu, B. Safarov, L. Yi, et al., "The adaptive evolution of cultural ecosystems along the Silk Road and cultural tourism heritage: A case study of 22 cultural sites on the Chinese section of the Silk Road World Heritage," Sustainability, vol. 15, no. 3, pp. 2465-2485, 2023, doi: 10.3390/su15032465.
[2] X. Xu, J. Zhang, S. Liu, et al., "Entropy change of historical and cultural heritage in traditional tibetan area of china based on spatial-temporal distribution pattern," Buildings, vol. 13, no. 12, pp. 2995-3023, 2023, doi: 10.3390/buildings13122995.
[3] C. Chen, H. Yang, X. Li, et al., "Hyperspectral estimation method for deterioration of rock carvings in the humid regions of southern China," Heritage Science, vol. 12, no. 1, pp. 1-19, 2024, doi: 10.1186/s40494-024-01226-0.
[4] I. Domingo and G. Barreda-Uso, "Knowledge-building in open-air rock art conservation: sharing the history and experiences with Levantine rock art," Studies in Conservation, vol. 68, no. 2, pp. 258-282, 2023, doi: 10.1080/00393630.2021.1996092.
[5] Y. Alshawabkeh and A. Baik, "Integration of photogrammetry and laser scanning for enhancing scan-to-HBIM modeling of Al Ula heritage site," Heritage Science, vol. 11, no. 1, pp. 1-17, 2023, doi: 10.1186/s40494-023-00997-2.
[6] I. Drofova, W. Guo, H. Wang, et al., "Use of scanning devices for object 3D reconstruction by photogrammetry and visualization in virtual reality," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 2, pp. 868-881, 2023, doi: 10.11591/ eei.v12i2.4584.
[7] C. Liu, R. Cui R, and Z. Wang, "Digital virtual simulation for cultural clothing restoration: Case study of tang dynasty mural diplomatic envoys from crown prince zhang huais tomb," Journal of Theoretical and Applied Electronic Commerce Research, vol. 19, no. 2, pp. 1358-1391, 2024, doi: 10.3390/jtaer19020069.
[8] L. Feng, J. Sun, T. Zhai, et al., "Planning for cultural connectivity: modeling and strategic use of architectural heritage corridors in Heilongjiang Province, China," Buildings, vol. 15, no. 12, pp. 1970-2012, 2025, doi: 10.3390/buildings15121970.
[9] W. Li, H. Lv, Y. Liu, et al., "An investigating on the ritual elements influencing factor of decorative art: based on Guangdong's ancestral hall architectural murals text mining," Heritage Science, vol. 11, no. 1, pp. 1-22, 2023, doi: 10.1186/ s40494-023-01069-1.
[10] J. Zhang and X. Yang, "Exploring the archaeological treasures of the silk road: Tibetan arts and crafts, traditional decorative patterns, and pattern symbols analysis," Mediterranean Archaeology and Archaeometry, vol. 23, no. 2, pp. 197-212, 2023, doi: 10.5281/zenodo.10958830.
[11] G. Germinario, L. Logiodice A, P. Mezzadri, et al., "Integrated investigations to study the materials and degradation issues of the urban mural Painting Ama Il Tuo Sogno by Jorit Agoch," Sustainability, vol. 16, no. 12, pp. 5069-5092, 2024, doi: 10.3390/su16125069.
[12] H. Wu X, J. Bai X, M. Chen D, et al., "Reversible multifunctional mural protective material with high durability, anti-aging, breathability, and harsh-environment resistance," Journal of Building Engineering, vol. 107, no. 1, pp. 112729-112749, 2025, doi: 10.1016/j.jobe.2025.112729.
[13] N. Zhang and G. Lv, "Digital Image Analysis and Pattern Recognition of the Artistic Expression and Evolution of Lotus Patterns in Goguryeo Murals in Jian," Mediterranean Archaeology and Archaeometry, vol. 24, no. 3, pp. 314-325, 2024.
[14] L. P. Kumar, N. B. Thomas, and N. A. George, "Network analysis of research articles on mural paintings," The European Physical Journal Plus, vol. 139, no. 12, pp. 1-20, 2024, doi: 10.1140/epjp/s13360-024-05887-5.
[15] B. Plengdeesakul, "Luang Prabang Mural Paintings: Social and Cultural Reflection, Lao PDR," Heranca, vol. 7, no. 4, pp. 51-63, 2024, doi: 10.52152/heranca.v7i4.969.
[16] M. Jia, J. Hu, Z. Yang, et al., "Automatic restoration of dunhuang murals

and process visualization method based on deep learning," Applied Sciences, vol. 15, no. 3, pp. 1422-1438, 2025, doi: 10.3390/app15031422.
[17] Y. Kang and C. Yang K C, "Using AI-Assisted Image Analysis Techniques to Examine Hollywood Movie Murals: A Zone of Intensive Mural Experience (ZIME) Perspective," Intercultural Communication Studies, vol. 34, no. 1, pp. 5-21, 2026, doi: 10.53941/ics.2025.100005.
[18] Y. Gao, Q. Zhang, X. Wang, et al., "Multidimensional knowledge discovery of cultural relics resources in the Tang tomb mural category," The Electronic Library, vol. 42, no. 1, pp. 1-22, 2024, doi: 10.1108/EL-04-2023-0091.
[19] Z. Zhou, H. Wang, Z. Zhao, et al., "Chartkg: A knowledge-graph-based representation for chart images," IEEE Transactions on Visualization and Computer Graphics, vol. 31, no. 9, pp. 5854-5868, 2024, doi: 10.1109/TVCG.2024.3476508.
[20] Y. Zhang, Y. Wang, Z. Guo, et al., "A Knowledge Graph-Guided and Multimodal Data Fusion-Driven Rapid Modeling Method for Digital Twin Scenes: A Case Study of Bridge Tower Construction," ISPRS International Journal of Geo-Information, vol. 15, no. 1, pp. 27-46, 2026, doi: 10.3390/ijgi15010027.
[21] S. Liu, "Research on the Conservation and Display of Mural Paintings in Tomb Chambers Based on Digital Twin Technology," J. COMBIN. MATH. COMBIN. COMPUT, vol. 127, no. 1, pp. 2429-2442, 2025, doi: 10.61091/jcmcc127a-138.
[22] J. Wang, F. Mo, S. Qiao, et al., "Spatial computing in digital twins," Digital Twin, vol. 2, no. 2, pp. 2508268-2508296, 2025, doi: 10.1080/27525783.2025.2508268.
[23] W. Luther, N. Baloian, D. Biella, et al., "Digital twins and enabling technologies in museums and cultural heritage: An overview," Sensors, vol. 23, no. 3, pp. 1583-1610, 2023, doi: 10.3390/s23031583.
[24] S. Kartal, Emre, and Gülder, "The Importance of Graphic Documentation on the Conservation of Mural Paintings: Examinations on Neet Günal's Mural Painting," Journal of Academic Social Science Studies, Art. no. 17(101), 2024, doi: 10.29228/JASSS.77804.
[25] N. Wang, J. Dong, H. Fang, et al., "3D reconstruction and segmentation system for pavement potholes based on improved structure-from-motion (SFM) and deep learning," Construction and Building Materials, vol. 398, no. 1, pp. 132499-132521, 2023, doi: 10.1016/j.conbuildmat.2023.132499.
[26] P. Wang, "Animation generation of traditional ethnic elements based on memory-enhanced self-supervised networks," International Journal of Information and Communication Technology, vol. 27, no. 7, pp. 1-20, 2026, doi: 10.1504/ IJICT.2026.151562.
[27] T. Luo, X. Sun, W. Zhao, et al., "Ethnic architectural heritage identification using low-altitude UAV remote sensing and improved deep learning algorithms," Buildings, vol. 15, no. 1, pp. 15-31, 2024, doi: 10.3390/buildings15010015.
[28] V. Ivanova Y U, "Dunhuang in the Digital Age: The Mediatization of Sacred Heritage and Chinas Cultural Diplomacy," Culture and Art, vol. (10), pp. 66-78, 2025, doi: 10.7256/2454-0625.2025.10.76334.
[29] K. Wen, "Literature review of digital protection of material cultural heritage," Environment, Social and Governance, vol. 1, no. 2, pp. 9-15, 2024, doi: 10.70267/08pazx48.
[30] J. Cao, X. Wang, F. Wang, et al., "Two-stage mural image restoration using an edge-constrained attention mechanism," PLoS One, vol. 19, no. 9, Art. no. e0307811-e0307831, 2024, doi: 10.1371/journal.pone.0307811.
[31] P. Zhang and C. Wang, "Research on the Characteristics of English Text Based on Computer and Its Translation," Journal of Physics: Conference Series, vol. 1992, no. 3, Art. no. 032004-, 2021, doi: 10.1088/1742-6596/1992/3/032006.
[32] V. Saba, D. Borggaard, C. Caracappa J, et al., "NOAA fisheries research geared towards climate-ready living marine resource management in the northeast United States," PLoS Climate, vol. 2, no. 12, Art. no. e0000323-e0000364, 2023, doi: 10.1371/ journal.pclm.0000323.
[33] E. Bayler, S. Chang P, L. De La Cour J, et al., "Satellite oceanography in NOAA: Research, development, applications, and services enabling societal benefits from operational and experimental missions," Remote Sensing, vol. 16, no. 14, pp. 2656-2690, 2024, doi: 10.3390/rs16142656.
[34] X. Zhang, S. Zhang, and J. Yan, "Pcp-mae: Learning to predict centers for point masked autoencoders," Advances in Neural Information Processing Systems, vol. 37, no. 1, pp. 80303-80327, 2024, doi: 10.52202/079017-2553.
[35] Y. Xiao, Y. Dou, and S. Yang, "Pointblip: Zero-training point cloud classification network based on blip-2 model," Remote Sensing, vol. 16, no. 13, pp. 2453-2470, 2024, doi: 10.3390/rs16132453.
[36] F. Sabatini, F. Albertin, B. Doherty, et al., "Unveiling street art: A multimodal and multitechnique approach for analyzing and mapping painting materials on large murals," Proceedings of the National Academy of Sciences, vol. 122, no. 35, Art. no. e2504918122-e2504918141, 2025, doi: 10.1073/pnas.2504918122.
[37] M. Yue, H. Liu, X. Chang, et al., "TGN: A Temporal Graph Network for Physics Prediction," Applied Sciences, vol. 14, no. 2, pp. 863-879, 2024, doi: 10.3390/app14020863.