AI-Enhanced Multimodal Computational Modeling and Cross-Media Analytics for the Living Transmission of Fish-Skin Painting

Zhaoli Wang*, Lili Zhang, Jiahui Wang, Na Li

Harbin Institute of Information Technology
No. 9, University Town, Binxi Town, Harbin City, Heilongjiang Province, China

*hxciwangzhaoli@126.com

Abstract. Fish-skin painting is an important form of Hezhe intangible cultural heritage, yet its preservation faces challenges such as fragile materials, tacit manual skills, limited teaching resources, and geographically constrained exhibitions. This paper proposes an AI-assisted multimodal reconstruction framework integrating high-resolution image acquisition, 3D surface recording, process-video annotation, feature encoding, skill-comparison modeling, and cross-media deployment. Based on the Fuyuan case, the study analyzes more than 200 digitized works, records of 5 inheritors, over 2,000 student participants, 100,000+ annual visitors, 500,000+ livestream viewers, 80+ cultural creative products, and 50+ study-tour groups. Indicator calculation shows that the ratios of student participants to digitized works, livestream viewers to offline visitors, and cultural creative products to digitized works are about 10:1, 5:1, and 0.4:1, respectively. The results indicate that computational reconstruction can transform fragile craft objects and tacit knowledge into searchable, measurable, teachable, and reusable digital resources for the living transmission of fish-skin painting.

Keywords: fish-skin painting; intangible cultural heritage; computational reconstruction; AI-assisted comparison; multimodal representation; cross-media communication.

1 Introduction

Intangible cultural heritage is maintained through continuous practice, teaching, interpretation, and public participation[1]. For fish-skin painting, traditional preservation methods based on exhibition, oral instruction, and physical storage are insufficient because fish skin is fragile, sensitive to humidity and fading, and difficult to teach through text alone. The craft process, including material selection, tanning, cutting, collage, sewing, and finishing, depends heavily on embodied knowledge and long-term apprenticeship.

Existing studies have discussed image archiving, 3D reconstruction, digital storytelling, and online dissemination in intangible cultural heritage protection[2-4]. However,

R. Zur et al. (eds.), *Proceedings of the 2026 5th International Conference on Art Design and Digital Technology (ADDT 2026)*, Advances in Computer Science Research 131,
https://doi.org/10.2991/978-94-6239-737-8_37

many projects still treat digitalization as static storage rather than computational representation. For computer-related research, the key question is not only how to digitize fish-skin painting, but how to encode, compare, retrieve, evaluate, and reuse its visual and procedural features.

This paper therefore proposes an AI-assisted reconstruction model. The main contributions are as follows:

1. A multimodal data model is built for fish-skin painting, integrating image, 3D, video, metadata, and cultural-semantic information.
2. A feature representation method is proposed to encode texture, color, contour, motif, and process information.
3. An AI-assisted skill evaluation model is designed to compare learner works with reference works.
4. A case-based evaluation is conducted using Fuyuan fish-skin painting data.

2 Multimodal Data Representation and Reconstruction Framework

Fish-skin painting contains material texture, manual traces, symbolic motifs, and process knowledge. Therefore, a single image archive cannot fully represent its cultural and technical value. As shown in Fig. 1, this study constructs a multimodal reconstruction framework covering data acquisition, feature extraction, structured storage, skill comparison, and cross-media reuse.

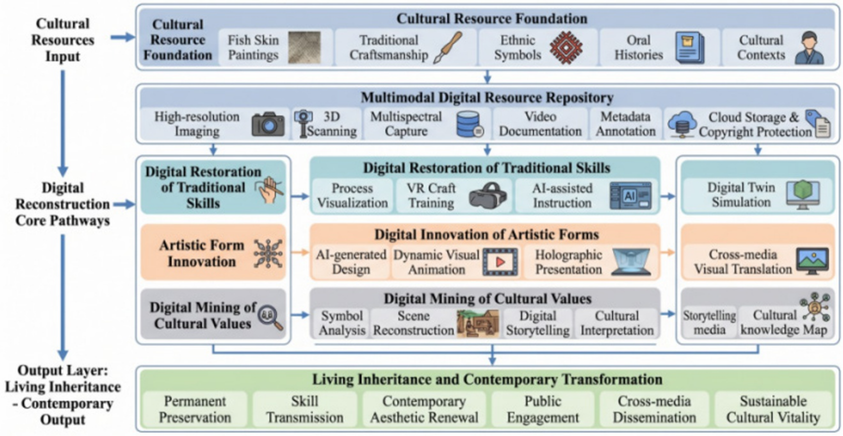

Fig. 1. Digital Reconstruction Framework of Fish Skin Painting.

Each digital object is defined as a multimodal unit:

$$O_i = \{I_i, P_i, V_i, M_i, S_i\}$$

where I_i denotes high-resolution image data, P_i denotes 3D point-cloud or surface-relief data, V_i denotes process-video data, M_i denotes metadata, and S_i denotes semantic cultural annotation.

For computational processing, each work is further encoded as a feature vector:

$$F_i = [T_i, C_i, G_i, R_i, Q_i]$$

where T_i represents texture features, C_i color features, G_i contour and geometric composition, R_i motif category, and Q_i process-stage features. This structure allows fish-skin painting resources to support retrieval, classification, similarity comparison, teaching feedback, and derivative design.

In the Fuyuan case, the dataset includes more than 200 representative works and process records from 5 inheritors. These data provide the basis for computational modeling, indicator calculation, and heritage teaching.

3 AI-Assisted Feature Extraction

The feature extraction layer converts visual and procedural information into machine-readable representations. Texture direction and roughness can be extracted using Gabor filters, local binary patterns, or convolutional neural networks. Color relationships can be represented in HSV or Lab color space. Contour and composition can be extracted by edge detection, segmentation models, or vision transformers[5]. Motif categories can be identified using image classification models or image-text models such as CLIP[6], while object and region segmentation can be supported by large segmentation models such as SAM[7]. For 3D surface data, point-cloud models such as PointNet can be used to describe relief and material-layering features[8-10].
A simplified extraction process is shown below:

Algorithm 1. Multimodal Feature Extraction for Fish-Skin Painting

Input:	image I, 3D data P, process video V
Output:	feature vector F
1.	Segment fish-skin painting region from I
2.	Extract texture feature T from local patches
3.	Extract color feature C in Lab/HSV space
4.	Extract contour and layout feature G
5.	Classify motif category R
6.	Annotate process sequence Q from video frames
7.	Return F = [T, C, G, R, Q]

Through this process, fish-skin painting changes from isolated craft objects into structured computational resources.

4 AI-Assisted Skill Comparison Model

Traditional craft teaching depends strongly on inheritors' experience. To support early-stage learning, this paper introduces an AI-assisted comparison score between learner work L and reference work R:

$$D(L,R) = w_t D_t + w_c D_c + w_g D_g + w_r D_r + w_q D_q$$

where D_t, D_c, D_g, D_r, and D_q represent differences in texture, color, contour, motif, and process sequence. The weights w_t, w_c, w_g, w_r, w_q can be adjusted by inheritors according to teaching priorities.

For process comparison, dynamic time warping can be used to compare learner operation sequences with expert demonstrations:

$$D_q = DTW(Q_L, Q_R)$$

The model does not replace human judgment. Instead, it provides interpretable feedback such as texture-direction deviation, edge-contour mismatch, color-coordination difference, motif-category inconsistency, or process-order error. This can reduce the uncertainty of beginner practice and improve teaching efficiency.

In practical teaching, comparison results can generate graded feedback: texture deviation, contour deviation, color deviation, motif deviation, and process deviation . Thus, the model provides step-by-step diagnostic feedback beyond a single similarity score. This is crucial for intangible cultural heritage education, as beginners struggle to identify discrepancies from inheritors' work. By translating expert experience into interpretable indicators, the system supports repeated practice while leaving final evaluation to human inheritors.

5 Cross-Media Reuse and Evaluation

After reconstruction and feature encoding, digital resources can be reused in exhibitions, online communication, school education, cultural creative design, and tourism. As shown in Fig. 2, cross-media presentation forms a feedback loop: digital resources generate public contact, user participation produces behavioral data, and these data further optimize teaching, exhibition, annotation, and design reuse.

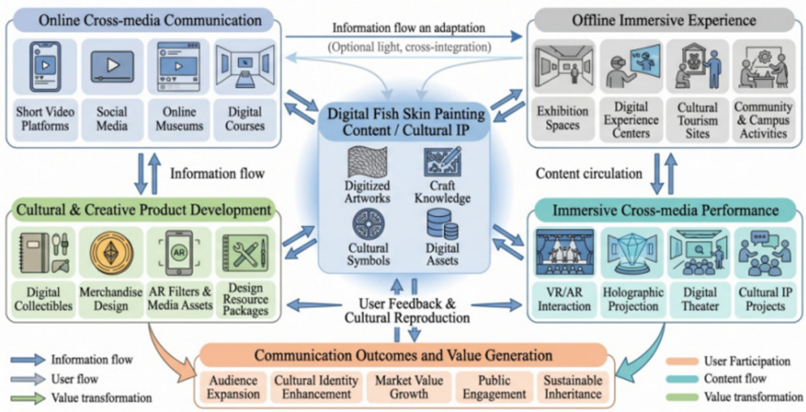

Fig. 2. Cross-media Communication Framework of Fish Skin Painting.

Compared with ordinary digital display, the computational framework supports evaluation-oriented reuse. The evaluation indicators are shown in Table 1.

Table 1. Evaluation indicators of Fuyuan fish-skin painting digitalization.

Dimension	Indicator	Case result
Digital preservation	Digitized representative works	200+
Digital preservation	Recorded inheritors	5
Skill transmission	Student participants	2,000+
Public dissemination	Annual visitors	100,000+
Public dissemination	Livestream viewers	500,000+
Value transformation	Cultural creative products	80+
Value transformation	Study tour groups	50+

Based on these data, several derived indicators can be calculated to connect the case statistics with the computational reconstruction framework. The ratio of student participants to digitized works is about 10:1, indicating that each digital work can support repeated teaching use. The ratio of livestream viewers to annual offline visitors is about 5:1, showing that online platforms significantly expand the communication range. The ratio of cultural creative products to digitized works is about 0.4:1, suggesting that reconstructed motifs and cultural resources have been partly transformed into reusable design assets. These indicators make the evaluation results more consistent with the paper's data-driven and model-based research design.

These results show that the Fuyuan project has moved beyond simple archiving. It forms a chain from data acquisition and feature representation to indicator calculation, teaching feedback, public dissemination, and cultural value transformation.

6 Discussion

From a computer-science perspective, the value of the proposed framework lies in three aspects.

First, it transforms fish-skin painting into multimodal cultural data. Image, 3D, video, and semantic annotations are integrated into a unified data structure, which supports later retrieval, classification, comparison, and reuse.

Second, it introduces interpretable feature modeling. Instead of using AI as a black-box generation tool, the framework connects computational features with craft concepts, such as texture direction, contour accuracy, motif category, and process sequence. This makes the model more suitable for heritage teaching.

Third, it links reconstruction with application evaluation. The framework does not stop at model construction but uses preservation, teaching, dissemination, and cultural-creative indicators to evaluate practical effects.

However, the current case data are still project-level statistics rather than controlled experimental results. Future work should build a labeled fish-skin painting dataset, conduct segmentation and classification experiments, and evaluate model performance using accuracy, precision, recall, F1-score, and user-study indicators.

7 Conclusion

This paper proposes an AI-assisted multimodal reconstruction framework for fish-skin painting. The framework represents fish-skin painting through image, 3D, video, metadata, and semantic annotations, and encodes texture, color, contour, motif, and process features. Based on this representation, an interpretable skill-comparison model supports teaching feedback and craft transmission, while reconstructed resources can be reused in exhibitions, education, online communication, cultural creative design, and tourism.

The Fuyuan case provides a practical data basis, including 200+ digitized works, 5 inheritor records, 2,000+ student participants, 100,000+ annual visitors, 500,000+ livestream viewers, 80+ cultural creative products, and 50+ study-tour groups. Derived indicators show ratios of about 10:1 for student participants to digitized works, 5:1 for livestream viewers to offline visitors, and 0.4:1 for cultural creative products to digitized works. The study demonstrates that AI-assisted methods can make fragile materials and tacit craft knowledge more preservable, measurable, teachable, and reusable while maintaining inheritors and cultural context as the core of living transmission.

Disclosure of Interests

The authors have no competing interests to declare.

References

1. UNESCO. Convention for the Safeguarding of the Intangible Cultural Heritage[S]. Paris: UNESCO, 2003. Available: https://ich.unesco.org/doc/src/01852-EN.pdf
2. Hou Y, Kenderdine S, Picca D, Egloff M, Adamou A. Digitizing Intangible Cultural Heritage Embodied: State of the Art[J]. Journal on Computing and Cultural Heritage, 2022, 15(3): 55:1–55:20. DOI: https://doi.org/10.1145/3494837
3. Skublewska-Paszkowska M, Milosz M, Powroznik P, Lukasik E. 3D Technologies for Intangible Cultural Heritage Preservation—Literature Review for Selected Databases[J]. Heritage Science, 2022, 10: 3. DOI: https://doi.org/10.1186/s40494-021-00633-x
4. Kim S, Im D U, Lee J, Choi H. Utility of Digital Technologies for the Sustainability of Intangible Cultural Heritage (ICH) in Korea[J]. Sustainability, 2019, 11(21): 6117. DOI: https://doi.org/10.3390/su11216117
5. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[C]//International Conference on Learning Representations. 2021. Available: https://openreview.net/forum?id=YicbFdNTTy
6. Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I. Learning Transferable Visual Models From Natural Language Supervision[C]//Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021, 139: 8748–8763. Available: https://proceedings.mlr.press/v139/radford21a.html

7. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg A C, Lo W Y, Dollár P, Girshick R. Segment Anything[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 4015–4026. DOI: https://doi.org/10.1109/ICCV51070.2023.00371
8. Qi C R, Su H, Mo K, Guibas L J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 77–85. DOI: https://doi.org/10.1109/CVPR.2017.16
9. Qi C R, Yi L, Su H, Guibas L J. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space[C]//Advances in Neural Information Processing Systems. 2017, 30. Available: https://papers.nips.cc/paper/7095-pointnet-deep-hierarchical-feature-learning-on-point-sets-in-a-metric-space
10. Wang Y, Sun Y, Liu Z, Sarma S E, Bronstein M M, Solomon J M. Dynamic Graph CNN for Learning on Point Clouds[J]. ACM Transactions on Graphics, 2019, 38(5): 1–12. DOI:https://doi.org/10.1145/3326362