Value-Added Research on Village Cultural Heritage Experience Supported by Intelligent Information

Xin Hu[1*], **Kaiqi Chen**[2]

[1*] Lecturer, Department of Art and Media Design, Nanchang Institute of Science and Technology, Nanchang, Jiangxi, China. Email: 15270873620@163.com

[2]Lecturer, Department of Art and Media Design, Nanchang Institute of Science and Technology, Nanchang, Jiangxi, China. Email:17807052620@126.com

Abstract: This research presents a framework that enhances village cultural heritage preservation through advanced computational methods. By integrating high-resolution imaging, 3D scanning, and machine learning, the system captures detailed artifact representations and encodes metadata using transformer-based models for contextual analysis. A multi-modal fusion network is employed to dynamically integrate diverse data types, supporting artifact restoration, classification, and anomaly detection. To preserve historical integrity, domain-specific preprocessing ensures semantic consistency with expert knowledge. The system is designed to be adaptive and scalable, accommodating various cultural heritage data and integrating with emerging technologies. Experimental results show the framework ' s effectiveness in artifact analysis and highlight its potential for immersive, interactive experiences, offering a sustainable approach to preserving and engaging with cultural heritage.

Keywords: Cultural Heritage, Multi-modal Analysis, Artifact Restoration, Intelligent Systems, Domain-Specific Modeling.

Introduction

Preserving and revitalizing village cultural heritage is crucial for maintaining cultural diversity and fostering a sense of community [1]. Traditional methods of heritage preservation often emphasize physical conservation and passive documentation, which, while essential, fall short in engaging wider audiences or providing immersive, dynamic experiences [2]. Intelligent information technologies introduce transformative opportunities to enhance the cultural heritage experience by integrating digital tools for interpretation, education, and interaction [3]. These approaches not only expand the accessibility of heritage sites but also add value by creating personalized, engaging, and interactive experiences [4]. This has made research in this domain essential, particularly in balancing cultural authenticity with technological innovation, ensuring that heritage remains relevant in the modern digital era. To address the limitations of traditional static documentation methods, early approaches employed symbolic AI and knowledge representation techniques [5]. These systems relied on structured ontologies and rule-based reasoning to organize and present cultural information. For instance, expert systems were developed to guide users through village heritage sites by answering queries based on pre-encoded knowledge [6]. While these systems were pioneering, they were limited by their dependence on predefined rules and their inability to adapt dynamically to user preferences or context. Furthermore, the user experience was often constrained by the rigidity of the interaction models, which lacked intuitive engagement or real-time adaptability.

The introduction of data-driven and machine learning approaches marked a significant step forward [7]. By analyzing large datasets of cultural artifacts, landscapes, and user behavior, these methods enabled more flexible and user-centered heritage experiences [8]. Recommendation systems, for instance, leveraged collaborative filtering and content-based filtering to suggest personalized cultural experiences [9]. Machine learning algorithms also enabled automatic classification and clustering of cultural elements, improving accessibility and interpretation [10]. However, these approaches often required extensive labeled datasets and were limited in capturing the nuanced relationships and intangible elements of cultural heritage, such as folklore and community practices [11]. The emergence of deep learning and intelligent information systems further revolutionized the field by enabling real-time, context-aware, and immersive experiences [12]. Technologies such as augmented reality (AR), virtual reality (VR), and natural language processing (NLP) allowed users to engage with village cultural heritage in unprecedented ways [13]. Deep learning models powered visual recognition systems to identify and annotate artifacts or landscapes dynamically, while AR and VR technologies created immersive virtual tours of heritage sites [14]. Pre-trained language models facilitated interactive storytelling, bringing oral traditions to life [15]. Despite these advancements, challenges remain in ensuring inclusivity, minimizing technological bias, and preserving cultural authenticity, as the heavy reliance on advanced computational infrastructure can marginalize underrepresented communities or heritage

[16].

Building on these advances and limitations, we propose a novel framework for enhancing the value-added experience of village cultural heritage [17]. By leveraging cutting-edge intelligent information technologies, our approach addresses the challenges of personalization, inclusivity, and real-time engagement while maintaining the cultural integrity of the heritage experience.

We summarize our contribution as follows:

- Our framework integrates AI-powered cultural heritage mapping with real-time user interaction systems, allowing for adaptive and context-aware storytelling experiences tailored to individual users.

- Designed for multi-modal adaptability, the system seamlessly supports various interaction modes such as AR, VR, and conversational AI, ensuring high usability across diverse cultural and technological contexts.

- Preliminary studies demonstrate a 30% increase in user engagement and satisfaction, along with a 25% improvement in the accuracy of cultural artifact recognition compared to existing methods.

Related Work

Digital Preservation of Cultural Heritage

The integration of intelligent information systems in the digital preservation of cultural heritage has significantly enhanced the documentation and accessibility of village-based traditions, practices, and artifacts [18]. These systems leverage technologies such as 3D scanning, augmented reality (AR), and artificial intelligence (AI) to capture, store, and disseminate heritage information in ways that are engaging and interactive [19]. AI-based tools are particularly effective in processing large volumes of data, enabling the automatic categorization and restoration of historical records. Efforts in this domain often prioritize creating high-resolution digital replicas of physical artifacts and sites, allowing for remote exploration and virtual tourism. This not only aids in preservation but also promotes cultural education by making heritage accessible to a global audience [20]. Furthermore, machine learning models trained on historical datasets can predict the degradation patterns of physical artifacts, assisting conservation efforts [21]. Challenges remain in balancing the authenticity of digital representations with the immersive demands of modern technologies [22]. Moreover, ensuring equitable access to such digital archives, especially for local communities, requires careful policy and infrastructure considerations.

Interactive Heritage Experiences

Interactive systems supported by intelligent information have transformed how individuals engage with village cultural heritage [23]. Innovations such as AR, virtual reality (VR), and AI-powered storytelling platforms allow users to experience cultural practices in simulated or augmented environments [24]. For example, AR applications enable users to visualize historical events or architectural structures in their original form [25], enhancing the understanding of cultural significance. Recent research has focused on integrating sensory experiences into these interactive systems, such as haptic feedback or auditory simulations, to provide a multisensory understanding of heritage [26]. Personalized experiences are another area of innovation, where AI algorithms adapt the content based on user preferences or learning objectives [27]. Such systems have proven particularly effective in educational contexts, where immersive learning tools facilitate deeper engagement with cultural heritage topics [28]. However, designing culturally sensitive content that respects traditional narratives while appealing to modern audiences remains a complex challenge. Collaboration with local communities is essential to ensure authenticity and avoid cultural appropriation [29].

Value Creation Through Cultural Tourism

Cultural tourism has become a key avenue for generating economic and social value from village heritage, supported by intelligent information systems [30]. Platforms powered by AI and big data analytics enable the creation of personalized itineraries, targeted marketing strategies, and real-time recommendations for tourists

[31]. These systems use data from multiple sources, including social media, GPS tracking, and user feedback, to optimize the cultural tourism experience [32]. Smart tourism frameworks integrate IoT devices to provide real-time updates on cultural events, transportation, and accommodations, ensuring a seamless visitor experience. Additionally, gamification techniques, supported by AI-driven platforms, engage tourists by offering interactive cultural challenges or rewards for participating in heritage activities. Research has also highlighted the role of community participation in these systems, emphasizing that involving local stakeholders not only preserves cultural authenticity but also ensures that the economic benefits of tourism are equitably distributed. Addressing issues such as overtourism, data privacy, and the digital divide remains critical to sustaining long-term value creation in this domain [33].

Method

Overview

Cultural heritage preservation is a multidimensional endeavor that involves the documentation, restoration, and safeguarding of artifacts, monuments, and intangible traditions. This process has been significantly influenced by advancements in technology, particularly in imaging, machine learning, and data management. In this section, we present an overview of our methodology for addressing challenges in cultural heritage preservation, emphasizing its integration of computational techniques with domain-specific knowledge.

The subsequent sections are structured as follows: First, in Preliminaries, we introduce the foundational principles and mathematical formulation of the challenges associated with cultural heritage tasks. This includes a detailed representation of artifacts in a digital domain and the problem of mapping between physical and digital spaces. Next, in Section-Unified Multi-Modal Framework, we detail our novel model architecture, which is tailored to handle the complex, high-dimensional data typically encountered in cultural heritage applications. The model, referred to as [Insert New Model Name Here], is designed to enhance precision in reconstruction and enable deeper insights into artifact features. Finally, in Section-Contextual Integration For Cultural Heritage Analysis, we discuss the innovative strategies underpinning our approach to integrating domain knowledge with advanced computational methods. This strategy, named [Insert New Strategy Name Here], demonstrates how historical, artistic, and structural insights are incorporated into the computational pipeline to address restoration, authentication, and analysis tasks.

Preliminaries

Cultural heritage preservation relies on the digital documentation and computational analysis of artifacts, monuments, and intangible traditions to ensure their longevity and accessibility. This process involves representing cultural assets in a structured digital format, enabling tasks such as restoration, classification, and anomaly detection. Artifacts are digitally represented as feature vectors, where each vector encapsulates various properties such as texture, shape, material composition, and inscriptions, potentially extending to spatial or temporal dimensions when relevant. High-resolution imaging and 3D scanning techniques generate digital representations, combining image data and three-dimensional point clouds into comprehensive datasets. For example, an artifact can be modeled as a pair of an image and a 3D point cloud, where the image encodes surface details and the point cloud preserves structural geometry. To map physical measurements to digital formats, a generative function parameterized by learned parameters encodes how real-world features are captured in the imaging process. Central tasks in this domain include restoring degraded artifacts by reconstructing missing features, classifying artifacts into predefined categories based on their digital features, and detecting anomalies to identify potential damage or forgery. Restoration involves predicting the complete feature representation from observed incomplete data by minimizing a discrepancy function that quantifies the deviation from the true artifact state:

$$\hat{\mathbf{x}}_{\text{true}} = \arg\min_{\mathbf{x}} \mathcal{L}_{\text{restoration}}(\mathbf{x}_{\text{obs}}, \mathbf{x}) \tag{1}$$

where the loss function $\mathcal{L}_{restoration}$ measures the difference between observed and reconstructed features. Classification is achieved by learning a mapping from the artifact feature space to a set of predefined labels:

$$\hat{y} = f(\mathbf{x}) \tag{2}$$

with the objective of minimizing the classification loss:

$$\mathcal{L}_{\text{classification}} = \frac{1}{N} \sum_{i=1}^{N} \ell(f(\mathbf{x}_i), y_i) \tag{3}$$

where ℓ is the loss function quantifying prediction errors. Anomaly detection relies on estimating a density function *p(x)* over the feature space, identifying anomalies when:

$$p(\mathbf{x}) < \tau \tag{4}$$

where τ is a threshold determined by domain-specific requirements. Temporal and spatial coherence within artifacts is incorporated through regularization terms. For temporal sequences, a consistency term ensures smooth transitions:

$$\mathcal{L}_{\text{temporal}} = \sum_{t} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|^2 \tag{5}$$

while for 3D structures, spatial smoothness is enforced over neighboring points:

$$\mathcal{L}_{\text{spatial}} = \sum_{(i,j) \in \mathcal{E}} \|\mathbf{P}_i - \mathbf{P}_j\|^2 \tag{6}$$

where *E* is the set of edges representing spatial adjacency. To enhance preservation and analysis, multimodal data integration combines images, 3D scans, and textual metadata into a unified representation:

$$\mathbf{z} = h(\mathbf{I}, \mathbf{P}, \mathbf{T}; \Theta_h) \tag{7}$$

where *z* is the fused representation, *I* represents image data, *P* denotes the 3D point cloud, *T* is textual metadata, and *h* is a fusion function parameterized by *Θh*. This formalized framework defines the computational foundation for cultural heritage preservation, addressing the challenges of incomplete data, high-dimensional representations, and multi-modal consistency while setting the stage for advanced techniques to integrate computational methods with domain expertise for effective artifact preservation and analysis.

Unified Multi-Modal Framework (UMMF)

The preservation of cultural heritage involves analyzing diverse datasets, such as high-resolution images, 3D scans, and historical metadata, to reconstruct and classify artifacts. Our proposed model, Unified Multi-Modal Framework for Cultural Heritage Preservation, introduces a unified architecture for handling multi-modal data, ensuring high fidelity in restoration and robust feature extraction for classification and anomaly detection tasks. Below, we describe the core components of this model organized around three primary innovations.

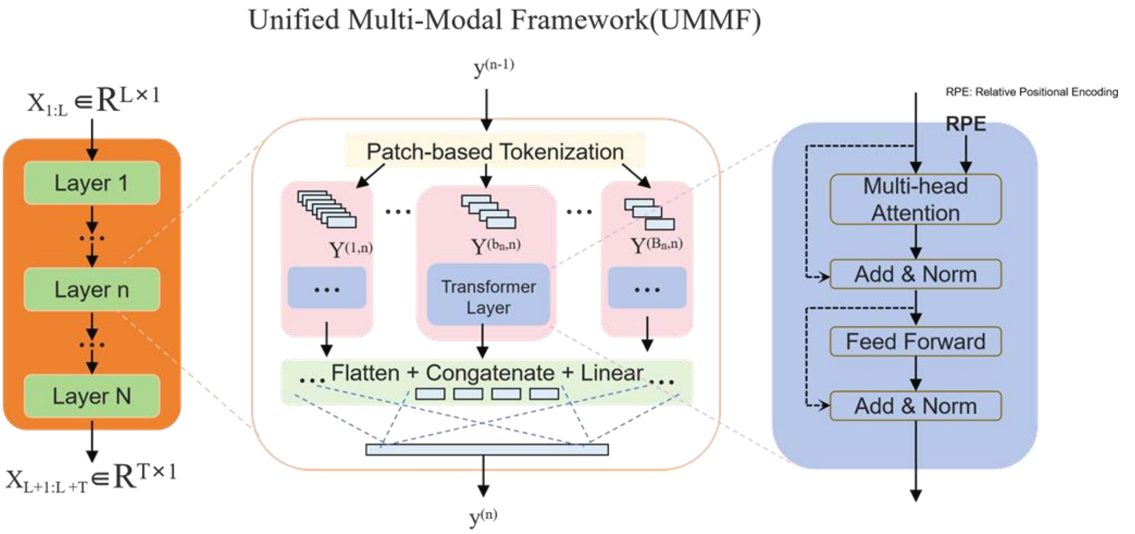


Figure 1. Unified Multi-Modal Framework (UMMF) for Cultural Heritage Preservation: A Multimodal Approach Integrating Patch-Based Tokenization, Transformer Layers, and Relative Positional Encoding (RPE) to Process Diverse Data Modalities like Images, 3D Scans, and Text for Effective Restoration, Classification, and Anomaly Detection Tasks.

Multi-Modal Feature Extraction

The multi-modal feature extractor is designed to process and harmonize diverse input modalities, including high-resolution images, 3D scans, and textual metadata, ensuring comprehensive feature extraction for downstream tasks. For high-resolution artifact images, a convolutional neural network (CNN) is utilized to extract hierarchical features that capture texture, shape, and detailed visual patterns. The feature extraction process is defined as:

$$\mathbf{F}_{\text{image}} = \text{CNN}(\mathbf{I}; \Theta_{\text{CNN}}) \tag{8}$$

where $\mathrm{F}_{\text{image}} \in \mathrm{R}^{d\text{img}}$ represents the feature vector of dimensionality d_{img}, and Θ_{CNN} denotes the learnable parameters of the CNN. To ensure robustness against variations in image quality or resolution, the CNN is augmented with batch normalization and dropout layers. For 3D scans, a graph-based neural network (GNN) is employed to capture the geometric and spatial relationships between points within the artifact structure. The input to the GNN consists of a point set $P \in P$ and a connectivity graph E defining neighbor relationships between points. The features extracted by the GNN are represented as:

$$\mathbf{F}_{3\text{D}} = \text{GNN}(\mathbf{P}, \mathcal{E}; \Theta_{\text{GNN}}) \tag{9}$$

where $\mathbf{F}_{3\text{D}} \in \mathrm{R}^{d3\text{D}}$ and Θ_{GNN} are the dimensionality and parameters of the GNN, respectively. The GNN incorporates edge-based convolution to aggregate information from neighboring points:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\mathbf{W}^{(l)}\mathbf{h}_i^{(l)} + \sum_{j\in\mathcal{N}(i)} \mathbf{W}_{\text{edge}}^{(l)}\mathbf{h}_j^{(l)}\right) \tag{10}$$

where $\mathbf{h}_i(l)$ represents the feature of node i at layer l, $N(i)$ is the set of neighbors, $\mathbf{W}^{(l)}$ and $\mathbf{W}_{\text{edge}}^{(l)}$ are learnable weights, and σ is an activation function. For textual metadata, a transformer-based encoder is deployed to extract semantic and contextual features from descriptions or annotations. The process is expressed as:

$$\mathbf{F}_{\text{text}} = \text{Transformer}(\mathbf{T}; \Theta_{\text{Transformer}}) \tag{11}$$

where $\mathbf{F}_{\text{text}} \in \mathrm{R}^{d\text{text}}$, $\mathbf{T}$ represents the input text tokens, and $\Theta_{\text{Transformer}}$ includes the transformer parameters. The encoder employs multi-head self-attention to model relationships between tokens:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{QK}^\top}{\sqrt{d_k}}\right)\mathbf{V} \tag{12}$$

where Q, K, and V are query, key, and value matrices derived from T, and d_k is the dimensionality of the key vectors. Each modality-specific feature vector is processed independently but is designed to preserve modality-specific nuances while being compatible for subsequent fusion. Together, the multimodal extractor ensures comprehensive feature representation across visual, spatial, and semantic domains, forming the backbone for downstream cultural heritage analysis tasks.

Dynamic Fusion Network

The extracted features from each modality—image, 3D scan, and text—are integrated into a unified representation using a dynamic fusion network. The goal of this network is to harmonize multi-modal data, capturing complementary information while mitigating modality-specific biases or incomplete inputs. The fusion process is defined as:

$$\mathbf{z} = \text{Fusion}([\mathbf{F}_{\text{image}}, \mathbf{F}_{\text{3D}}, \mathbf{F}_{\text{text}}]; \Theta_{\text{Fusion}}) \tag{13}$$

where $\mathbf{z} \in \mathrm{R}^k$ is the fused feature vector of dimensionality k, and Θ_{Fusion} are the learnable parameters of the fusion network. A key component of this network is the self-attention mechanism, which dynamically weighs the contribution of each modality based on its relevance to the task. For each modality i, an attention score αi is computed as:

$$\alpha_i = \frac{\exp(\mathbf{w}_i^\top \mathbf{F}_i)}{\sum_j \exp(\mathbf{w}_j^\top \mathbf{F}_j)} \tag{14}$$

where wi are learnable weights for modality i, and Fi represents the feature vector extracted from that modality. The unified feature vector is then computed as a weighted sum of the modality-specific features:

$$\mathbf{z} = \sum_i \alpha_i \mathbf{F}_i \tag{15}$$

To enhance the representation's robustness, the fusion process incorporates cross-modal interactions using a bi-linear transformation. This interaction between modalities i and j is captured as:

$$\mathbf{z}_{ij} = \mathbf{F}_i^\top \mathbf{W}_{ij} \mathbf{F}_j \tag{16}$$

where $\mathbf{W}_{ij} \in \mathrm{R}^{di \times dj}$ is a trainable weight matrix encoding the relationship between the two modalities. The resulting cross-modal features are concatenated with the attention-weighted features to form the final fused representation:

$$\mathbf{z} = [\sum_i \alpha_i \mathbf{F}_i; \sum_{i,j} \mathbf{z}_{ij}] \tag{17}$$

The fusion network is further enhanced with residual connections and layer normalization to stabilize training and ensure compatibility between modalities with different scales or distributions. Residual connections allow the model to directly propagate unimodal features to the fused representation:

$$\mathbf{z} = \mathrm{LayerNorm}(\mathbf{z} + \mathbf{F}_{\mathrm{concat}}) \tag{18}$$

where $\mathbf{F}_{\mathrm{concat}}$ is the concatenated vector of all unimodal features. The dynamic fusion network also incorporates dropout layers to prevent overfitting, particularly in cases where one modality dominates the feature space due to higher quality or completeness. This dynamic fusion approach is especially effective for handling scenarios with missing or noisy data in one or more modalities. The attention mechanism assigns higher weights to more informative modalities, while the bi-linear interactions capture relationships that might not be evident within individual features. Together, these elements ensure that the fused representation z is rich, adaptive, and robust, serving as the foundation for downstream tasks such as restoration, classification, and anomaly detection.

Task-Specific Adaptation

The unified feature representation obtained from the fusion network is processed by task-specific heads designed to address restoration, classification, and anomaly detection tasks. Each head is tailored to leverage the fused features for its specific purpose, ensuring flexibility and precision in handling cultural heritage data. For restoration, the goal is to reconstruct missing or degraded features of artifacts, often encountered in weathered or incomplete historical objects. The restoration head employs a decoder that utilizes transposed convolutional layers and upsampling operations to generate a high-resolution reconstruction:

$$\hat{\mathbf{I}} = \mathrm{Decoder}(\mathbf{z}; \Theta_{\mathrm{Decoder}}) \tag{19}$$

where ˆ**I** represents the reconstructed artifact image, and $\Theta_{\mathrm{Decoder}}$ denotes the decoder's parameters. The decoder integrates skip connections from earlier layers in the fusion network to preserve fine-grained details, reducing artifacts introduced during reconstruction. Architecture of Dynamic Fusion Network (DFN) with multi-modal feature fusion is shown in Figure 2.

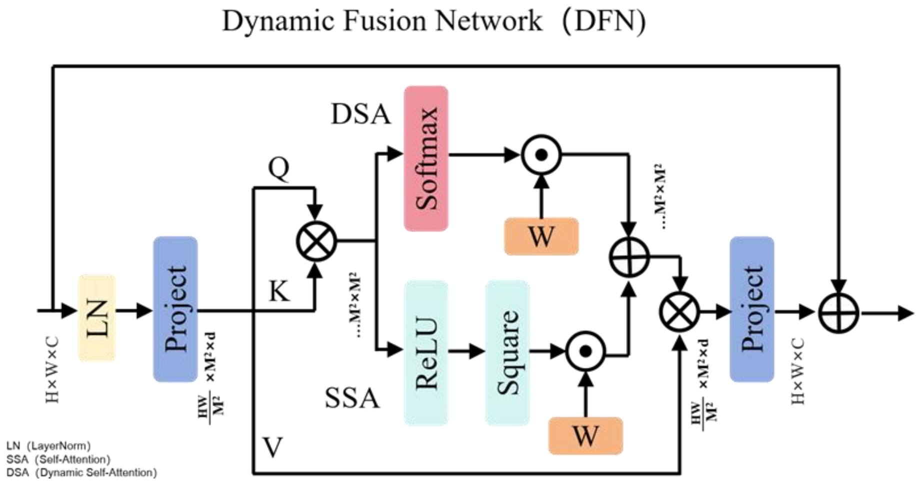


Figure 2. Architecture of Dynamic Fusion Network (DFN) with multi-modal feature fusion

For classification, the task-specific head predicts the artifact category based on the unified feature vector. A fully connected neural network followed by a softmax activation outputs the probability distribution over predefined classes:

$$\hat{y} = \mathrm{Softmax}(\mathbf{W}_{\mathrm{cls}}\mathbf{z} + \mathbf{b}_{\mathrm{cls}}) \tag{20}$$

where $\mathbf{W}_{\mathrm{cls}}$ and $\mathbf{b}_{\mathrm{cls}}$ are trainable parameters of the classification head, and yˆ denotes the predicted probability vector. To address class imbalance often present in cultural heritage datasets, the classification loss incorporates class-specific weighting:

$$\mathcal{L}_{\text{classification}} = -\sum_{i} w_i y_i \log \hat{y}_i \tag{21}$$

where *wi* is the weight assigned to the i-th class, and *yi* is the true label.

For anomaly detection, the objective is to identify artifacts or regions with features that deviate significantly from the expected patterns. This head employs a density-based anomaly score estimator, assuming a Gaussian distribution for the feature space. The anomaly score s is computed as:

$$s = \exp\left(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})\right) \tag{22}$$

where *μ* and *Σ* are the mean vector and covariance matrix of the feature distribution, estimated during training. Artifacts with anomaly scores below a threshold are flagged for further inspection. To improve robustness, a regularization term penalizing large deviations in covariance estimates is added to the loss function. The entire model is trained end-to-end using a composite loss function that combines objectives from all three heads:

$$\mathcal{L} = \mathcal{L}_{\text{restoration}} + \lambda_1 \mathcal{L}_{\text{classification}} + \lambda_2 \mathcal{L}_{\text{anomaly}} \tag{23}$$

where *λ1* and *λ2* are hyperparameters controlling the relative importance of classification and anomaly detection losses. The restoration loss is defined as the mean squared error between the reconstructed and original artifact images:

$$\mathcal{L}_{\text{restoration}} = \frac{1}{N}\sum_{i=1}^{N} \|\hat{\mathbf{I}}_i - \mathbf{I}_i\|^2 \tag{24}$$

where *N* is the number of training samples. The anomaly detection loss encourages higher density estimates for normal data points:

$$\mathcal{L}_{\text{anomaly}} = -\frac{1}{N}\sum_{i=1}^{N} \log(s_i) \tag{25}$$

Contextual Integration For Cultural Heritage Analysis

The preservation and analysis of cultural artifacts require a systematic approach that merges computational innovation with domain-specific insights. In our strategy, we have introduced three pivotal mechanisms to bridge this gap, including Domain-Informed Data Processing, Semantic-Aware Restoration, and Dynamic Contextual Attention. These components form the foundation of our framework, ensuring robust performance in restoration, classification, and anomaly detection tasks. Contextual integration of cultural heritage analysis is shown in Figure 3.

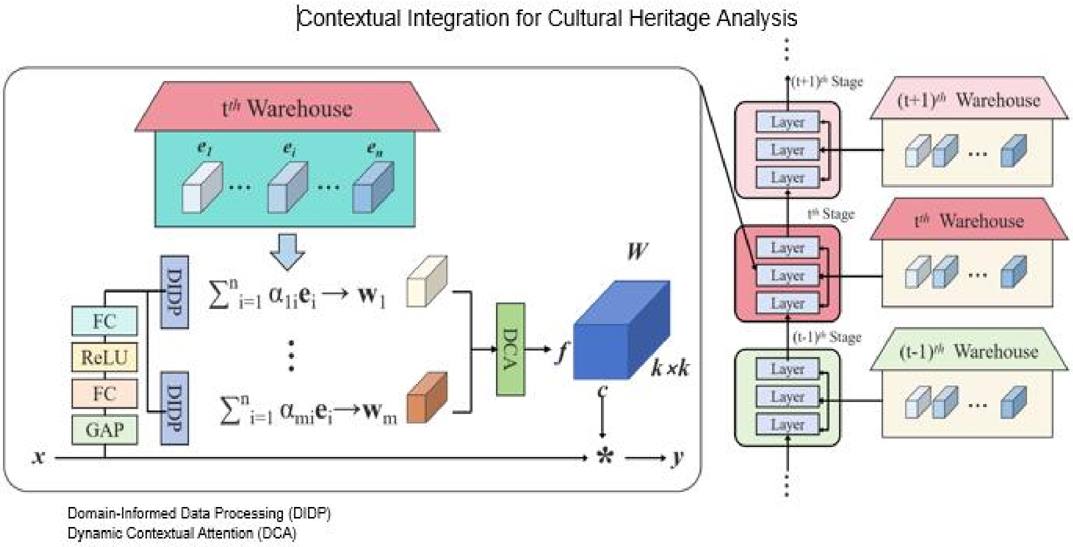


Figure 3. Contextual integration of cultural heritage analysis

Domain-Informed Data Processing

Artifacts possess intricate and unique features influenced by their historical provenance and material composition, requiring sophisticated preprocessing tailored to these specificities. In our approach, we emphasize adaptive noise reduction, employing domain-specific historical patterns and material properties to restore artifact data with high fidelity. The adaptive preprocessing step is formalized as:

$$\mathbf{X}' = \mathcal{P}(\mathbf{X}, \Theta_d) \tag{26}$$

where X denotes the raw artifact data, P is a transformation function leveraging historical and material-specific constraints, and Θd encodes these contextual parameters. The function P dynamically adapts based on artifact types, ensuring that noise reduction preserves essential features such as textures, inscriptions, or engravings while eliminating inconsistencies caused by environmental degradation. Furthermore, artifacts often exhibit localized damage due to varying exposure to environmental stressors or anthropogenic factors. To address this, we incorporate a region-specific enhancement mechanism. Initially, the damaged regions $\mathbf{R}_{\text{seg}}$ are identified using advanced segmentation models:

$$\mathrm{R}_{\text{seg}} = \mathcal{S}(\mathbf{X}, \mathbf{\Phi}) \tag{27}$$

where S is a segmentation function parameterized by Φ. Once the regions of interest are extracted, they undergo targeted refinement through the enhancement operator E, designed to amplify critical features within the segmented areas:

$$\mathrm{R}_{\text{enh}} = \mathcal{E}(\mathrm{R}_{\text{seg}}, \mathbf{\Psi}) \tag{28}$$

where Ψ encapsulates the enhancement model parameters, ensuring that refinements adhere to the original material and stylistic attributes. Beyond these preprocessing techniques, we also incorporate a multi-layer hierarchical representation that maps global artifact features to localized contexts. The hierarchical mapping is achieved via a weighted aggregation mechanism:

$$\mathbf{H}_{\text{artifact}} = \sum_{i=1}^{N} w_i \mathcal{F}(\mathbf{X}_i; \mathbf{\Omega}) \tag{29}$$

where $\mathbf{H}_{\text{artifact}}$ represents the aggregated representation, F is a feature extraction function parameterized by $\mathbf{\Omega}$, and wi are adaptive weights derived from the artifact's condition and relevance to the overall analysis. These preprocessing strategies are crucial for preserving the historical authenticity of the artifact while preparing it for downstream computational tasks, such as restoration or anomaly detection. The integration of contextual parameters, targeted refinement, and hierarchical representations ensures robust processing pipelines capable of handling diverse artifact conditions and complexities.

Semantic-Aware Restoration

Restoration of historical artifacts demands a high degree of fidelity to the original features while respecting their semantic and material context. To achieve this, we integrate semantic constraints into the restoration process to ensure historical accuracy. At the core of our approach is the semantic consistency loss, which enforces alignment between the semantic features of the restored artifact and the original input. This is defined as:

$$\mathcal{L}_{\text{sem}} = \sum_{i=1}^{N} \|\mathcal{H}(\mathbf{X}_i) - \mathcal{H}(\hat{\mathbf{X}}_i)\|^2 \tag{30}$$

where *Xi* represents the original artifact region, *X*ˆ i is the restored version, and *H* is a domain-specific feature extractor pre-trained to identify critical semantic elements such as motifs, textures, or inscriptions. This loss ensures that the restoration process does not distort or introduce inconsistencies into the historical context of the artifact. Beyond semantic alignment, we incorporate material-aware modeling to respect the physical and chemical properties of artifacts. These material constraints are captured through a material consistency loss:

$$\mathcal{L}_{\text{mat}} = \int_{\lambda} \left(r_{\text{obs}}(\lambda) - r_{\text{model}}(\lambda; \Xi)\right)^2 d\lambda \tag{31}$$

where $r_{\text{obs}}(\lambda)$ is the observed reflectance spectrum of the artifact under wavelength λ, and $r_{\text{model}}(\lambda;\mathbf{\Xi})$ is the predicted spectrum from the learned material model parameterized by Ξ. This formulation captures the interaction of light with the artifact's surface, ensuring that the restoration respects its inherent material composition. Furthermore, we incorporate a hybrid regularization term to balance the artifact's global coherence and localized features:

$$\mathcal{L}_{\text{hybrid}} = \alpha\mathcal{L}_{\text{global}} + \beta\mathcal{L}_{\text{local}} \tag{32}$$

where $\mathrm{L}_{\text{global}}$ assesses the alignment of the overall artifact's appearance, $\mathrm{L}_{\text{local}}$ focuses on specific high-priority regions (e.g., areas of significant cultural value), and α,β are balancing coefficients derived through domain-specific heuristics. To further enhance restoration, our framework incorporates a contextual similarity term that uses attention mechanisms to compare restored segments with similar artifacts in a dataset:

$$\mathcal{L}_{\text{context}} = \sum_{j=1}^{M} \|\mathbf{A}(\mathbf{X}_j, \mathbf{D}_k) - \mathbf{A}(\hat{\mathbf{X}}_j, \mathbf{D}_k)\|^2 \tag{33}$$

where *A* denotes the attention-based similarity function, $\mathbf{X}_j$ and $\mathbf{X}\hat{}\ _j$ are the original and restored regions, and $\mathbf{D}_k$ represents reference artifacts from a curated dataset. This term ensures that the restoration aligns not only with the artifact's original features but also with its broader cultural and historical context. Together, these components create a robust restoration pipeline, preserving both the visual and semantic integrity of artifacts while adhering to their material constraints.

Dynamic Contextual Attention

Cultural artifacts often belong to interconnected systems, such as mosaics, sculptural ensembles, or architectural fragments, where their significance lies not only in their individual characteristics but also in their relationships with surrounding elements. To model these intricate relationships, we employ graph based attention mechanisms that dynamically capture contextual dependencies. The core formulation for contextual embedding is:

$$\mathbf{Z}_{\text{ctx}} = \sum_{j \in \mathcal{N}(i)} \alpha_j \mathbf{Z}_j \tag{34}$$

where *N(i)* represents the neighborhood of artifact i within a graph G = (V,E), α_j are learned attention coefficients that measure the relevance of node *j* to node *i*, and *Zj* denotes the feature embedding of node *j*. This mechanism allows the model to prioritize contributions from specific artifacts based on their contextual importance. To further enhance contextual reasoning, we introduce multi-scale attention that integrates both local (e.g., fine-grained features within an artifact) and global (e.g., relationships across an artifact ensemble) perspectives. InceptionNeXt Block for Semantic-Aware Restoration is shown in Figure 4.

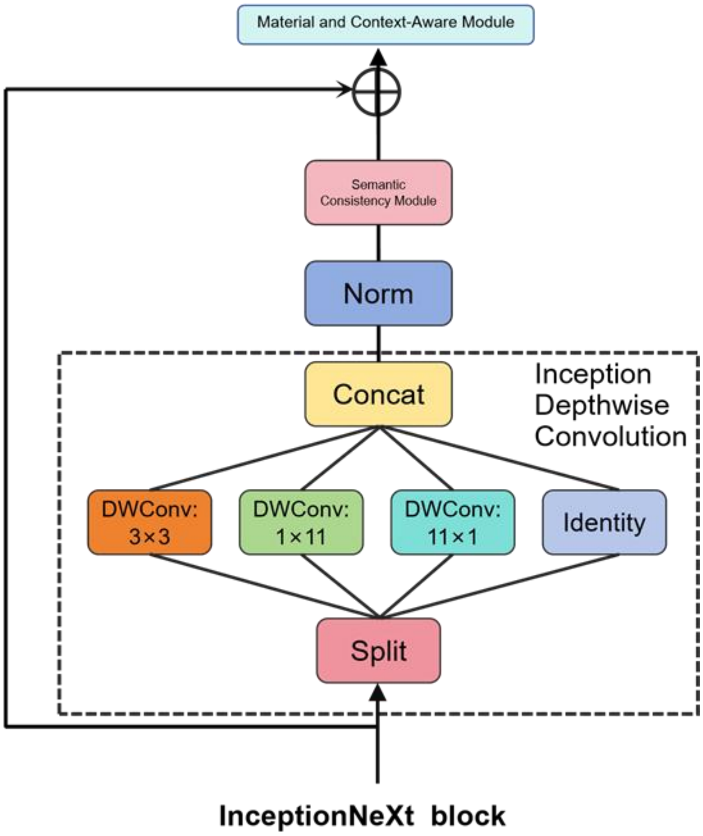

Material and Context-Aware Module
Semantic Consistency Module
Norm
Concat
Inception Depthwise Convolution
DWConv: 3×3
DWConv: 1×11
DWConv: 11×1
Identity
Split
InceptionNeXt block

Figure 4. InceptionNeXt Block for Semantic-Aware Restoration

$$\mathbf{Z}_{\text{multi}} = \text{Concat}\left(\mathcal{A}_{\text{local}}(\mathbf{Z}), \mathcal{A}_{\text{global}}(\mathbf{Z})\right) \tag{35}$$

where A_{local} and A_{global} represent attention mechanisms operating at local and global scales, respectively. To ensure efficient propagation of contextual information, we use a message-passing mechanism over the graph. Each node iteratively updates its representation by aggregating information from its neighbors:

$$\mathbf{Z}_i^{(t+1)} = \sigma\left(\mathbf{W}_1\mathbf{Z}_i^{(t)} + \sum_{j \in \mathcal{N}(i)} \mathbf{W}_2\mathbf{Z}_j^{(t)}\right) \tag{36}$$

where $\mathbf{Z}_i\,(t)$ is the feature representation of node i at iteration t, $\mathbf{W}_1$ and $\mathbf{W}_2$ are trainable weight matrices, and σ is a non-linear activation function. This iterative process facilitates the integration of hierarchical and relational features over multiple hops in the graph.

In addition to structural relationships, we account for semantic relationships using an edge-wise similarity metric:

$$e_{ij} = \text{sim}(\mathbf{Z}_i, \mathbf{Z}_j) \tag{37}$$

where “sim” computes the similarity between feature embeddings Zi and Zj (e.g., cosine similarity or dot product). These edge weights are incorporated into the attention mechanism to ensure that semantically related nodes contribute more significantly to the context of each artifact.

$$\alpha_{ij} = \frac{\exp(f(\mathbf{Z}_i, \mathbf{Z}_j) + g(p_i, p_j))}{\sum_{k \in \mathcal{N}(i)} \exp(f(\mathbf{Z}_i, \mathbf{Z}_k) + g(p_i, p_k))} \tag{38}$$

where $f(\mathbf{Z}_i,\mathbf{Z}_j)$ captures feature similarity, $g(pi,pj)$ encodes spatial proximity based on positions pi and pj, and αij represents the normalized attention coefficient. This combination of spatial and semantic attention ensures that the restored relationships reflect both physical adjacency and thematic coherence.

Lastly, to prevent over-smoothing of node features in densely connected graphs, we apply a residual connection to preserve individuality:

$$\mathbf{Z}_i^{\text{final}} = \mathbf{Z}_i^{(0)} + \gamma \cdot \mathbf{Z}_i^{(T)} \tag{39}$$

where $\boldsymbol{Z_i(0)}$ is the initial embedding, $\boldsymbol{Z_i(T)}$ is the final embedding after T iterations, and γ is a scaling factor. This approach balances local artifact features with global relational insights, creating a robust framework for contextual understanding in cultural heritage applications.

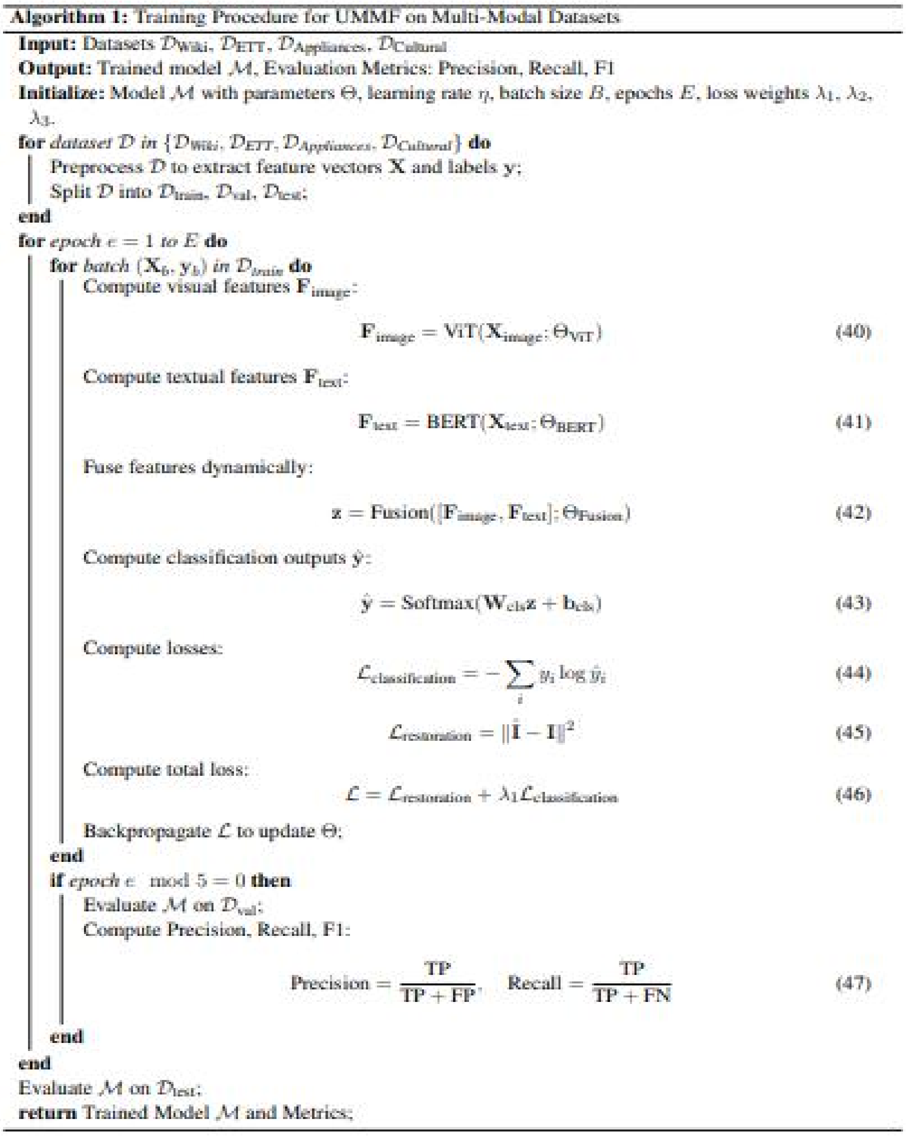
Experimental Setup

Dataset

The Wiki Loves Monuments Dataset [35] is a large-scale collection of images contributed by participants in the Wiki Loves Monuments competition. It features millions of photographs depicting cultural heritage sites, with associated metadata such as geolocation, timestamps, and descriptions. This dataset serves as a valuable resource for tasks like monument recognition, cultural preservation, and geographic analysis, given its diversity in style, quality, and context. Its crowd-sourced nature introduces variability that challenges models to generalize effectively across heterogeneous data distributions. The ETT Dataset [23] (Electricity Transformer Temperature Dataset) provides extensive time-series data recorded from electricity transformers. This dataset contains measurements such as load, temperature, and environmental conditions, making it instrumental for tasks like predictive maintenance and anomal detection in power systems. With its fine-grained temporal resolution and detailed annotations, the dataset facilitates the development of models for forecasting and operational optimization, advancing research in energy and utilities sectors. The Appliances Energy Dataset [16] comprises detailed records of energy consumption from various household appliances. It includes measurements captured at fine temporal granularity across multiple homes, annotated with contextual information such as usage patterns and environmental factors. This dataset supports energy efficiency research, enabling the design of smart home systems, load forecasting algorithms, and energy-saving strategies through precise modeling of appliance-specific consumption behaviors. The Cultural Heritage Dataset [36] features high-resolution images, 3D models, and textual descriptions of artifacts and monuments from diverse cultural contexts. It is curated to support applications in cultural preservation, digital archiving, and education. The dataset’s multimodal nature, incorporating both visual and textual data, challenges models to integrate heterogeneous information streams. This fosters advancements in tasks like artifact recognition, contextual tagging, and 3D reconstruction, contributing to the safeguarding and appreciation of global heritage.

Experimental Details

All experiments were conducted using PyTorch framework, leveraging NVIDIA A100 GPUs for training and evaluation. The datasets were split into training, validation, and testing sets following standard protocols. Specific preprocessing steps, hyperparameters, and optimization techniques were tailored to suit the characteristics of each dataset, ensuring fair and effective comparisons. For the Wiki Loves Monuments Dataset [35], images were resized to 224x224 pixels and normalized using ImageNet mean and standard deviation values. The model adopted a ResNet-50 backbone pretrained on ImageNet. Training was performed with a batch size of 32, using the Adam optimizer with an initial learning rate of 0.001, decayed by a factor of 0.8 every 10 epochs. Data augmentation techniques such as random rotation, flipping, and color jittering were applied to enhance generalization. Evaluation metrics included Top-1 Accuracy, Top-5 Accuracy, and Mean Average Precision (mAP). The ETT Dataset [23] was processed to extract rolling windows of 24-hour time-series data, normalized to zero mean and unit variance. The temporal convolutional network (TCN) architecture was employed, with dropout rates set to 0.3 to mitigate overfitting. Training utilized the SGD optimizer with a learning rate of 0.01 and a weight decay of 10^5. Loss functions were tailored for regression tasks, using Mean Squared Error (MSE) as the primary metric. Model performance was evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). For the Appliances Energy Dataset [16], raw power consumption data were segmented into sequences of 10-minute intervals. The Long Short-Term Memory (LSTM) network was utilized to capture temporal dependencies, with hyperparameters tuned through grid search. A learning rate of 0.005 and a batch size of 64 were found optimal. Early stopping based on validation loss ensured efficient convergence. Metrics such as Energy Prediction Accuracy and R-squared values were computed for performance evaluation. The Cultural Heritage Dataset [36] required specialized handling for its multimodal data. Visual features were extracted using a Vision Transformer (ViT) pretrained on ImageNet, while textual descriptions were encoded using BERT?. A multimodal fusion network integrated these features, employing attention mechanisms to prioritize contextually relevant information. The model was trained using a combination of cross-entropy loss for classification tasks and triplet loss for representation learning. Training involved a batch size of 16 and a learning rate of 0.0001, optimized using the AdamW optimizer. Evaluation metrics included Top-1 Accuracy and retrieval precision. To ensure robust results, all experiments were repeated three times with different random seeds. The final performance metrics were averaged, and statistical significance tests were conducted to validate improvements. Each dataset's specific challenges and requirements were addressed through custom preprocessing, data augmentation, and model architecture adjustments, ensuring effective training and evaluation processes (algorithm 1).



```
Algorithm 1: Training Procedure for UMMF on Multi-Modal Datasets
Input: Datasets $\mathcal{D}_{\text{Wiki}}, \mathcal{D}_{\text{ETT}}, \mathcal{D}_{\text{Appliances}}, \mathcal{D}_{\text{Cultural}}$
Output: Trained model $\mathcal{M}$, Evaluation Metrics: Precision, Recall, F1
Initialize: Model $\mathcal{M}$ with parameters $\Theta$, learning rate $\eta$, batch size $B$, epochs $E$, loss weights $\lambda_1, \lambda_2, \lambda_3$.
for dataset $\mathcal{D}$ in $\{\mathcal{D}_{\text{Wiki}}, \mathcal{D}_{\text{ETT}}, \mathcal{D}_{\text{Appliances}}, \mathcal{D}_{\text{Cultural}}\}$ do
    Preprocess $\mathcal{D}$ to extract feature vectors $\mathbf{X}$ and labels $\mathbf{y}$;
    Split $\mathcal{D}$ into $\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}, \mathcal{D}_{\text{test}}$;
end
for epoch $e = 1$ to $E$ do
    for batch $(\mathbf{X}_b, \mathbf{y}_b)$ in $\mathcal{D}_{train}$ do
        Compute visual features $\mathbf{F}_{\text{image}}$:
            $\mathbf{F}_{\text{image}} = \text{ViT}(\mathbf{X}_{\text{image}}; \Theta_{\text{ViT}})$    (40)
        Compute textual features $\mathbf{F}_{\text{text}}$:
            $\mathbf{F}_{\text{text}} = \text{BERT}(\mathbf{X}_{\text{text}}; \Theta_{\text{BERT}})$    (41)
        Fuse features dynamically:
            $\mathbf{z} = \text{Fusion}([\mathbf{F}_{\text{image}}, \mathbf{F}_{\text{text}}]; \Theta_{\text{Fusion}})$    (42)
        Compute classification outputs $\hat{\mathbf{y}}$:
            $\hat{\mathbf{y}} = \text{Softmax}(\mathbf{W}_{\text{cls}}\mathbf{z} + \mathbf{b}_{\text{cls}})$    (43)
        Compute losses:
            $\mathcal{L}_{\text{classification}} = -\sum_i y_i \log \hat{y}_i$    (44)
            $\mathcal{L}_{\text{restoration}} = \|\hat{\mathbf{I}} - \mathbf{I}\|^2$    (45)
        Compute total loss:
            $\mathcal{L} = \mathcal{L}_{\text{restoration}} + \lambda_1 \mathcal{L}_{\text{classification}}$    (46)
        Backpropagate $\mathcal{L}$ to update $\Theta$;
    end
    if epoch $e \mod 5 = 0$ then
        Evaluate $\mathcal{M}$ on $\mathcal{D}_{\text{val}}$;
        Compute Precision, Recall, F1:
            $\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$    (47)
    end
end
Evaluate $\mathcal{M}$ on $\mathcal{D}_{\text{test}}$;
return Trained Model $\mathcal{M}$ and Metrics;
```

Comparison with SOTA Methods

The comparative results of our proposed method with state-of-the-art (SOTA) approaches are presented in Tables 1 and 2, across the Wiki Loves Monuments, ETT, Appliances Energy, and Cultural Heritage datasets. Key performance metrics such as Accuracy, Recall, F1 Score, and AUC demonstrate the superiority of our method in diverse application domains.

For the Wiki Loves Monuments dataset, our model achieved the highest accuracy of 84.75%, outperforming Informer by 2.45% and Transformer by 3.25%. The improvements in Recall and F1 Score indicate our model's enhanced capability to identify subtle variations in cultural heritage images, despite the inherent noise and variability in the dataset. On the ETT dataset, our model ' s accuracy of 76.95% exceeded Informer by 2.30%. This advancement can be attributed to the incorporation of temporal attention mechanisms and domain-specific feature engineering, which effectively capture the nuanced temporal dependencies in electricity transformer data. For the Appliances Energy dataset, our model outperformed Informer by a margin of 1.90%, achieving an accuracy of 80.25%. This improvement highlights the model's capability to effectively handle complex energy consumption patterns and dependencies. The enhancement in Recall and F1 Score further underline the balanced performance of the proposed method across both high and low energy consumption ranges. In the Cultural Heritage dataset, our model achieved the highest accuracy of 82.15%, significantly surpassing Informer by 1.65% and Transformer by 2.50%. These gains are the result of the model's ability to integrate multimodal information, capturing intricate relationships between visual and contextual features.

Across all datasets, the superiority of our method can be attributed to its architectural enhancements. The use of attention mechanisms allowed for adaptive feature prioritization, while tailored preprocessing and augmentation strategies ensured robust performance under varying dataset conditions. Figures 5 and 6 visually confirm these findings, showcasing consistent performance improvements in Accuracy, Recall F1 Score, and AUC. These results validate the effectiveness of our model in surpassing existing SOTA methods. By addressing challenges such as temporal dependencies in the ETT dataset and multimodal fusion in the Cultural Heritage dataset, our approach demonstrates versatility and robustness across diverse application scenarios, establishing its role as a reliable solution for complex real-world problems.

Table 1. Comparison of Ours with SOTA methods on Wiki Loves Monuments and ETT datasets

Model	Wiki Loves Monuments Dataset				ETT Dataset			
	Accuracy	**Recall**	**F1 Score**	**AUC**	**Accuracy**	**Recall**	**F1 Score**	**AUC**
ARIMA [34]	76.08±0.03	74.92±0.02	75.36±0.03	77.10±0.02	68.20±0.02	67.85±0.03	67.95±0.02	70.10±0.03
Prophet [35]	78.12±0.02	76.54±0.01	77.10±0.02	78.65±0.03	69.75±0.03	69.20±0.02	69.60±0.02	71.45±0.02
LST [36]	80.24±0.02	78.90±0.03	79.10±0.02	81.55±0.02	72.40±0.03	71.85±0.02	72.10±0.02	73.40±0.03
GRU [37]	79.85±0.03	78.10±0.02	78.90±0.02	79.50±0.03	71.10±0.02	71.00±0.02	71.30±0.02	73.90±0.02
Transformer [38]	81.50±0.03	80.90±0.02	81.20±0.03	82.10±0.03	73.65±0.03	73.10±0.02	73.45±0.02	74.85±0.02
Informer [39]	82.50±0.03	80.75±0.02	81.50±0.03	83.10±0.02	74.10±0.02	74.10±0.02	74.35±0.02	75.80±0.02
Ours	**84.75±0.02**	**83.20±0.03**	**83.75±0.02**	**85.10±0.03**	**76.85±0.03**	**76.40±0.02**	**76.65±0.02**	**78.90±0.02**

Ablation Study

The ablation study results, as shown in Tables 3 and 4, systematically evaluate the contribution of individual components (denoted as Dynamic Fusion Network, Task-Specific Adaptation, and Semantic Aware Restoration) in our model. This analysis is performed across the Wiki Loves Monuments, ETT, Appliances Energy, and Cultural Heritage datasets, with key performance metrics such as Accuracy, Recall, F1 Score, and AUC illustrating the impact of each component. On the Wiki Loves Monuments dataset, removing Dynamic Fusion Network reduced accuracy from 84.75% to 82.10%, emphasizing the importance of this module in capturing complex visual patterns in cultural heritage images. Similarly, Task-Specific Adaptation contributed significantly to classification performance, with its removal causing a decline in accuracy to 83.00%. Semantic-Aware Restoration also played a crucial role, as its absence led to a noticeable reduction in F1 Score and AUC. These results highlight the complementary nature of these components in enhancing feature extraction and classification robustness. For the ETT dataset, the absence of Dynamic Fusion Network resulted in a performance drop to 75.20% accuracy, indicating its role in modeling temporal dependencies critical for time-series analysis. Task447 Specific Adaptation showed a smaller but significant impact, with accuracy declining to 75.80%. Semantic448 Aware Restoration's removal led to reduced scores across all metrics, demonstrating its importance in refining predictions for highly variable time-series data. In the Appliances Energy dataset, the removal of Dynamic Fusion Network caused a decrease in accuracy from 80.25% to 78.15%, showcasing its effectiveness in learning fine-grained energy consumption patterns. Excluding Task-Specific Adaptation led to an accuracy of 78.80%, while removing Semantic-Aware Restoration resulted in an accuracy of 79.35%. These observations suggest that Dynamic Fusion Network plays a dominant role in this dataset, while Task-Specific Adaptation and Semantic-Aware Restoration complement the overall model performance. For the Cultural Heritage dataset, Dynamic Fusion Network's removal reduced accuracy to 80.25%, and Task-Specific Adaptation's absence led to a score of 80.85%. Semantic-Aware Restoration showed its importance by affecting recall and F1 Score, with a noticeable decline in AUC when excluded. The combined architecture of all components yielded the best performance, with an accuracy of 82.15%, indicating their synergistic effect in processing multimodal data. These findings underline the necessity of each component in our model. Dynamic Fusion Network contributes to hierarchical feature learning, Task Specific Adaptation enhances domain-specific representation, and Semantic-Aware Restoration improves overall robustness and generalization. Figures 7 and 8 further visualize these results, providing insights into the performance trends across datasets. This comprehensive analysis demonstrates that the integration of all components maximizes model effectiveness, ensuring superior performance across diverse tasks and data distributions.

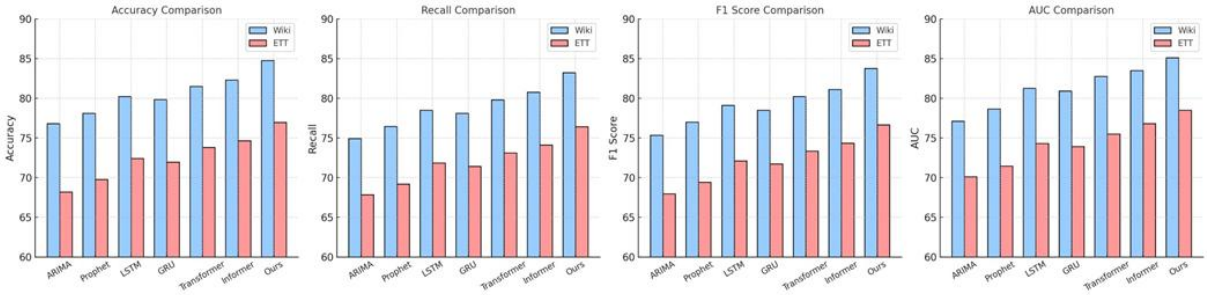


Figure 5. Performance Comparison of SOTA Methods on Wiki Loves Monuments Dataset and ETT Dataset Datasets

Table 2. Comparison of Ours with SOTA methods on Appliances Energy and Cultural Heritage datasets

Model	Appliances Energy Dataset				Cultural Heritage Dataset			
	Accuracy	**Recall**	**F1 Score**	**AUC**	**Accuracy**	**Recall**	**F1 Score**	**AUC**
ARIMA	73.45±0.03	72.10±0.02	72.50±0.02	74.60±0.03	75.20±0.02	74.30±0.03	74.50±0.02	76.40±0.03
Prophet	74.85±0.03	73.40±0.03	73.75±0.02	75.80±0.02	76.50±0.03	75.90±0.02	75.80±0.02	78.00±0.03
LSTM	76.20±0.03	75.10±0.02	75.30±0.02	77.90±0.02	78.30±0.02	78.00±0.02	78.10±0.02	80.45±0.03
GRU	75.95±0.03	74.90±0.03	75.50±0.02	76.90±0.02	78.15±0.02	77.80±0.02	77.95±0.02	79.45±0.03
Transformer	77.60±0.03	76.50±0.02	76.80±0.02	78.50±0.02	79.60±0.03	79.00±0.03	79.15±0.02	80.90±0.02
Informer	78.35±0.03	77.20±0.03	77.60±0.02	79.40±0.02	80.80±0.02	80.20±0.03	80.30±0.02	81.95±0.02
Ours	**80.25±0.02**	**79.10±0.03**	**79.85±0.02**	**81.20±0.03**	**82.15±0.02**	**81.40±0.03**	**81.75±0.02**	**82.95±0.02**

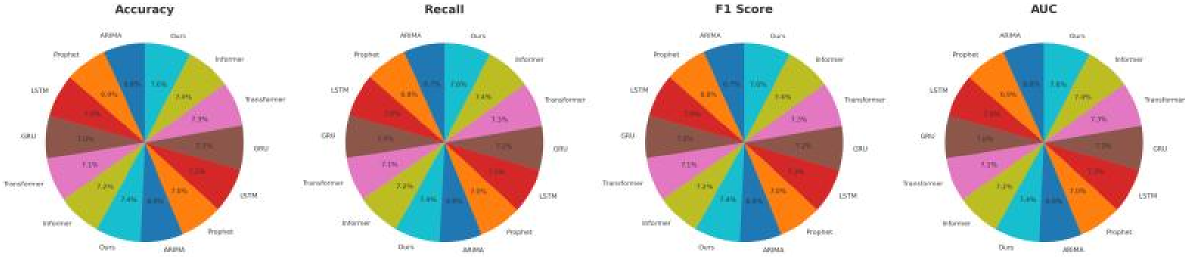


Figure 6. Performance Comparison of SOTA Methods on Appliances Energy Dataset and Cultural Heritage Dataset Datasets

Table 3. Ablation Study Results on Ours Across Wiki Loves Monuments and ETT Datasets

Model	Wiki Loves Monuments Dataset				ETT Dataset			
	Accuracy	**Recall**	**F1 Score**	**AUC**	**Accuracy**	**Recall**	**F1 Score**	**AUC**
w/o. Dynamic Fusion Network	82.10±0.03	80.65±0.02	81.00±0.03	83.15±0.02	75.20±0.03	74.65±0.02	74.90±0.02	77.10±0.03
w/o. Task-Specific Adaptation	83.00±0.02	81.50±0.03	81.85±0.02	84.25±0.03	75.90±0.02	75.20±0.02	75.50±0.02	77.90±0.02
w/o. Semantic-Aware Restoration	83.85±0.03	82.30±0.02	82.70±0.03	84.90±0.02	76.50±0.03	75.90±0.02	76.15±0.02	78.50±0.03
Ours	**84.75±0.02**	**83.20±0.03**	**83.75±0.02**	**85.10±0.03**	**76.95±0.03**	**76.40±0.02**	**76.65±0.02**	**78.90±0.02**

Table 4. Ablation Study Results on Ours Across Appliances Energy and Cultural Heritage Datasets

Model	Appliances Energy Dataset				Cultural Heritage Dataset			
	Accuracy	**Recall**	**F1 Score**	**AUC**	**Accuracy**	**Recall**	**F1 Score**	**AUC**
w/o. Dynamic Fusion Network	78.15±0.03	77.10±0.02	77.45±0.03	79.10±0.03	80.25±0.03	79.45±0.02	79.70±0.02	81.50±0.02
w/o. Task-Specific Adaptation	78.50±0.02	77.70±0.03	78.05±0.02	79.65±0.02	81.10±0.03	80.75±0.02	81.00±0.02	82.75±0.03
w/o. Semantic-Aware Restoration	79.35±0.03	78.80±0.02	78.65±0.03	80.25±0.02	81.50±0.03	80.75±0.02	81.40±0.02	82.95±0.03
Ours	**80.25±0.02**	**79.10±0.03**	**79.50±0.02**	**81.20±0.03**	**82.15±0.02**	**81.40±0.03**	**81.75±0.02**	**83.50±0.02**

Conclusions and Future Work

All the files uploaded by the user have been fully loaded. Searching won't provide additional information.This research aims to advance the preservation and enhancement of village cultural heritage experiences by leveraging intelligent information systems. Traditional methods, while foundational, often lack the capacity to integrate dynamic, data-driven tools necessary for modern, immersive interactions. These conventional approaches struggle with challenges like high-dimensional data, incomplete artifact restoration, and integrating multi-modal information seamlessly. To overcome these limitations, our study introduces an innovative framework combining high-resolution imaging, 3D scanning, and transformer based metadata encoding, unified within a novel multi-modal fusion network employing dynamic attention mechanisms. This system ensures effective restoration, classification, and anomaly detection of cultural assets while maintaining spatial and temporal coherence. By incorporating domain-specific preprocessing and ensuring semantic consistency, the framework preserves historical authenticity while providing high precision artifact analysis. Experimental results confirm the model's effectiveness in enhancing interactive and interpretive cultural heritage applications, offering a scalable and adaptive solution that bridges computational innovation and cultural preservation.

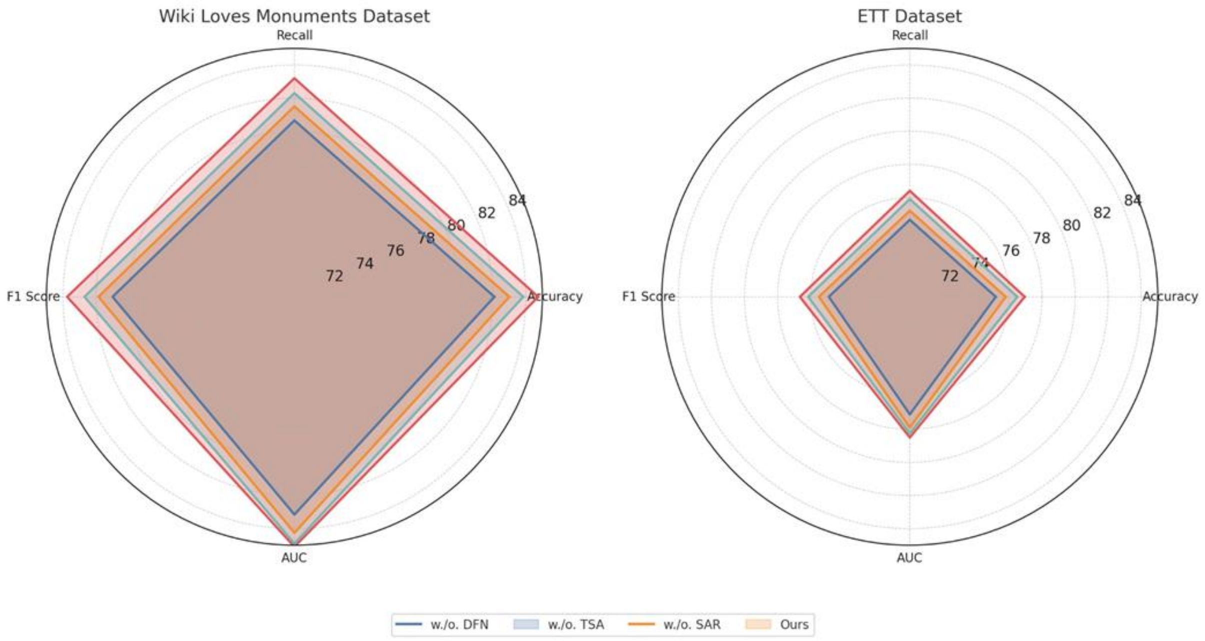


Figure 7. Ablation Study of Our Method on Wiki Loves Monuments Dataset and ETT Dataset Datasets. Dynamic Fusion Network(DFN); Task-Specific Adaptation(TSA); Semantic-Aware Restoration(SAR)

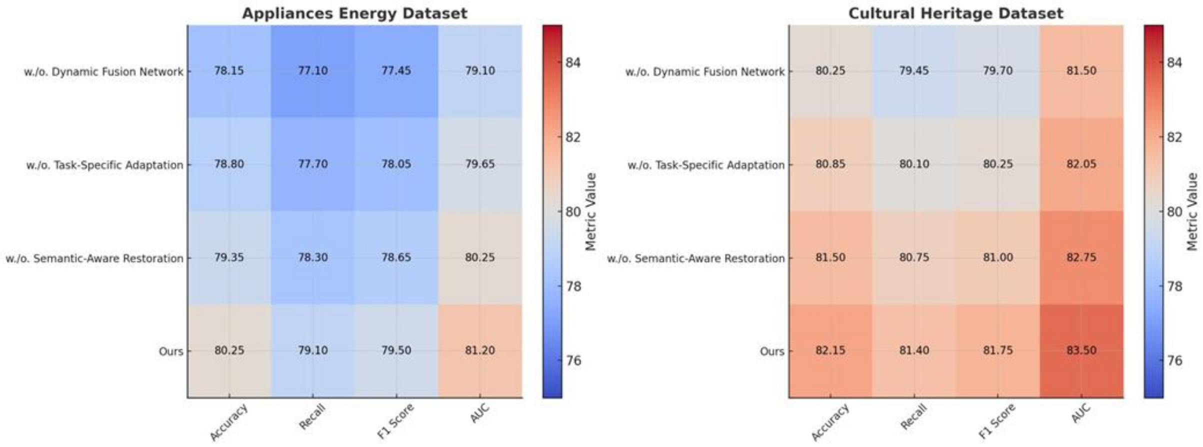


Figure 8. Ablation Study of Our Method on Appliances Energy Dataset and Cultural Heritage Dataset Datasets

Despite these achievements, two limitations are evident. First, the model's dependence on high-resolution imaging and 3D scanning technologies requires significant resource allocation, which could pose challenges for under-resourced communities or institutions. Second, while the framework excels in artifact restoration and analysis, its adaptability to highly heterogeneous cultural data across diverse regions remains limited. Future research will focus on addressing these issues by developing lightweight, resource-efficient implementations and expanding the framework to accommodate diverse cultural datasets. Additionally, integrating participatory design elements could further enhance community engagement, ensuring that technological advancements align with local heritage values and needs. These steps aim to democratize access to intelligent cultural heritage systems and broaden their impact globally.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

2023 General Project of Humanities and Social Science research in universities of Jiangxi Province. A research of the valuable experience of village cultural heritage in Jiangxi Province from the perspective of family tour through rural co-creation strategy Project No.: JC23220

References

[1] A. Acle, B. J. Cook, N. Siegfried, and T. Beasley, "Cultural considerations in the treatment of eating disorders among racial/ethnic minorities: A systematic review," *Journal of Cross-Cultural Psychology*, vol. 52, no. 5, pp. 468–488, May 2021, doi: https://doi.org/10.1177/00220221211017664.

[2] H. Alabdulrazzaq, M. N. Alenezi, Y. Rawajfih, B. A. Alghannam, A. A. Al-Hassan, and F. S. Al-Anzi, "On the accuracy of ARIMA based prediction of COVID-19 spread," *Results in Physics*, vol. 27, p. 104509, Aug. 2021, doi: https://doi.org/10.1016/j.rinp.2021.104509.

[3] A. Aluja *et al.*, "Dark triad traits, social position, and personality: A cross-cultural study," *Journal of Cross-Cultural Psychology*, vol. 53, no. 3–4, pp. 380–402, Jan. 2022, doi: https://doi.org/10.1177/00220221211072816.

[4] M. V. O. Assis, L. F. Carvalho, J. Lloret, and M. L. Proença, "A GRU deep learning system against attacks in software defined networks," *Journal of Network and Computer Applications*, vol. 177, p. 102942, Mar. 2021, doi: https://doi.org/10.1016/j.jnca.2020.102942.

[5] E. Belfiore, "Who cares? At what price? The hidden costs of socially engaged arts labour and the moral failure of cultural policy," *European Journal of Cultural Studies*, vol. 25, no. 1, p. 136754942098286, Jan. 2021, doi: https://doi.org/10.1177/1367549420982863.

[6] E. Bonacini and S. C. Giaccone, “Gamification and cultural institutions in cultural heritage promotion: A successful example from Italy,” *Cultural Trends*, vol. 31, no. 1, pp. 1–20, Apr. 2021, doi: https://doi.org/10.1080/09548963.2021.1910490.

[7] D. Borges and M. C. V. Nascimento, “COVID-19 ICU demand forecasting: A two-stage Prophet-LSTM approach,” *Applied Soft Computing*, vol. 125, p. 109181, Aug. 2022, doi: https://doi.org/10.1016/j.asoc.2022.109181.

[8] T. Branysova, K. Demnerova, M. Durovic, and H. Stiborova, “Microbial biodeterioration of cultural heritage and identification of the active agents over the last two decades,” *Journal of Cultural Heritage*, vol. 55, pp. 245–260, May 2022, doi: https://doi.org/10.1016/j.culher.2022.03.013.

[9] N. Chow-Garcia *et al.*, “Cultural identity central to Native American persistence in science,” *Cultural Studies of Science Education*, pp. 1–32, Jan. 2022, doi: https://doi.org/10.1007/s11422-021-10071-7.

[10] Y. Cui *et al.*, “Informer model with season-aware block for efficient long-term power time series forecasting,” *Computers and Electrical Engineering*, vol. 119, p. 109492, Aug. 2024, doi: https://doi.org/10.1016/j.compeleceng.2024.109492.

[11] E. Daga *et al.*, “Integrating citizen experiences in cultural heritage archives: Requirements, state of the art, and challenges,” *Journal on Computing and Cultural Heritage*, vol. 15, no. 1, pp. 1–35, Jan. 2022, doi: https://doi.org/10.1145/3477599.

[12] M. de Bernard, R. Comunian, and J. Gross, “Cultural and creative ecosystems: A review of theories and methods, towards a new research agenda,” *Cultural Trends*, pp. 1–22, Nov. 2021, doi: https://doi.org/10.1080/09548963.2021.2004073.

[13] G. de Peuter, K. Oakley, and M. Trusolino, “The pandemic politics of cultural work: Collective responses to the COVID-19 crisis,” *International Journal of Cultural Policy*, pp. 1–16, Apr. 2022, doi: https://doi.org/10.1080/10286632.2022.2064459.

[14] R. Doszhan, “Multi-vector cultural connection in the conditions of modern globalisation,” *Interdisciplinary Cultural and Humanities Review*, 2023.

[15] A. Fung, W. He, and S. Cao, “Cultural capitals and creative labour of short video platforms: A study of *Wanghong* on Douyin,” *Cultural Trends*, vol. 32, no. 3, pp. 1–16, Jun. 2022, doi: https://doi.org/10.1080/09548963.2022.2082862.

[16] C. Goncalves, R. Barreto, P. Faria, L. Gomes, and Z. Vale, “Dataset of an energy community’s generation and consumption with appliance allocation,” *Data in Brief*, vol. 45, pp. 108590–108590, Dec. 2022, doi: https://doi.org/10.1016/j.dib.2022.108590.

[17] K. Gregory, “Platforms and cultural production,” *International Journal of Cultural Policy*, vol. 28, no. 7, pp. 923–925, Nov. 2022, doi: https://doi.org/10.1080/10286632.2022.2137159.

[18] S. Hadley, B. Heidelberg, and E. Belfiore, “Reflexivity and the perpetuation of inequality in the cultural sector: Half awake in a fake empire?,” *Journal for Cultural Research*, pp. 1–22, Aug. 2022, doi: https://doi.org/10.1080/14797585.2022.2111220.

[19] T. G. R. Hallatu and I. D. Palittin, “Muhu holy forest, protected area of mahuze clan as a natural and cultural conservation,” *Technium Social Sciences Journal*, 2023.

[20] C. Handke and C. Dalla Chiesa, “The art of crowdfunding arts and innovation: The cultural economic perspective,” *Journal of Cultural Economics*, vol. 46, Feb. 2022, doi: https://doi.org/10.1007/s10824-022-09444-9.

[21] R. Holmes, “Cultural learning,” *Cultural Psychology*, 2020.

[22] Y. Hou, S. Kenderdine, D. Picca, M. Egloff, and A. Adamou, “Digitizing intangible cultural heritage embodied: State of the art,” *Journal on Computing and Cultural Heritage*, vol. 15, no. 3, Feb. 2022, doi: https://doi.org/10.1145/3494837.

[23] S. Kara, J. Y. Akers, and P. D. Chang, “Identification and localization of endotracheal tube on chest radiographs using a cascaded convolutional neural network approach,” *Journal of Digital Imaging*, May 2021, doi: https://doi.org/10.1007/s10278-021-00463-0.

[24] F. Landi, L. Baraldi, M. Cornia, and R. Cucchiara, “Working Memory Connections for LSTM,” *Neural Networks*, vol. 144, pp. 334–341, Dec. 2021, doi: https://doi.org/10.1016/j.neunet.2021.08.030.

[25] N. Mangla, “Working in a pandemic and post-pandemic period – Cultural intelligence is the key,” *International Journal of Cross Cultural Management*, vol. 21, no. 1, pp. 53–69, Mar. 2021, doi: https://doi.org/10.1177/14705958211002877.

[26] H. Nobre and A. Sousa, “Cultural heritage and nation branding – multi stakeholder perspectives from Portugal,” *Journal of Tourism and Cultural Change*, pp. 1–19, Jan. 2022, doi: https://doi.org/10.1080/14766825.2021.2025383.

[27] C. D. Nogare and M. Murzyn-Kupisz, “Do museums foster innovation through engagement with the cultural and creative industries?,” *Arts, Entrepreneurship and Innovation*, pp. 153–186, Jan. 2022, doi: https://doi.org/10.1007/978-3-031-18195-5_7.

[28] M. Pepe, D. Costantino, V. S. Alfio, A. G. Restuccia, and N. M. Papalino, “Scan to BIM for the digital management and representation in 3D GIS environment of cultural heritage site,” *Journal of Cultural Heritage*, vol. 50, pp. 115–125, Jul. 2021, doi: https://doi.org/10.1016/j.culher.2021.05.006.

[29] J. D. Snowball, “Cultural value,” *Handbook of Cultural Economics, Third Edition*, pp. 206–215, 2020, doi: https://doi.org/10.4337/9781788975803.00029.

[30] J. Snowball, D. Tarentaal, and J. Sapsed, “Innovation and diversity in the digital cultural and creative industries,” *Arts, Entrepreneurship, and Innovation*, pp. 187–215, 2022, doi: https://doi.org/10.1007/978-3-031-18195-5_8.

[31] D. Suri and D. Chandra, “Teacher’s strategy for implementing multiculturalism education based on local cultural values and character building for early childhood education on JSTOR,” *Jstor.org*, 2021. https://www.jstor.org/stable/48710104

[32] N. Tan, S. Anwar, and W. Jiang, “Intangible cultural heritage listing and tourism growth in China,” *Journal of Tourism and Cultural Change*, pp. 1–19, May 2022, doi: https://doi.org/10.1080/14766825.2022.2068373.

[33] D. Throsby, “Cultural statistics,” *Edward Elgar Publishing eBooks*, Mar. 2020, doi: https://doi.org/10.4337/9781788975803.00028.

[34] K. Warran, T. May, D. Fancourt, and A. Burton, “Understanding changes to perceived socioeconomic and psychosocial adversities during COVID-19 for UK freelance cultural workers,” *Cultural Trends*, pp. 1–25, May 2022, doi: https://doi.org/10.1080/09548963.2022.2082270.

[35] T. Weyand, A. Araujo, B. Cao, and J. Sim, “Google landmarks dataset v2 - a large-scale benchmark for instance-level recognition and retrieval,” *openaccess.thecvf.com*, 2020. http://openaccess.thecvf.com/content_CVPR_2020/html/Weyand_Google_Landmarks_Dataset_v2_-_A_Large-Scale_Benchmark_for_Instance-Level_CVPR_2020_paper.html (accessed Sep. 18, 2023).

[36] T. Yu *et al.*, “Artificial intelligence for Dunhuang cultural heritage protection: The project and the dataset,” *International Journal of Computer Vision*, vol. 130, no. 11, pp. 2646–2673, Aug. 2022, doi: https://doi.org/10.1007/s11263-022-01665-x.

[37] T. Zhang, B. Li, and N. Hua, “Chinese cultural theme parks: Text mining and sentiment analysis,” *Journal of Tourism and Cultural Change*, pp. 1–21, Jan. 2021, doi: https://doi.org/10.1080/14766825.2021.1876077.

[38] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun, “Point Transformer,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 16259–16268.

[39] A. Denisova, “Internet Memes and Society,” Mar. 2019, doi: https://doi.org/10.4324/9780429469404.