ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on Assisted analysis of ceramic Process Evolution and Cultural Relic Restoration based on multimodal learning

Fan Xu[1,*] *and Jiannan Zhang*[2]

[1] Xi'an Siyuan University, Xi'an 710038, Shaanxi, China
[2] Shaanxi Yonghua Ceramic Art Culture Co., Ltd., Xi'an 710000, Shaanxi, China

SUMMARY: *Aiming at the problem that ceramic process evolution recognition relies on empirical interpretation and lacks multi-source collaboration in assisted analysis of cultural relic restoration, a multi-modal data system that fuses visible light images, microscopic images, spectral detection, three-dimensional morphology and text description is constructed, and an integrated model consisting of feature coding, cross-modal fusion, process evolution analysis and assisted restoration recommendation is proposed. In this study, convolutional neural network, Transformer, spectral line coding network and pre-trained language model are used to realize heterogeneous information unified representation, and multi-task learning is used to complete generation recognition, kiln mouth classification, decoration style discrimination and disease analysis. The experiment was divided into training set, validation set and test set according to 70%, 15% and 15%. The complete model achieved 93.8%, 90.6% and 89.8% in process evolution recognition, disease detection and repair assistant recommendation tasks, respectively, which were significantly better than 78.6%, 71.4% and 69.8% of traditional feature methods. The results show that multimodal learning can effectively improve the ability of process knowledge extraction, damage diagnosis and repair assistant decision-making of ceramic cultural relics.*

KEYWORDS: *Multimodal learning; Ceramic process evolution; Aided analysis of cultural relic restoration; Cross-modal fusion*

1 Introduction

Ceramic cultural relics not only record the evolution track of raw material ratio, forming method, glaze application process, firing system and decorative style in different periods, but also are important physical carriers for carrying out process history research, kiln entrance identification, age determination and protection and restoration. For a long time, ceramic research has mainly relied on instrument observation, decoration comparison and empirical interpretation. This path has strong applicability in the analysis of small samples, complete samples and typical samples. However, when faced with damaged samples, fragmented samples, cross-regional style mixed samples, and complex scenes containing images, spectra, microscopic, three-dimensional and text, the research on ceramics has been greatly improved. There are often problems such as strong subjectivity, limited processing efficiency and difficult collaborative utilization of multi-source information. Di Angelo et al. (2022) systematically reviewed the classification and reconstruction method of archaeological pottery based on 3D high-density scanning [1]. Cardarelli (2022) studied the unsupervised feature

*18092436188@163.com
https://doi.org/10.65102/is2026796
Published: 30 April, 2026

extraction method of pottery contour based on deep variational convolutional autoencoder [2]. Jin et al. (2023) proposed a non-destructive identification method for acoustic marks of ancient ceramics [3], and Towarek et al. (2024) studied the application progress of machine learning in the analytical chemistry of cultural heritage [4]. These results show that with the development of computer vision, pattern recognition, deep learning and data fusion technology, the research of ceramic cultural relics is gradually shifting from experience-driven to data-driven, which provides a new technical basis for process evolution analysis and restoration assistant decision-making.

From the research status at home and abroad, the existing research mainly focuses on two directions: ceramic process evolution recognition, cultural relic disease detection and digital restoration assistance. In terms of process evolution analysis, Pang et al. (2024) studied the pottery evolution pattern discovery method based on deep learning [5], Ao et al. (2024) respectively studied the morphological characteristics of export cups in Ming and Qing Dynasties and their social and cultural correlation from the perspectives of quantitative typological and influencing factors of morphological evolution [10, 11]. Chen et al. (2025) studied the systematic morphological analysis and style interpretation of Longquan bottle objects in Song Dynasty [19], and Yang et al. (2025) proposed an enhanced ResNet50 based classification model for Ming and Qing Dynasty ceramic images [20]. In terms of intelligent recognition and restoration assistance, Wang et al. (2024) studied the ceramic origin classification method based on the combination of microscopic images and integrated deep learning [6], Ling et al. (2024) reviewed the machine learning path in archaeological ceramic recognition and pointed out the advantages of deep learning in automatic classification [9]. Jin et al. (2024) studied the problem of accurate acoustic classification of visually approximate monochromatic porcelain pieces [12], Liu et al. (2024) proposed a prediction model of color matching of repair materials [7], Zheng et al. (2024) studied the restoration method of ancient ceramics based on image texture stitching [8]. Stoean et al. (2024) proposed a 3D restoration framework for degraded objects for museum digital display [13]. Hu et al. (2025) studied ceramic classification and value prediction based on the integration of deep learning and machine learning [14]. Cardarelli (2025) proposed the PyPotteryLens framework for automatic digitization of archaeological pottery drawings [16] and the PyPotteryInk diffusion model for automatic transfer of line drawings [15] respectively. Deng et al. (2025) studied the local fading feature extraction method of ancient ceramic decoration [17]. Liao et al. (2025) proposed a deep adversarial and inverse diffusion prediction method for the missing area of ancient ceramics [18]. In general, the existing results have made significant progress in shape recognition, material judgment, defect detection, color matching and digital reconstruction. However, most of the research still focuses on single-modal images, local tasks or static classification problems, and the collaborative representation of multi-source information such as instrument shape, ornament, glazing, microstructure, disease area and literature description is still insufficient. The ability to dynamically model process evolution is also relatively limited, and the comparative summary of related studies is shown in Table 1.

Table 1: Comparative analysis of related studies

Research Direction	Representative References	Main Content	Main Limitations
Digital Classification and Morphological Analysis of Ceramics	[1][2][5][10][11][19][20]	Three-dimensional scanning, deep learning, and quantitative typology are used for vessel-shape recognition, chronological classification, and evolutionary analysis.	Most studies focus on morphology or image information, while multimodal collaborative modeling remains insufficient.
Ceramic Detection and Intelligent Identification	[3][4][6][9][12][14]	Acoustic methods, microscopic imaging, analytical chemistry, and machine learning are combined for provenance identification, attribute discrimination, and value analysis.	The relationships among different modalities are relatively weak, and the interpretability of technological evolution is limited.
Auxiliary Analysis for Cultural Relic Restoration	[7][8][13][17][18]	Research covers color prediction, texture stitching, missing-region prediction, and three-dimensional restoration.	Existing studies mostly concentrate on local restoration tasks and lack an integrated auxiliary decision-making framework.
Digital Tools and Automatic Generation for Ceramics	[15][16]	Diffusion models and automatic digitalization frameworks are used to improve the efficiency of ceramic document processing.	Greater emphasis is placed on process automation, with insufficient integration with technological evolution analysis.

Based on the above analysis, this paper focuses on the construction of multimodal ceramic relic data, the design of process evolution analysis model, the construction of restoration auxiliary analysis method and experimental verification, and plans to construct a multimodal data system that fuses visible light image, microscopic texture, detection spectrum, three-dimensional shape and text description. The collaborative modeling of device structure, decoration style, material properties and disease characteristics is realized in a unified representation space. Furthermore, a cross-modal feature alignment and deep fusion mechanism is designed to improve the ability of generation identification, kiln mouth discrimination, process path analysis and damage understanding. On this basis, the mechanisms of similar objects matching, missing area prediction and restoration proposal generation are introduced to form an intelligent computing framework for auxiliary analysis of cultural relic restoration. The innovation of this paper is mainly reflected in three aspects.

Firstly, an integrated multi-modal data representation for ceramic process evolution and repair assistance is constructed. The second is a cross-modal deep fusion model that takes into account both process evolution identification and disease understanding. The third is to couple digital restoration prediction and auxiliary decision analysis into the same technical path, so as to enhance the computability, reusability and engineering application value of ceramic cultural relics research.

2 Data construction and feature representation of multi-modal ceramic relics

2.1 Key dimensions of ceramic process evolution and restoration aided analysis

The ceramic process evolution analysis and the auxiliary analysis of cultural relic restoration correspond to the joint representation problem of multi-source heterogeneous information, whose computational goal is to establish a unified mapping relationship between the shape, decorative style, glaze state, embryo composition, disease characteristics and historical semantics. A single image modality can only capture local appearance differences, and it is difficult to depict deep information such as kiln mouth process differences, era style transfer, and damage mechanism coupling. Therefore, it is necessary to construct a multi-modal feature space for cover type, grain, glaze color, fetal body composition, surface disease and history description. Let the original observation sample of the ith ceramic cultural relic be represented as follows.

$$X_i=\{x_i^{(s)},x_i^{(p)},x_i^{(g)},x_i^{(c)},x_i^{(d)},x_i^{(t)}\} \tag{1}$$

Here,$x_i^{(s)}$ represents the structural mode of the instrument, including the curvature along the mouth, the change rate of the abdominal diameter, the thickness of the instrument wall, the ratio of the bottom and foot, the parameters of the contour arc and 3D geometric descriptors. $x_i^{(p)}$ represents the pattern mode, including pattern unit, boundary shape, repetition cycle, topology layout and local stroke texture.$x_i^{(g)}$ represents the glaze mode, which describes the hue distribution, lightness gradient, saturation change, surface reflection intensity and micro texture. $x_i^{(c)}$ represents the composition mode of the fetal body, which is usually composed of XRF, XRD, Raman or hyperspectral detection results, and is used to characterize the element combination, mineral phase structure and firing system related characteristics. $x_i^{(d)}$ represents the surface disease mode, covering damage areas such as cracks, glaze stripping, defects, weathering, qin staining and attachment pollution; $x_i^{(t)}$ denotes the historical description modality, which contains textual semantic information such as archaeological records, expiation horizons, era labels, restoration archives, and expert annotations.

The structure of the instrument carries the information of forming process and proportion specification, which is suitable for identifying the difference of the era and the style of the kiln system. The pattern mode records the evolution law of decorative language, which can reflect the pattern organization way, aesthetic preference and process propagation path. Glaze color characteristics are related to glaze application system, firing atmosphere and temperature control, which are the important basis for distinguishing the same shape and heterogeneous objects. The matrix composition directly corresponds to the raw material system and mineral transformation process, which can provide material discrimination

information besides appearance characteristics. Surface disease modes describe the spatial distribution of crack propagation, glaze loss and local weathering, which have a direct constraint effect on disease diagnosis, damage assessment and repair path selection. The historical description modality provides semantic priors, chronological context, and repair records, which can compensate for insufficient visual evidence in the condition of damaged samples and incomplete samples. The above six dimensions are not independent of each other, but together constitute a coupling system of "process attribute-material attribute-damage attribute-knowledge attribute". The framework shown in Figure 1 takes six types of modalities as parallel inputs, which are converged to the cross-modal fusion layer after single-modal coding, and finally outputs the process evolution identification results and the repair auxiliary analysis results.

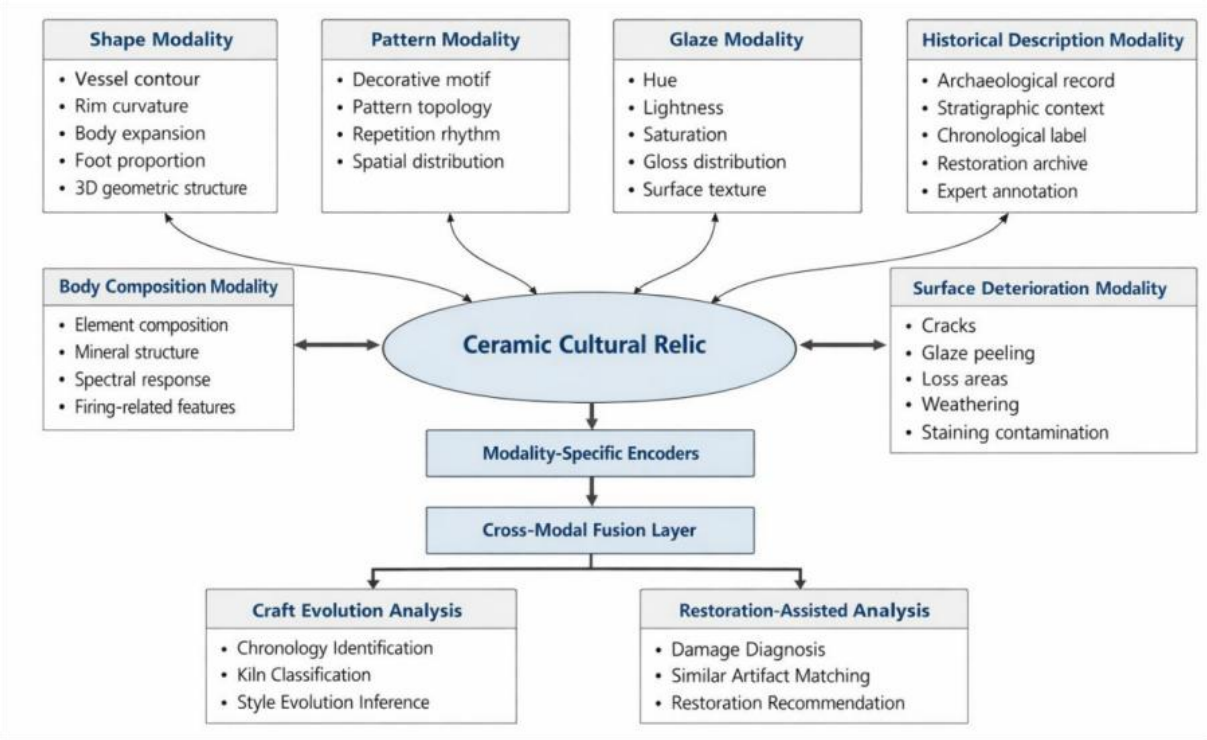

Figure 1: Key dimensions and feature representation framework for multimodal ceramic relics

In order to achieve unified representation, it is necessary to map each modality into a shared feature space. Let $\phi m(\cdot)$ be the encoding function of the MTH modality, then the fusion feature representation of the ith cultural relic is defined as follows.

$$z_i = \sum_{m=1}^{6} \alpha_i^{(m)} \phi_m\left(x_i^{(m)}\right), \quad \sum_{m=1}^{6} \alpha_i^{(m)} = 1 \tag{2}$$

Here, $\alpha_i^{(m)}$ represents the weight coefficient of the MTH mode, which is used to measure the contribution of the mode in the current task. This equation corresponds to weighted fusion rather than simple splicing, and can automatically highlight key dimensions according to different tasks such as generation identification, kiln mouth classification, disease segmentation or repair recommendation. In order to improve the adaptive ability of mode selection, the weights can be learned through the attention mechanism:

$$\alpha_i^{(m)}=\frac{\exp\left(w_m^{\top}\phi_m(x_i^{(m)})\right)}{\sum_{k=1}^{6}\exp\left(w_k^{\top}\phi_k(x_i^{(k)})\right)} \tag{3}$$

Here, w_m is the learnable parameter vector. This mechanism can dynamically adjust the modal contribution according to the morphological differences, material differences and damage differences within the sample, and weaken the interference of single modal noise on the final discrimination result.

In the implementation of feature coding, the instrument-like structural modes can be modeled by contour parameterization, point cloud descriptor or graph structure geometric coding. The decorative modality is suitable for using convolutional neural network or Vision Transformer to extract multi-scale pattern semantics. The glaze mode can combine color moments, gray level co-occurrence matrix, local binary pattern and micro-texture statistics to form a composite representation. The peak position combination relationship and element coupling mode can be learned by one-dimensional convolutional network or spectral line Transformer. Surface disease modalities usually rely on semantic segmentation networks to extract the spatial distribution characteristics of cracks, glazing and defect areas. The historical description modality can be used to complete text embedding with the help of BERT-like pre-trained language models. The key dimension system thus formed can not only serve the ceramic process evolution identification, but also support the auxiliary analysis of cultural relic restoration, and meet the integration path of multimodal learning and computer technology required by the topic.

2.2 Multimodal Data Acquisition and preprocessing

In order to support the joint modeling of ceramic process evolution recognition and auxiliary analysis of cultural relic restoration, a multimodal data system is constructed, which is composed of visible light images, microscopic images, spectral detection data, three-dimensional morphological information and text description information. Let the original sample of the ith ceramic cultural relic be denoted as follows.

$$S_i=\{I_i^{v},I_i^{m},Q_i,P_i,T_i,y_i\} \tag{4}$$

Among them, I_i^{v} represents the visible light image, which is used to describe the contour of the instrument, the glaze color distribution and the macroscopic decoration. I_i^{m} represents the microscopic image, which is used to characterize the glaze particle structure, microcracks and local abrasion texture. Q_i represents the spectral detection sequence, including XRF, hyperspectral, or Raman measurements; Pi represents the 3D point cloud or mesh model, which is used to record the shape parameters, surface changes and damage boundaries of the object. Ti represents textual information such as archaeological records, restoration archives, era labels, and expert descriptions. y_i is the supervision label, covering the task objectives such as generation category, kiln mouth type, disease type and repair status. The multi-modal acquisition process is shown in Figure 2, where the original data enters the preprocessing stage after unified numbering, instance registration and modal alignment to ensure that the data from different sources establish a one-to-one correspondence at the level of the same cultural relic entity.

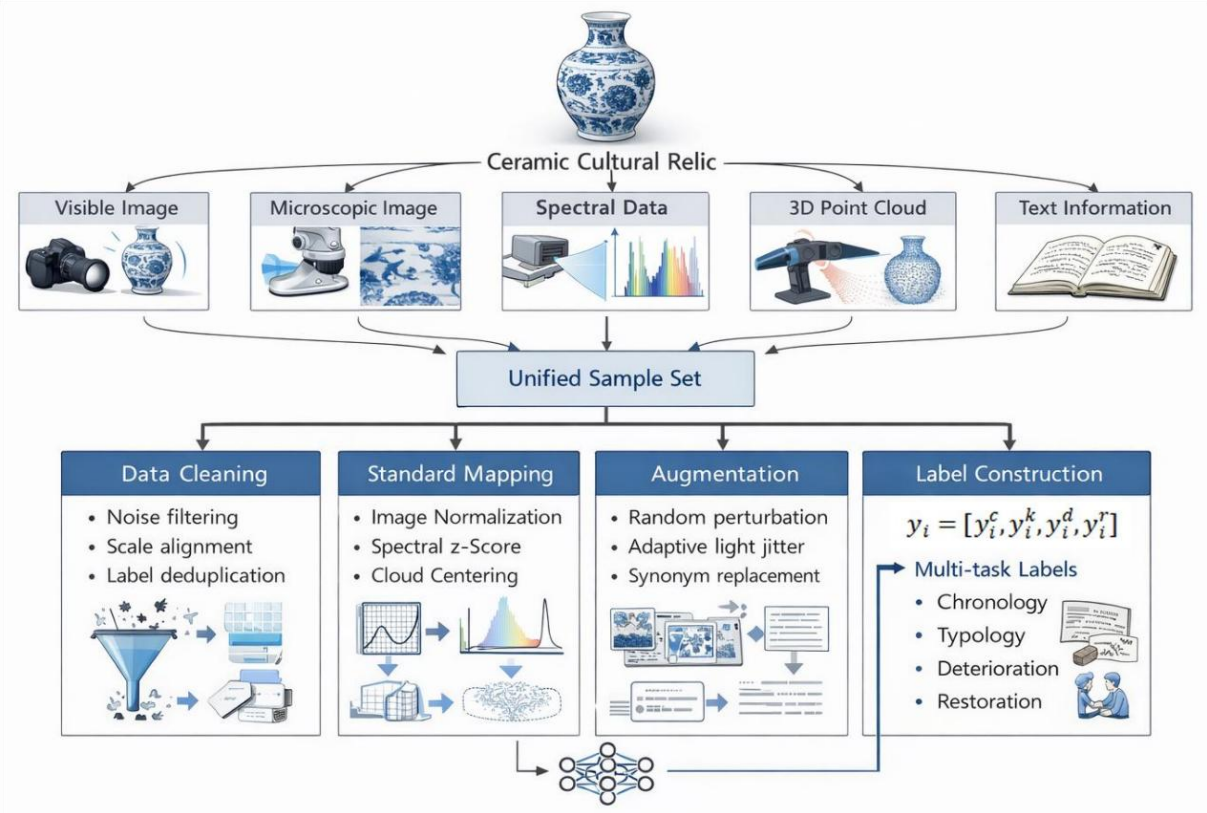

Figure 2: Multimodal ceramic relic data acquisition and preprocessing process

Data cleaning deals with noise, missing, scale inconsistency and annotation conflict. Background separation, distortion correction and noise suppression operations were used to remove the interference of shooting environment. Savitzky-golay smoothing and abnormal peak suppression were used to improve the signal-to-noise ratio of the spectral sequence. The geometric noise of 3D point cloud is controlled by outlier elimination, hole repair and resampling. For text data, terminology was standardized, duplicate records were merged, and invalid fields were removed. The sample set after cleaning can be expressed as follows.

$$D=\{(\hat{I}_i^v,\hat{I}_i^m,\hat{Q}_i,\hat{P}_i,\hat{T}_i,\hat{y}_i)\}_{i=1}^{N} \tag{5}$$

Among them, the data with "^" symbol represents the valid sample after cleaning. In order to weaken the influence of different acquisition devices and dimensional differences, this paper implements standardized mapping for each modality. Image pixels are linearly normalized:

$$\tilde{I}=\frac{I-I_{min}}{I_{max}-I_{min}} \tag{6}$$

Spectral sequences are normalized by z-score:

$$\tilde{Q}_{ij}=\frac{Q_{ij}-\mu_j}{\sigma_j} \tag{7}$$

Here, μ_j and σ_j represent the mean and standard deviation of the JTH band or element channel, respectively. Centralization and scale normalization are used for 3D point cloud:

$$\tilde{P}_i=\frac{P_i\text{-mean}(P_i)}{\max\limits_{p\in P_i}\|p\text{-mean}(P_i)\|_2} \tag{8}$$

The processing can map objects of different sizes into a unified geometric space, avoiding the deviation of morphological coding caused by scale differences.

In order to enhance the robustness of the model to damaged samples, illumination perturbation and local abrasion, this paper introduces a modal enhancement strategy in the training phase. Rotation, cropping, brightness perturbation, color jitter and partial occlusion were performed on visible and microscopic images. The spectral data were subjected to slight Gaussian noise and bands were randomly discarded. The point cloud data performs random rotation, jitter and density perturbation. Synonymous term substitution and fragment truncation were used for textual data to model record differences. The augmented samples can be expressed as follows.

$$A(S_i)=\{A_v(I_i^v),A_m(I_i^m),A_q(Q_i),A_p(P_i),A_t(T_i)\} \tag{9}$$

Here, A^* denotes the mode-specific enhancement operator. Label construction adopts a multi-task encoding method to map the generation category, kiln type, disease level and repair status into a joint label vector:

$$y_i=[y_i^c,y_i^k,y_i^d,y_i^r] \tag{10}$$

Here, y_i^c represents the era category label, y_i^k represents the kiln mouth label, y_i^d represents the disease label, and y_i^r represents the repair status label. The proposed structure can provide a unified supervision signal for subsequent cross-modal feature fusion, multi-task learning and repair assistance recommendation.

2.3 Multi-modal Feature Encoding and Unified Representation

After multimodal ceramic relic data enters the analysis model, the feature encoding and shared semantic alignment of heterogeneous modalities need to be completed first. Due to the significant differences in data structure, statistical distribution and semantic granularity among visible light images, microscopic images, spectral sequences, 3D morphology and text descriptions, we adopt a hierarchical representation strategy of "modal-specific encoder-linear projection mapping-shared space fusion". The core goal is to simultaneously preserve the effective discriminant information in the device structure, grain texture, material composition, damage state and historical semantics, and compress them into a unified low-dimensional representation space for subsequent process evolution recognition and repair auxiliary analysis.

The visible image and microscopic image modalities are modeled by convolutional neural network for local texture. Let the input images be Iiv and Iim respectively, then the visual features can be expressed as follows.

$$f_i^v=CNN_v(I_i^v),\quad f_i^m=CNN_m(I_i^m) \tag{11}$$

Among them, $CNN_v(\cdot)$ focuses on extracting the contour, glaze color distribution and macroscopic decorative features, and $CNN_m(\cdot)$ focuses on describing the microscopic particle structure, microcrack boundary and local abrasion texture. For 3D morphological information, this paper adopts the Transformer structure to model the global geometric relationship. After

dividing the point cloud or mesh fragment into a number of morphological units, a sequence of geometric tokens $\{p_{i1}, p_{i2}, \ldots, p_{in}\}$, and enter the morphological encoder:

$$f_i^{3d} = \mathrm{Trans}_{3d}(P_i) \tag{12}$$

The representation can preserve the geometric features of the object contour curvature, abdominal distension, mouth turning and damage boundary. The spectral detection data belongs to a one-dimensional continuous sequence, and the spectral line coding network is used to extract the peak position, peak width and band coupling relationship. Let the spectral input be Qi, then:

$$f_i^q = \mathrm{SpecEnc}(Q_i) = \mathrm{Trans}_q(\mathrm{Conv1D}(Q_i)) \tag{13}$$

Among them, Conv1D(·) is responsible for extracting local peak patterns, and $\mathrm{Trans}_q(\cdot)$ models long-distance band dependence, thereby enhancing the expressive power of element assemblages and mineral phase changes. In the text description mode, the pre-trained language model is used to complete the semantic coding, and the text of the marker is recorded as Ti, then:

$$f_i^t = \mathrm{PLM}(T_i) \tag{14}$$

Among them, PLM(·) can be implemented by BERT-like models to extract semantic priors in soil horizons, age labels, restoration records, and expert annotations. The coding methods and functional division of each modality are shown in Table 2.

Table 2: Multimodal feature encoding and unified representation design

Modality Type	Input Data	Encoding Model	Output Feature	Main Representational Content
Visible-Light Image	Overall artifact image	CNN	f_i^v	Vessel contour, glaze color distribution, and macroscopic decorative patterns
Microscopic Image	Local microscopic regions	CNN	f_i^m	Particle structure, micro-cracks, and abrasion textures
Spectral Data	XRF / hyperspectral / Raman sequences	1D-CNN + Transformer	f_i^q	Peak-position relationships, elemental combinations, and mineral phase variations
3D Morphology	Point clouds / mesh fragments	Transformer	f_i^{3d}	Curvature structure, proportional parameters, and damaged-boundary features
Text Description	Archaeological records and restoration archives	Pre-trained language model	f_i^t	Chronological semantics, excavation information, and restoration priors
Unified Representation	Projected features from all modalities	Attention fusion + contrastive constraints	z_i	Shared semantic representation of process attributes, material attributes, and damage attributes

The original feature dimensions of different modalities are not consistent, which need to

be mapped to a unified representation space through a projection layer. Let the encoded feature of the MTH modality be fi(m), then the shared space representation is defined as follows.

$$h_i^{(m)}=W_m f_i^{(m)}+b_m \tag{15}$$

Here, W_m and bm are the projection matrix and bias term, respectively, and all modes are mapped to satisfy $h_i^{(m)} \in R^d$. On this basis, the gated fusion mechanism is used to generate the unified representation vector of ceramic cultural relics:

$$z_i=\sum_{m=1}^{M} \alpha_i^{(m)} h_i^{(m)}, \quad \sum_{m=1}^{M} \alpha_i^{(m)} =1 \tag{16}$$

where $\alpha_i^{(m)}$ is the modal weight, which is adaptively learned by the attention network:

$$\alpha_i^{(m)}=\frac{\exp\left(u^{\top} \tanh\left(W_a h_i^{(m)}\right)\right)}{\sum_{k=1}^{M} \exp\left(u^{\top} \tanh\left(W_a h_i^{(k)}\right)\right)} \tag{17}$$

The proposed mechanism can dynamically adjust the contribution of visual, spectral, geometric and textual information according to the sample characteristics, and weaken the interference of missing and noise modes on the final discrimination results.

To enhance cross-modal consistency, contrastive constraints are further introduced in the unified representation stage. Assuming that different modes of the same cultural relic constitute positive sample pairs and different cultural relics constitute negative sample pairs, the contrastive loss is written as follows.

$$L_{con}=-\log \frac{\exp\left(\operatorname{sim}(h_i^a, h_i^b)/\tau\right)}{\sum_{j=1}^{N} \exp\left(\operatorname{sim}(h_i^a, h_j^b)/\tau\right)} \tag{18}$$

where sim(,) is the cosine similarity and τ is the temperature coefficient. Through shared space projection, attention fusion and contrast constraints, the model can establish a unified multimodal representation space of ceramic cultural relics, which can provide consistent feature input for subsequent process evolution analysis, disease recognition and restoration assistant recommendation.

3 Ceramic process evolution analysis model fusing multi-modal learning

3.1 The overall framework design of the model

In order to realize the multi-dimensional collaborative analysis of ceramic process evolution information, this paper constructs an overall computing framework of "multi-modal input-feature encoder-cross-modal fusion-evolution analysis output". The goal is to simultaneously model the correlation relationship between the structure of the tool, the decorative texture, the glaze state, the material composition, the spatial shape and the historical semantics in a unified network. It is mapped to the discriminative output for epoch identification, kiln mouth classification and style evolution inference. The overall framework

is shown in Figure 3.

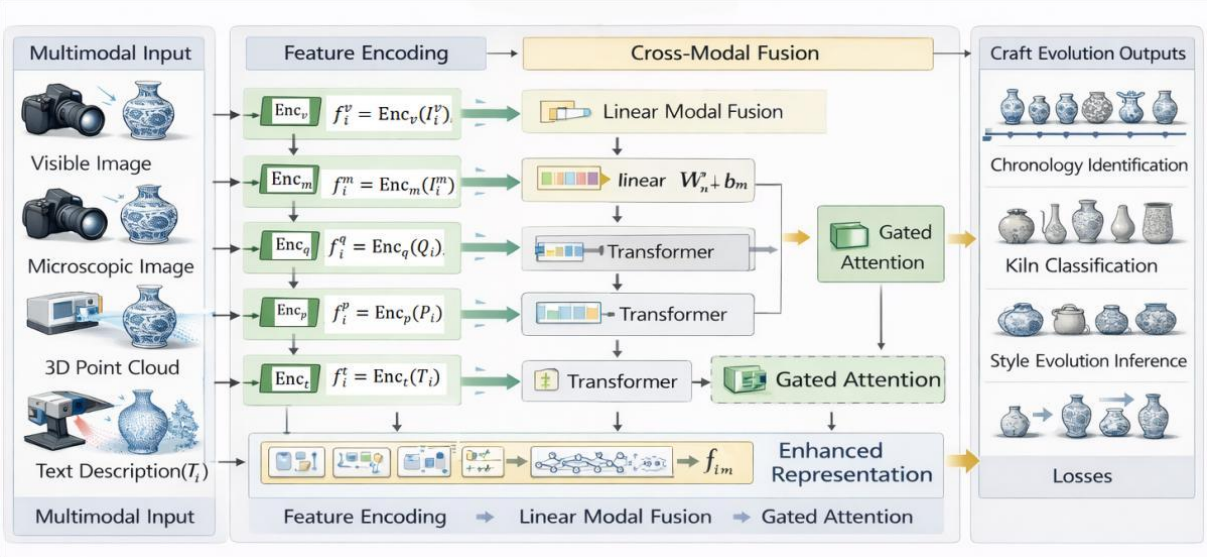

Figure 3: General framework of ceramic process evolution analysis model with multimodal learning

The input layer receives the multimodal sample of the ith ceramic artifact:

$$X_i=\{I_i^v,I_i^m,Q_i,P_i,T_i\} \tag{19}$$

Here, I_i^v is the visible light image, I_i^m is the microscopic image, Q_i is the spectral detection sequence, P_i is the 3D point cloud or grid representation, and T_i is the text description. Different modes enter the corresponding coding branch to complete feature extraction, and obtain visual features, microscopic texture features, spectral line component features, geometric features and semantic text features:

$$f_i^v=Enc_v(I_i^v),\quad f_i^m=Enc_m(I_i^m),\quad f_i^q=Enc_q(Q_i),\quad f_i^p=Enc_p(P_i),\quad f_i^t=Enc_t(T_i) \tag{20}$$

Here, $Enc_v(\cdot)$ and $Enc_m(\cdot)$ use convolutional neural networks to extract local texture and macroscopic structure information, $Enc_q(\cdot)$ uses one-dimensional convolution and Transformer coupling structure to model band peak dependence, $Enc_p(\cdot)$ uses geometric Transformer to extract 3D contour and curvature relationship. $Enc_t(\cdot)$ employs a pre-trained language model to generate semantic text embeddings. Since the output dimension of each modality is inconsistent, the model sets up a unified mapping module after the encoding layer to project the heterogeneous features into a shared latent space:

$$h_i^{(m)}=W_m f_i^{(m)}+b_m,\quad m\in\{v,m,q,p,t\} \tag{21}$$

Here, W_m and bm denote the linear projection parameters of the MTH mode, respectively. This mapping aligns image, spectral, geometric and textual information on the same feature dimension d and provides a basis representation for cross-modal interaction.

The cross-modal fusion layer assumes the coupling modeling function of process attributes. Its core is not simple splicing, but learning the contribution of different modes on the current sample through an attention-driven dynamic weighting mechanism. Let the set of modal features in the unified space be $\{h_i^v,h_i^m,h_i^q,h_i^p,h_i^t\}$, then the fusion representation is

defined as follows.

$$z_i=\sum_{m=1}^{5}\alpha_i^{(m)}h_i^{(m)},\quad \sum_{m=1}^{5}\alpha_i^{(m)}=1 \tag{22}$$

where $\alpha_i^{(m)}$ is the modal weight, which is adaptively calculated by the attention network:

$$\alpha_i^{(m)}=\frac{\exp\left(u^{\top}\tanh\left(W_a h_i^{(m)}\right)\right)}{\sum_{k=1}^{5}\exp\left(u^{\top}\tanh\left(W_a h_i^{(k)}\right)\right)} \tag{23}$$

where, Wa and u are learnable parameters. This mechanism can automatically adjust the modal contribution according to the process differences and information completeness within the sample. When the visual decoration is clear and the text information is missing, the model will increase the weight of the image branch.When the damage is severe but the unearthed records are complete, the discriminative effect of text and spectral branches will be enhanced. In order to further characterize the high-order association between modalities, the cross-modal interaction unit is connected after the fusion layer, and the multi-head self-attention is used to establish the coupling relationship of"instrument model - ornament - glaze - component - semantics", and the output enhanced representation $\tilde{z}_i$.

The output layer of evolution analysis adopts a multi-task learning structure, which synchronously completes the generation recognition, kiln mouth classification and style evolution inference on the basis of unified representation. The output can be written as follows:

$$\hat{y}_i^{c}=\mathrm{Softmax}(W_c\tilde{z}_i+b_c),\quad \hat{y}_i^{k}=\mathrm{Softmax}(W_k\tilde{z}_i+b_k),\quad \hat{y}_i^{s}=W_s\tilde{z}_i+b_s \tag{24}$$

Here, $\hat{y}_i^{c}$ represents the era category prediction result, $\hat{y}_i^{k}$ represents the kiln mouth category prediction result, and $\hat{y}_i^{s}$ represents the continuous style evolution score. The overall loss function is composed of classification loss, regression loss and cross-modal alignment loss.

$$L=\lambda_1 L_{cls}^{c}+\lambda_2 L_{cls}^{k}+\lambda_3 L_{reg}^{s}+\lambda_4 L_{align} \tag{25}$$

Among them, cross-entropy loss is used for Lclsc and Lclsk, mean square error loss is used for Lregs, and Lalign is used to constrain the semantic consistency of different modalities of the same cultural relic in the shared space. The overall framework thus formed connects multimodal input, exclusive coding, shared mapping, cross-modal fusion and multi-task output into an integrated computing link, which can provide a deep learning model with complete structure, information coupling and scalability for ceramic process evolution analysis.

3.2 Cross-modal feature fusion mechanism

The key of multi-modal ceramic process evolution analysis is not the feature extraction accuracy of a single branch, but the structural alignment and semantic collaboration between heterogeneous modalities. The image modality mainly carries the visual differences of the contour of the vessel, the organization of the decoration, and the glaze. The spectral line modality corresponds to the information of the element composition and the firing system.

The three-dimensional morphological modality describes the curvature distribution, proportion relationship, and damage boundary. If only a simple splicing method is used for fusion, although the original information of each modality can be retained, it is difficult to explicitly model the high-order association between "ornament-material", "instrument model-kiln mouth" and "disease-repair record". To this end, we design a deep fusion mechanism consisting of feature projection, attention weighting, bilinear interaction and cross-modal Transformer to complete multimodal information collaboration in a unified semantic space.

Let the image, spectral line, 3D shape and text features of the ith ceramic cultural relic obtained after coding by each branch be hiimg, hispec, higeo and hitxt, respectively. Since the statistical distribution and dimension scale of different modalities are not consistent, a linear projection layer is introduced to complete the shared space mapping before fusion:

$$\tilde{h}_i^{(m)}=W_m h_i^{(m)}+b_m, \quad m\in\{img,spec,geo,txt\} \tag{26}$$

Wm and bm are mode-specific projection parameters, which satisfy $\tilde{h}_i^{(m)}\in R^d$ after mapping. On this basis, the basic fusion vector is constructed in the form of feature concatenation:

$$u_i=[\tilde{h}_i^{img}\,\|\,\tilde{h}_i^{spec}\,\|\,\tilde{h}_i^{geo}\,\|\,\tilde{h}_i^{txt}] \tag{27}$$

This representation can completely preserve the first-order representation of each modality, but the concatenation itself does not contain interaction terms. Therefore, bilinear fusion is further introduced to characterize the second-order coupling relationship between pairs of modalities. Let the interaction terms of any two modes a and b be as follows.

$$\beta_i^{(a,b)}=(\tilde{h}_i^{(a)})^{\top}W_{ab}\tilde{h}_i^{(b)} \tag{28}$$

Here, W_{ab} is the bilinear weight matrix. The bilinear interactions of all modal combinations can be aggregated as follows.

$$b_i=\sum_{a<b}\beta_i^{(a,b)} \tag{29}$$

This term can explicitly learn the coupling strength between instrument type and composition, decoration and age semantics, three-dimensional shape and damage state, so as to enhance the ability of the model to distinguish the process evolution path.

To avoid the fixed modal contribution on different samples, an attention allocation mechanism is embedded in the fusion layer to dynamically model the importance of each modality. The modal weights are defined as follows.

$$\alpha_i^{(m)}=\frac{\exp\left(q^{\top}\tanh\left(W_a\tilde{h}_i^{(m)}\right)\right)}{\sum_k\exp\left(q^{\top}\tanh\left(W_a\tilde{h}_i^{(k)}\right)\right)} \tag{30}$$

Here, W_a and q are learnable parameters. The weighted modal representation is thus obtained as follows.

$$\hat{h}_i^{(m)}=\alpha_i^{(m)}\tilde{h}_i^{(m)} \tag{31}$$

This mechanism can automatically adjust the contribution ratio according to the completeness of the internal information of the sample and the task requirements. In the samples with clear grain and missing chronological records, the image and 3D morphological branches will obtain higher weights. In the samples with severe appearance wear and complete inspection records, the discrimination between spectral lines and text modes will be enhanced. Attention weighting solves the modal selection problem, and bilinear fusion solves the explicit interaction problem. However, both of them are still mainly at the level of vector magnitude representation, and it is difficult to deal with the finer-grained correspondence between cross-modal tokens. In order to further capture the deep dependencies between local grain area and keyword description, spectral line peak segment and kiln mouth label, 3D curvature segment and instrument type category, this paper introduces a cross-modal Transformer in the high-level fusion stage.

Suppose that four types of modes form a sequence after blocking or tokenization:

$$H_i=[X_i^{img};X_i^{spec};X_i^{geo};X_i^{txt}]\in R^{n\times d} \tag{32}$$

where n is the total number of tokens. Cross-modal Transformer computes the correlation between any tokens via multi-head self-attention:

$$Att(Q,K,V)=Softmax\left(\frac{QK^{\top}}{\sqrt{d}}\right)V \tag{33}$$

Here, Q=HiWQ, K=HiWK, and V=HiWV. The long head mechanism can be written as follows:

$$MHA(H_i)=Concat(head_1,\ldots,head_r)W_O \tag{34}$$

After multiple layers of cross-modal encoding, the enhanced representation HiL is obtained. Its global fusion vector is generated in the form of pooling:

$$c_i=Pool(H_i^L) \tag{35}$$

Finally, the concatenated features, bilinear interaction features and cross-modal Transformer output are combined to form a fusion representation:

$$z_i=W_u u_i+W_b b_i+W_c c_i \tag{36}$$

where W_u, W_b and W_c are learnable parameter matrices. The representation preserves first-order information, second-order interaction and global dependency structure simultaneously, and can be used as a unified input for epoch identification, kiln mouth classification and style evolution inference.

The above fusion mechanism is consistent with the ceramic process evolution analysis task. The image branch provides visual evidence of ornamentation morphology and glaze layer, the spectral line branch supplements material and firing information, the three-dimensional morphological branch describes the geometric evolution law of the instrument, and the text branch introduces archaeological knowledge constraints. Attention mechanism is used to allocate modal contributions, feature concatenation ensures the integrity

of basic information, bilinear fusion enhances modal coupling expression, and cross-modal Transformer is responsible for establishing fine-grained semantic correspondence. The resulting deep fusion path can realize the information collaboration among image, spectral line, 3D shape and text knowledge in a unified representation space, and provide stable and interpretable fusion features for the subsequent evolution analysis output.

3.3 Analysis method of ceramic process evolution

Ceramic process evolution analysis is not a single category discrimination problem, and the calculation object includes four related tasks: generation identification, kiln mouth classification, decoration style discrimination and process path inference. The four types of tasks share low-level evidence such as instrument type, ornamentation, glaze color, material and text semantics, but have differences in supervision granularity and output form. Generation identification and kiln mouth classification belong to discrete discriminant tasks, grain style discrimination has both category attribution and continuous change characteristics, and process path inference is corresponding to stage evolution relationship modeling. If the above tasks are separated, it is difficult for the model to take advantage of the coupling constraints between "time-kiln mouth -ornament-process system". If only the static classification network is used, it cannot describe the continuity and transfer law of process evolution. To this end, a joint analysis method of "shared representation layer -- multi-task output layer -- temporal relationship modeling layer" is constructed on the basis of unified fusion representation zi, and its structure is shown in Figure 4. Figure 4 takes the multimodal fusion features as input, and divides into three task branches: generation identification, kiln mouth classification and decoration style discrimination after the shared feature enhancement module. At the same time, the inter-period process relationship graph is constructed to carry out the propagation modeling of stage adjacency and style continuity between samples, and the final output process path inference results.

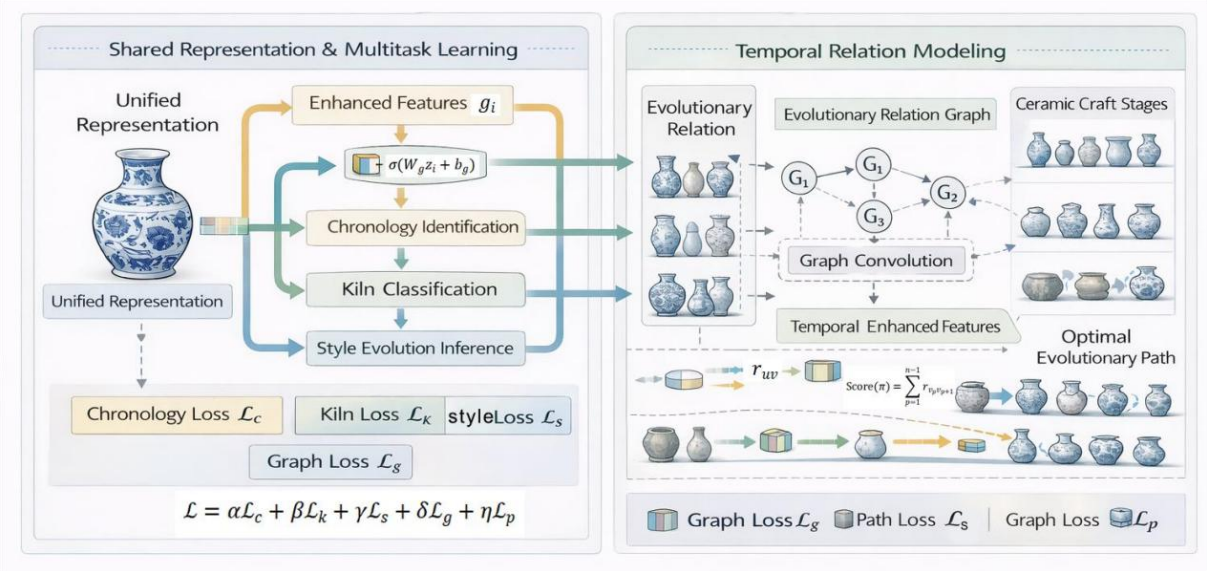

Figure 4: Ceramic process evolution analysis method integrating multi-task learning and temporal relationship modeling

Let the ith ceramic cultural relic be fused to obtain a unified representation zi∈Rd after being fused in Section 3.2. The shared representation layer first performs a nonlinear mapping on it:

$$g_i=\sigma(W_g z_i+b_g) \tag{37}$$

where W_g and bg are shared layer parameters and $\sigma(\cdot)$ is the activation function. Based on the shared feature gi, the generation identification, kiln mouth classification and decoration style discrimination are output by independent task heads:

$$\hat{y}_i^c=\text{Softmax}(W_c g_i+b_c),\quad \hat{y}_i^k=\text{Softmax}(W_k g_i+b_k),\quad \hat{y}_i^s=\text{Softmax}(W_s g_i+b_s) \tag{38}$$

Among them, $\hat{y}_i^c$ $\hat{y}_i^k$ and $\hat{y}_i^s$ represent the prediction results of age category, kiln mouth category and decorative style category, respectively. In this design, the common information of the instrument type and material level is reserved in the shared layer, and the task-specific information is transferred to the branch classifier to learn, which helps to alleviate the feature competition between different tasks.

In order to enhance the recognition ability of process continuity and phase transition law, this paper introduces the temporal relation graph to model the evolutionary adjacency relationship between samples. Suppose that all training samples constitute a graph G=(V,E), each node corresponds to a ceramic cultural relic, and the edge weight is jointly determined by the feature similarity, age proximity and kiln mouth correlation. The relationship weight between node i and node j is defined as follows.

$$a_{ij}=\lambda_1 \cos(g_i,g_j)+\lambda_2 \exp\left(-\frac{|t_i-t_j|}{\tau}\right)+\lambda_3\, 1(k_i=k_j) \tag{39}$$

where t_i and tj represent the time label or time interval center value of the sample, $1(\cdot)$ is the indicative function, and λ_1,λ_2, and λ_3 are the weight coefficients. This formula unifies the feature proximity of the decoration and the type level, the stage proximity of the time axis and the process inheritance relationship of the kiln system into the same side weight expression. Based on the relation matrix $A=[a_{ij}]$, graph convolution is used to propagate process evolution information:

$$g_i^{(l+1)}=\rho\left(\sum_{j\in N(i)} \tilde{a}_{ij}\, W^{(l)} g_j^{(l)}\right) \tag{40}$$

Here, $\tilde{a}_{ij}$ is the normalized edge weight, $W^{(l)}$ is the graph convolution parameter of the LTH layer, and $\rho(\cdot)$ is the nonlinear mapping. The enhanced representation gi∗ obtained after graph propagation preserves both the sample's own attributes and the inter-temporal neighborhood information, which can more stably represent the gradual relationship between process stages.

The process path inference is implemented using the transfer score function. Let the evolution transfer score from process stage u to stage v be as follows.

$$r_{uv}=\mu_1 \cos(\bar{g}_u^*,\bar{g}_v^*)+\mu_2\, \Delta s_{uv}+\mu_3\, \Delta m_{uv} \tag{41}$$

Among them, $\bar{g}_u^*$ and $\bar{g}_v^*$ represent the average representation of stage u and stage v, Δsuv represents the similarity change of decorative style, Δmuv represents the change range of material or glaze color characteristics, μ_1,μ_2,μ_3 are transfer parameters. Given a candidate process path $\pi=(v_1,v_2,\ldots,v_n)$, then its total path score is:

$$\text{Score}(\pi)=\sum_{p=1}^{n-1} r_{v_p v_{p+1}} \tag{42}$$

By finding the maximum score sequence in the candidate path set, the model can infer the process evolution direction and the key turning point. This strategy can incorporate instrument type variation, ornament variation and material evolution into the same sequence optimization framework, avoiding simplifying process evolution into disconnected static classification results.

Multi-task joint loss function is used for overall training:

$$L=\alpha L_c+\beta L_k+\gamma L_s+\delta L_g+\eta L_p \tag{43}$$

where L_c, L_k and L_s are the cross-entropy loss of generation identification, kiln port classification and decoration style discrimination, respectively; L_g is the constraint loss of timing relation graph, which is used to maintain the continuity of samples in adjacent stages in the feature space; L_p is the process path ranking loss, which is used to improve the rationality of path inference results. Through the joint optimization of shared representation, multi-task learning and temporal relation propagation, the model can simultaneously capture the stage characteristics and continuity characteristics of ceramic process, so as to improve the consistency and stability of generation identification, kiln mouth classification, grain style discrimination and process path inference.

4 Aided analysis model and experimental results of cultural relic restoration

4.1 Auxiliary analysis model for disease recognition and restoration of cultural relics

The key of the assistant analysis of cultural relics restoration is not to detect cracks, defects, enamel denuding and pollution deposits in isolation, but to integrate the spatial distribution of diseases, artifacts' technological attributes, similar artifacts' reference information and restoration knowledge rules into the unified computing link, forming a closed-loop model composed of "disease identification, damage assessment, similarity retrieval, knowledge matching and restoration suggestion output". Based on this, this paper constructs a restoration aided analysis framework for ceramic cultural relics, and its overall structure is shown in Figure 5.

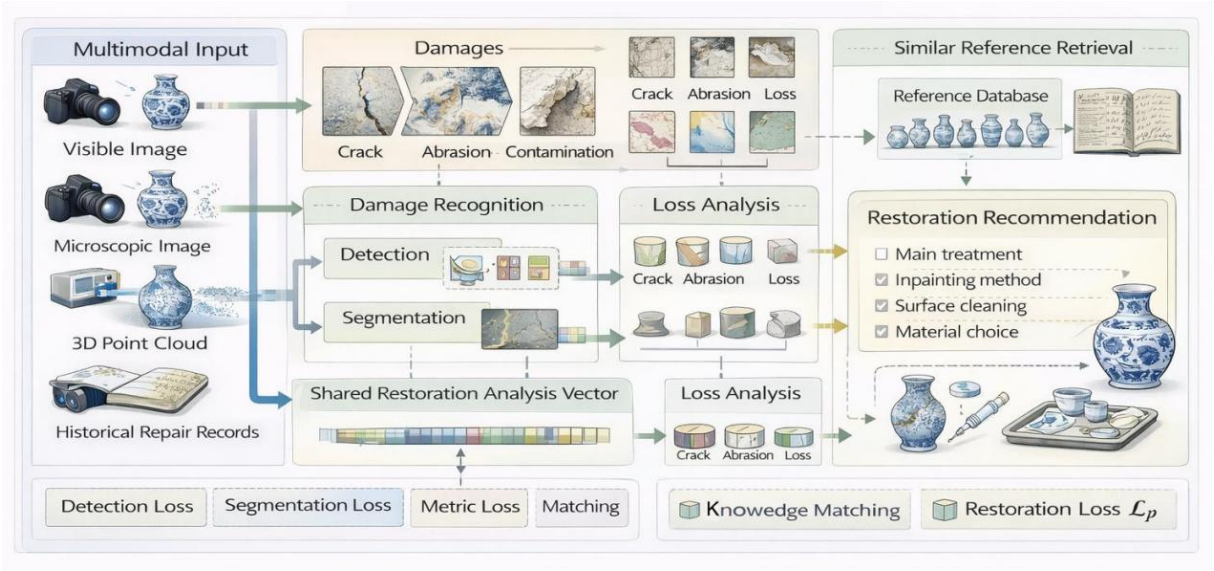

Figure 5: Aided analysis model for disease identification and restoration of cultural relics

The input terminal receives visible light images, microscopic images, three-dimensional morphological fragments and historical repair records, and generates disease representation vectors and process representation vectors after multi-modal coding. The disease branch adopts the collaborative modeling method of detection and segmentation, and performs pixel-level identification and region-level discrimination for cracks, defects, denudations and pollution deposition. Suppose the input sample is Xi, and the disease recognition branch outputs the disease class probability map Pi and the damage mask Mi, then:

$$P_i=\mathrm{Softmax}(f_{det}(X_i)),\quad M_i=\sigma(f_{seg}(X_i)) \tag{44}$$

Here, $f_{det}(\cdot)$ represents the disease classification and detection network, $f_{seg}(\cdot)$ represents the disease segmentation network, and $\sigma(\cdot)$ is the Sigmoid mapping. In order to enhance the representation ability of slender cracks and boundary broken areas, the model introduces a multi-scale dilated convolution and channel attention module in the encoding stage, so that the shallow texture information and deep semantic information are aligned in the feature pyramid, so as to improve the recognition accuracy of small diseases and low contrast damage areas.

In the damage analysis stage, the model does not stop at the output of the disease category directly, but further calculates the proportion of the disease area, the crack length density, the complexity of the defect boundary, and the dispersion of the denuding area to characterize the damage degree of cultural relics. Let the disease mask be Mi and the effective area of the image be Ωi, then the overall damage rate can be defined as follows.

$$D_i=\frac{\sum_{(x,y)\in\Omega_i} M_i(x,y)}{|\Omega_i|} \tag{45}$$

The structural complexity of crack diseases is jointly measured by skeleton length and branch number:

$$C_i^{crack}=\lambda_1 L_i+\lambda_2 B_i \tag{46}$$

Here, L_i represents the total length of the crack skeleton, B_i represents the number of

branch nodes, and λ_1,λ_2 are the weight coefficients. The defective region is evaluated by combining the 3D morphological incomplete boundary and the curvature change of the neighborhood to avoid the area deviation caused by only 2D projection. The above indicators, together with the shape, glazing and body composition characteristics, constitute the repair analysis vector:

$$r_i=[z_i \parallel D_i \parallel C_i^{crack} \parallel C_i^{loss} \parallel C_i^{erosion}] \tag{47}$$

Here, z_i is the multimodal fusion representation obtained in Chapter 3. The vector not only reflects the technical attributes of the current cultural relic, but also describes its damage state, which provides a unified input for similar artifacts retrieval and repair strategy matching.

The similar artifacts retrieval module uses metric learning mechanism to search the reference samples closest to the target cultural relic in the collection sample library in terms of vessel type, decoration, glaze color and material attributes. Let the query cultural relic be characterized as ri and the JTH sample in the library be characterized as rj, then the similarity function is written as follows.

$$S_{ij}=\frac{r_i^{\top} r_j}{\|r_i\|_2 \|r_j\|_2} \tag{48}$$

The top K reference samples NK(i) are selected according to the similarity ranking, and the candidate repair set is constructed by combining their repair records, disease types and processing results. Only relying on similarity retrieval may still be affected by sample noise and historical repair differences. Therefore, this paper further introduces a repair knowledge matching module, and represents the repair rules as a quaternary association structure of "disease type - damage level - process attribute - suggested operation". Let the MTH restoration knowledge rule be denoted as κm, then the matching score between the target cultural relic and the rule is defined as follows.

$$R_{im}=\eta_1 \, sim(r_i,\kappa_m^f)+\eta_2 \, sim(h_i,\kappa_m^d)+\eta_3 \, sim(z_i,\kappa_m^c) \tag{49}$$

Here, κ_m^f represents the disease feature template in the rule, κ_m^d represents the damage level template, κ_m^c represents the process constraint template, and η_1,η_2, and η_3 are the matching weights. The final repair proposal is generated by the combination of the similar object retrieval results and the knowledge matching results:

$$\hat{y}_i^{res}=\arg\max_m \left(\alpha \cdot \frac{1}{K} \sum_{j \in N_K(i)} S_{ij} + \beta \cdot R_{im} \right) \tag{50}$$

This output corresponds to auxiliary conclusions such as repair priority, replantation mode, surface cleaning strategy, and material selection suggestions.

To ensure the synergy between recognition and recommendation, a joint loss function is used for model training:

$$L=\mu_1 L_{det}+\mu_2 L_{seg}+\mu_3 L_{metric}+\mu_4 L_{match} \tag{51}$$

Here, L_{det} is the disease classification detection loss, L_{seg} is the segmentation loss, L_{metric} is the metric learning loss for similarity retrieval, and L_{match} is the repair knowledge

matching loss. The model connects disease identification, damage quantification, similar object retrieval and knowledge rule matching into an integrated calculation process, which can not only realize the fine identification of diseases such as cracks, defects, enamel denudation and pollution deposition, but also output interpretable repair auxiliary suggestions under the constraints of process attributes, which provides a complete model basis for subsequent experimental verification.

4.2 Experimental design and parameter setting

The experimental data set consists of visible light images, microscopic images, spectral detection data, three-dimensional morphological data and text descriptions. The training set, validation set and test set are divided according to cultural relic instances, and the proportions are set to 70%, 15% and 15% to avoid the leakage of different modal samples of the same artifact across the collection. The experimental platform uses Python 3.10, PyTorch 2.1 and CUDA 12.1, and the hardware environment is Intel Xeon Silver 4314 CPU, NVIDIA RTX 4090 GPU and 128 GB memory. The model training batch size is set to 16, the training rounds are 150, the optimizer uses AdamW, the initial learning rate is set to 2×10^{-4}, the weight decay coefficient is 1×10^{-5}, and the cosine annealing scheduling and early stopping strategy are combined to control the convergence process. The image input size is 224×224, the spectral sequence length is intercepted to 256 dimensions according to the effective band, the 3D point cloud sampling points are set to 2048, and the maximum text length is set to 128 tokens. Accuracy, Precision, Recall, F1-score, mAP and IoU were used to evaluate the disease recognition task, and Top-1/Top-3 retrieval accuracy, NDCG and recommendation matching rate were used to measure the repair aided analysis part. The comparison model Settings include single-modal CNN, image-text dual-modal model, feature direct stitching model, bilinear fusion model and cross-modal Transformer model, to verify the advantages of the proposed method in disease recognition accuracy, multi-modal collaboration ability and effectiveness of repair assistance.

4.3 Experimental results and analysis

The experimental results show that the proposed model achieves better performance on the three tasks of process evolution recognition, disease detection and repair assistant recommendation. The accuracy of the traditional feature plus SVM method in the ceramic process evolution recognition task is 78.6%, the mAP of disease detection is 71.4%, and the matching rate of repair recommendation is 69.8%. Although the single-modal CNN is improved to 84.9%, 80.2% and 77.5%, the utilization of decorative semantics, material information and historical records is still insufficient. After introducing image and text bimodality, the three indexes reach 86.7%, 82.6% and 80.9%, respectively, indicating that cross-modal semantic complement has a positive effect on complex sample discrimination. Direct fusion and bilinear fusion further improve the collaborative ability of multi-source information, the accuracy of process evolution recognition increases to 89.3% and 91.1%, the mAP of disease detection increases to 85.9% and 87.4%, and the matching rate of repair recommendation reaches 84.7% and 86.2%, respectively. The proposed model achieves 93.8%, 90.6% and 89.8% on the three tasks, respectively, and the overall performance is the best. As shown in Figure 6, the proposed method always maintains the highest level in each comparison model, indicating that the designed cross-modal attention fusion and unified representation mechanism can effectively enhance the information collaboration between instrument type, ornament, spectral line, 3D shape and text knowledge, which not only improves the accuracy of process stage recognition, but also improves the stability of disease analysis and repair assistant recommendation.

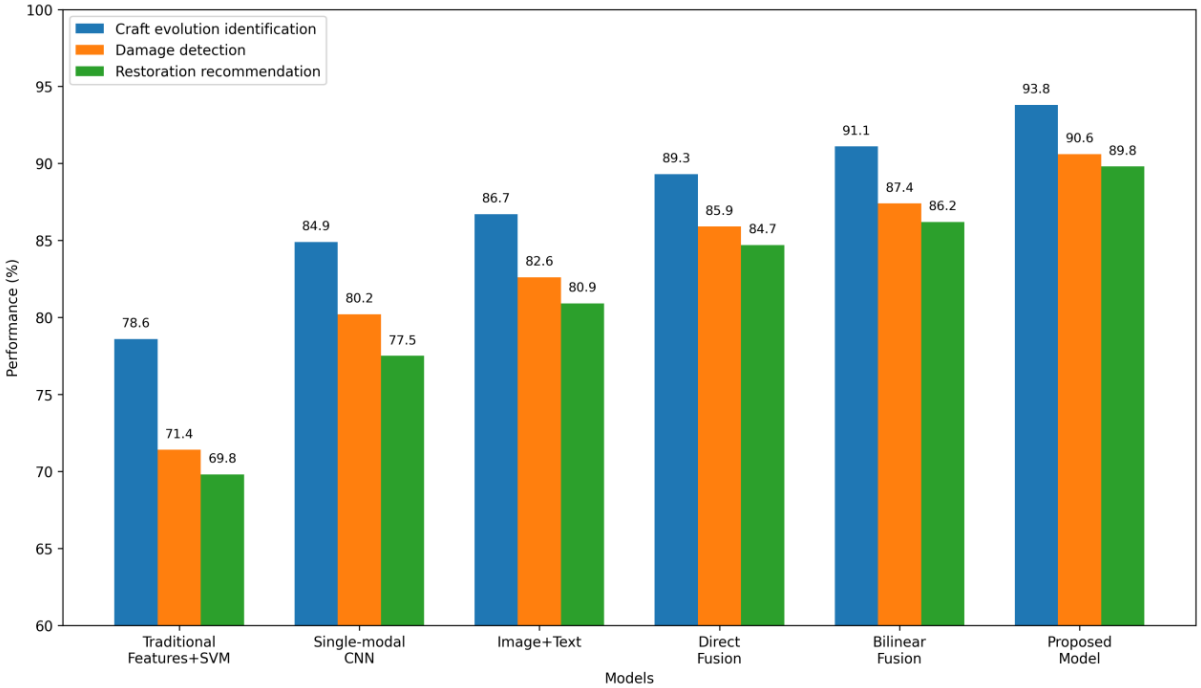

Figure 6: Performance comparison of different models in process evolution identification, disease detection and repair assisted recommendation tasks

4.4 Ablation experiments and discussion of results

In order to test the actual contribution of each component module to the performance of the model, this paper removes the image modality, spectral modality, text modality and deep fusion module respectively based on the complete model, and keeps the training set division, optimization strategy and evaluation index consistent. The ablation results are shown in Table III. The complete model achieves 93.8%, 90.6% and 89.8% in process evolution recognition, disease detection and repair assistant recommendation tasks, respectively. After removing the image modality, the three indexes decreased to 88.7%, 84.1% and 82.5%, with the most significant decrease, indicating that the visible light and microscopic images are still the main information sources for pattern style recognition, crack boundary extraction and surface disease discrimination. After removing the spectral line mode, the accuracy of process evolution identification decreases to 91.2%, and the matching rate of repair recommendation decreases to 85.8%, which indicates that the information of material composition and firing system plays an important role in the distinction of kiln opening, the judgment of age boundary and the matching of repair materials. After removing the text modality, the process evolution identification, disease detection and repair recommendation indexes are reduced to 90.8%, 86.9%and 84.6%, respectively, indicating that history records, repair archives and era labels provide effective semantic constraints in complex damaged samples. After removing the deep fusion module and replacing it with direct stitching, the model performance decreases to 89.3%, 85.9%and 84.7%respectively, indicating that it is difficult to fully model the high-order association between images, spectral lines, 3D shapes and texts by relying only on simple feature stacking, and cross-modal attention and interaction mechanism play a decisive role in the construction of unified representation.

Table 3: Results of ablation experiments

Model Configuration	Technological Evolution Recognition Accuracy (%)	Disease Detection mAP (%)	Restoration Recommendation Matching Rate (%)
Without Image Modality	88.7	84.1	82.5
Without Spectral Modality	91.2	87.3	85.8
Without Text Modality	90.8	86.9	84.6
Without Deep Fusion Module	89.3	85.9	84.7
Full Model	93.8	90.6	89.8

The results show that the image modality directly determines the separability of the disease area and the decorative structure, the spectral line modality strengthens the discrimination ability of material properties and process systems, the text modality makes up for the lack of visual evidence in the damaged samples, and the deep fusion module is responsible for compressing the information from different sources into a consistent semantic space. The lack of any link will weaken the overall inference ability of the model. Under the condition of complex samples, severe denudation, pollution deposition and superposition of multiple diseases will lead to local texture fragmentation, fuzzy boundaries and inconsistent modal information. The model still has misclassification in the overlap area of cracks and pollution. Under the condition of small sample size, the distribution of rare kiln mouth, rare instrument type and special decoration category is insufficient, which is easy to cause the deviation of classification boundary and over-fitting of feature space. Under the condition of cross-era samples, there is a significant inheritance relationship in the proportion of utensils, glaze color performance and decorative language of adjacent periods, and the stage boundaries are not strictly discrete. The judgment of the model on the transitional period samples is still affected by style mixing and knowledge label sparsity. The above phenomena indicate that the follow-up research still needs to be further optimized in the enhancement of scarce samples, cross-era continuous modeling, and fine segmentation of complex disease regions.

5 Conclusion

Focusing on the needs of ceramic process evolution recognition and auxiliary analysis of cultural relic restoration, this paper constructs a multi-modal data system that fuses visible light images, microscopic images, spectral detection, three-dimensional morphology and text description, and forms an integrated computing link from data acquisition, feature coding, cross-modal fusion to evolution analysis and restoration recommendation. In the experiment, the dataset was divided into 70% training set, 15% validation set and 15% test set according to cultural relic instances. The image input size was 224×224, the spectral length was 256 dimensions, the point cloud sampling points were 2048, and the training rounds were 150. The results show that the indicators of the complete model in process evolution recognition, disease detection and repair assistant recommendation tasks reach 93.8%, 90.6% and 89.8%, respectively, which are significantly better than 78.6%, 71.4% and 69.8% of traditional features plus SVM, and 84.9%, 80.2% and 77.5% of single-modal CNN. Ablation experiments further show that multi-modal collaboration and deep fusion module are the key to improve the performance. After removing image modalities, the three indicators are reduced to 88.7%, 84.1% and 82.5%. After removing the spectral line mode, the number decreases to 91.2%, 87.3% and 85.8%. After removing the text mode, the number decreases to

90.8%, 86.9% and 84.6%. After removing the fusion module, it decreases to 89.3%, 85.9% and 84.7%. This indicates that the value of computer technology in the research of ceramic cultural relics has been extended from a single recognition tool to a comprehensive analysis method oriented to process knowledge refining, disease diagnosis and restoration assistant decision-making. In the future, it is still necessary to continue to deepen the direction of few-shot learning, knowledge graph fusion, cross-era continuous modeling and generative repair assistance, so as to improve the adaptability of the model to rare samples, complex diseases and high uncertainty repair scenarios.

Author's Profile

Xu Fan (1990.01-), female, Han ethnicity, born in Ankang, Shaanxi Province. Lecturer at Xi'an Siyuan University, Senior Craft Artist. Research interests: Archaeology, Arts and Crafts.

Zhang Jiannan (1992.10–), male, Han ethnicity, native of Weinan, Shaanxi Yonghua Ceramic Art Culture Co., Ltd., Senior Craft Artist. Research directions: ceramic form, production techniques.

References

[1] Di Angelo L, Di Stefano P, Guardiani E. A review of computer-based methods for classification and reconstruction of 3D high-density scanned archaeological pottery[J]. Journal of Cultural Heritage, 2022, 56: 10-24.

[2] Cardarelli L. A deep variational convolutional Autoencoder for unsupervised features extraction of ceramic profiles. A case study from central Italy[J]. Journal of Archaeological Science, 2022, 144: 105640.

[3] Jin X, Wang X, Xue C. Nondestructive characterization and artificial intelligence recognition of acoustic identifiers of ancient ceramics[J]. Heritage Science, 2023, 11(1): 1-10.

[4] Towarek A, Halicz L, Matwin S, et al. Machine learning in analytical chemistry for cultural heritage: a comprehensive review[J]. Journal of Cultural Heritage, 2024, 70: 64-70.

[5] Pang H, Qi X, Xiao C, et al. Pottery evolution pattern discovery based on deep learning: case study of Miaozigou culture in China[J]. Heritage Science, 2024, 12(1): 1-13.

[6] Wang Q, Xiao X, Liu Z. Using microscopic imaging and ensemble deep learning to classify the provenance of archaeological ceramics[J]. Scientific Reports, 2024, 14(1): 32024.

[7] Liu X, Liu Y, Wang K, et al. A color prediction model for mending materials of the Yuquan Iron Pagoda in China based on machine learning[J]. Heritage Science, 2024, 12(1): 183.

[8] Zheng Q, Yang H, Yang J, et al. Ancient ceramics restoration method based on image processing texture stitching[J]. Heritage Science, 2024, 12(1): 423.

[9] Ling Z, Delnevo G, Salomoni P, et al. Findings on machine learning for identification of archaeological ceramics: A systematic literature review[J]. IEEE Access, 2024, 12: 100167-100185.

[10] Ao J, Xu Z, Li W, et al. Quantitative typological analysis applied to the morphology of export mugs and their social factors in the Ming and Qing dynasties from the perspective of East–West trade[J]. Heritage Science, 2024, 12(1): 1-24.

[11] Ao J, Xu Z, Li W, et al. Analysis of factors related to the morphological evolution of Lingnan export mugs in the 18th-20th centuries in the context of one belt and one road[J]. PLoS One, 2024, 19(8): e0304104.

[12] Jin X, Wang X, Zhang X, et al. Accurate acoustic classification research of visually similar monochrome porcelain fragments[J]. Heritage Science, 2024, 12(1): 1-13.

[13] Stoean R, Bacanin N, Stoean C, et al. Bridging the past and present: AI-driven 3D restoration of degraded artefacts for museum digital display[J]. Journal of Cultural Heritage, 2024, 69: 18-26.

[14] Hu Y, Wu S, Ma Z, et al. Integrating deep learning and machine learning for ceramic artifact classification and market value prediction[J]. npj Heritage Science, 2025, 13(1): 306.

[15] Cardarelli L. PyPotteryInk: One-step diffusion model for sketch to publication-ready archaeological drawings[J]. Journal of Cultural Heritage, 2025, 74: 300-310.

[16] Cardarelli L. PyPotteryLens: An Open-Source Deep Learning Framework for Automated Digitisation of Archaeological Pottery Documentation[J]. Digital Applications in Archaeology and Cultural Heritage, 2025: e00452.

[17] Deng Z, Wu D, Xia M, et al. Localized fading feature extraction method of ancient ceramic decoration[J]. npj Heritage Science, 2025, 13(1): 505.

[18] Liao D, Zeng T, Yang J, et al. Digital prediction of ancient ceramic images missing areas based on deep adversarial and reverse diffusion[J]. npj Heritage Science, 2025, 13(1): 242.

[19] Chen H, Jia H, Liu L. Systematic morphological analysis and stylistic cultural interpretation of Song Dynasty Longquan Porcelain Vases[J]. npj Heritage Science, 2025, 13(1): 423.

[20] Yang L H, Zhou W B, Qiu W R. Chronological classification of Ming and Qing dynasty ceramics images based on an enhanced ResNet50 model[J]. STAR: Science & Technology of Archaeological Research, 2025, 11(1): e2498260.