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

A Research Regarding the Deep-Level Discussion of Non-Physical Cultural Heritage Components within Fine Art Creation

Jianlei Jia[1,*] *and Dan Huang*[2]

[1] School of Arts, Wuxi Taihu University, Wuxi 214064, Jiangsu, China
[2] School of Education, Guangdong Huali College, Jiangmen 529152, Guangzhou, China

SUMMARY: *When components of non-material cultural heritage walk into modern artistic creation, they often experience at the same time pattern recombination, color re-arrangement, the shifting of handicraft marks, and the movement of cultural meanings. Methods which only depend on image similarity or category identification have difficulty in steadily catching these deep translational connections. To address this issue, this paper constructs a source-work paired corpus based on source ICH images, images of creative works, and their associated textual descriptions, and proposes the H-DFA hierarchical deep feature analysis framework.This method establishes a comprehensive representation across three dimensions-form retention, color-craft coupling, and semantic transfer-and converges the final deep feature scores through creation-adaptive weighting to identify the degree of retention, translation intensity, and cultural consistency of ICH elements across different works. The experimental section focuses on unified data partitioning, comparison methods, and evaluation interfaces, comparing methods such as HOG-SVM, CLIP, MICMLF, and WuMKG-guided, and verifying the model's stability under occlusion perturbations and cross-scenario creative conditions.The results show that H-DFA achieves 90.8%, 89.9%, 0.844, and 0.810 for Accuracy, Macro-F1, mAP@10, and Spearman ρ, respectively, outperforming all comparison methods overall. Layer-by-layer ablation results indicate that semantic correction and creative adaptation weighting are critical components for improving ranking quality and expert consistency; Results under different scenarios show that this method can keep a leading level of performance in six creative work items: China traditional painting, decoration painting, picture drawing, wall picture design, electronic painting and mixed material art. This research makes clear that to build a stable corresponding relationship between ICH source elements and creative outcomes, and to arrange visual, technical, semantic information by using a layered deep feature framework, is able to more effectively support the quantitative analysis, creative assessment, digital utilization of ICH cultural elements within the creation of fine art.*

KEYWORDS: *Intangible cultural heritage elements; Fine art creation; Deep feature analysis; Multimodal corpus; Creative translation*

1 Introduction

In current fine art making, public showings, and cultural spreading activities, the method that ICH components are put into works has great change happened. Creators no longer deal solely with traditional patterns that can be directly copied, but rather with composite visual resources composed of pattern frameworks, color schemes, material textures, craftsmanship traces, and

*jiajianleiwxu@163.com
https://doi.org/10.65102/is2026760 Published: 30 April, 2026

cultural connotations. Exhibition institutions, university curricula, cultural and creative development platforms, and digital art communities are also continuously driving these resources into new creative contexts. Current requirements mainly centralize on two facets: firstly, the demand to maintain the cultural particularity of ICH in creative productions, avoiding problems such as unclear sources, meaning shift, and the reduction of craft marks after visual taking; second, the requirement for effectively transferring tradition components through media like Chinese ink wash picture, decoration picture, illustration, digital picture, and mixed media. Pursuant to the definition given in the UNESCO Convention for the Protection of Intangible Cultural Heritage, ICH includes practices, expressions, knowledge, skills, together with related objects and cultural spaces, and is handed down through many fields such as oral traditions, performing arts, social customs, knowledge of nature, and traditional handcrafts; the Convention also puts stress that confirming, writing records, studying, protecting, passing on, teaching, and bringing back to life are all necessary requirements for ICH to keep going [1, 2]. Therefore, it is extremely clear that the bringing of ICH into fine art creation has close relation with special paths of cultural cognition, knowledge passing-down, and creative transformation.

Under this background, digital study has clearly developed in the past years from "keeping data" to a stage that pays attention to "arranging knowledge, giving support to creation, and building interaction." Taking Suzhou Wu-Luo Weaving as the example, the concerned research has put visualization and digital continuable development as the core questions, exploring how traditional handcrafts can obtain new spreading channels and sustainable development conditions under the digital environment [3]. With urban cultural spreading and public use situations as the targets, Liu et al. brought deep learning into the interface making for Chongqing's ICH apps, directly placing conservation and spread requirements inside the smart city situation [4]. At the same time, visual analysis and interactive systems have been utilized by people for the quantitative observation and online comprehension of cultural objects, hence this enables ICH research to gradually complete transition from static display to an analytical framework which possesses the properties of searchable, comparable, and interpretable.

In the research directions which are more directly connected with the problems discussed in this paper, an obvious tendency is that the step-by-step substitution of single-mode visual identification by multi-mode expression. The authors Fan et al. put forward the MICMLF model, this model carries out joint modeling on ICH images and relevant text descriptions, and captures complementary information between images and text via multimodal attention mechanism and hierarchical integration [5, 6]. Wan et al. built the WuMKG multi-mode knowledge graph which is for Chinese painting and calligraphy work, they attempt to put together scattered images, words, and knowledge connections inside one single graph frame to give a basis for cross-mode search, question reply, and visible display [7]. The authors Xia and other persons, at the same time, on the level of curriculum design, has proved that digital technology and interdisciplinary workshops have the promotion function in intangible cultural heritage education and creative production [8]. These researches which are connected together together point out that the technology core of ICH research has already changed from object gathering and simple category dividing to the cooperation between pictures and words, knowledge model building, and all-round support which is made for special use situations. To the study of art making, this change is especially important, because the process of creation itself includes visual shapes, text explanation, and culture meanings.

However, there remains a significant gap between existing research and the "in-depth feature analysis of ICH cultural elements in artistic creation." The first category of research focuses on digital preservation, platform dissemination, or curriculum organization. While these studies illustrate how ICH is preserved, displayed, and adapted for educational purposes, they rarely address which core features are retained, which formal relationships are rewritten, or which cultural connotations are weakened when a specific ICH element is incorporated into a concrete creative work [9-12]. The second category of research focuses on image classification, cross-modal recognition, or knowledge graph organization, with methodological objectives centered on classification accuracy, retrieval performance, or knowledge connectivity, while insufficient attention is paid to the correspondence between "source elements" and "artistic works" [13-15]. The third category of research has begun to explore how elements of intangible cultural heritage can be integrated into contemporary design and the development of derivative products; however, the scope of such research often remains limited to product design or individual case studies. This category of research lacks a detailed analysis of the partial borrowing, cross-media reconfiguration, and transfer of symbolic meaning commonly found in artistic creation. [16].

This separation therefore brings about four straight problems. Firstly, the data arrangement ways do not match the real requirements of creation-oriented research. The majority of researches utilize one sole ICH image, one sole text description, or one sole kind of cultural object to serve as the basic unit, therefore it leads to the absence of stable matching between source objects and creative results. Therefore, it has difficulty to confirm the accurate origin of design patterns, color matchings, and craft traces which are appeared in the creation process. Second, the evaluation standards do not match the creation logic. Image classification work put stress on which category an image belongs to, while search work put stress on ordering that bases on similarity; however, artistic creation puts more emphasis on getting balance among keeping formal things, continuous meaning, and change for creation. Third, although current models have already started to add in textual descriptions, knowledge relations, or visual attention, they frequently put these information parts together into one single united feature, hence it is hard to clearly express the level-by-level relations among pattern structures, color techniques, and cultural semantics. Fourth, a great number of researches are more suitable for proving whether an object has been recognized or whether the system can work normally, but they cannot with reliability explain why a work is divided into high, medium, or low fusion.

Furthermore, the complexity that is contained within the course of putting ICH elements into fine art therefore determines that simple methods which depend on resemblance are not enough. A work can have very similar shape outlines to the original source part, but it can show big differences in color arrangements and symbolic meanings; On the opposite side, it may have passed through big visual re-understanding but still keep clear culture clues by means of craft marks, structure arrangement, and character carvings. If a person only makes comparison on the whole visual similarity, hence it is easy to mix up effective translation with superficial imitation; if a person only depends on the text that the creator made, or the subjective judgment that experts give, it is very easy to bring about a deviation of evaluation standards. Hence, this paper places emphasis on a question that has stronger operability: the way to build a steady corresponding relation between ICH source elements and artistic creation outcomes, and, on the basis of this corresponding relation, find out their deep-layer characteristics, carry out quantification on their fusion degree, and guarantee that model outcomes can be checked and approved through expert evaluation.

According to the above considerations, this thesis regards the "source element-artistic creation" matching as its starting point and establishes an overall research framework which centers on the deep analysis of ICH cultural elements in artistic creation. The research scope is limited to ICH categories with clear potential for visual translation, and ICH source images, creative images, and related textual descriptions are organized into a unified corpus. On this basis, a hierarchical deep feature analysis method is constructed across three dimensions-pattern retention, color-craft coupling, and semantic transfer consistency-ensuring that the methodological structure directly corresponds to the subsequent interpretation of results.

Empirical validation is then conducted regarding overall performance, layer-by-layer optimization, cross-scenario generalization, and expert agreement. The contributions of this paper are primarily reflected in three aspects: First, we constructed a paired corpus of ICH source works tailored for fine arts creation research; second, we proposed a tiered deep feature analysis framework that integrates form, craftsmanship, and semantics into a unified evaluation logic; third, we established an empirical interface capable of directly supporting creation analysis, course feedback, and digital curation, providing a more stable operational approach for quantitative research on the application of ICH elements in fine arts creation.

2 Methods

2.1 Construction and Hierarchical Annotation of ICH Source-Work Paired Corpora

The ICH source data used in this study was drawn from publicly available ICH text and image entries, open digital museum collections, publicly released exhibition catalogs, and thematic publications; the creative output data was sourced from publicly exhibited works, outcomes of university fine arts courses, and works on digital painting platforms accompanied by explicit creative descriptions.To ensure a high degree of alignment between the research subjects and the research topic, this study retained only twelve categories of ICH elements with significant potential for visual translation, including woodblock New Year prints, paper-cutting, embroidery, batik, blue-printed cloth, bamboo weaving, lacquer art, shadow puppetry, clay sculpture, kites, enamel decoration, and seal carving.After deduplication, clarity screening, segmentation, and manual verification, the initial sample was transformed into a unified corpus suitable for source-work matching. The composition of the data is shown in Table 1.

Table 1: Composition and Classification of the ICH Source-Work Pairing Corpus

Data Layer	Content	Quantity
Intangible Cultural Heritage Images	12 Types of Visual and Craft Objects	4,860
Creative Images	Traditional Chinese Painting, Decorative Painting, Illustration, Mural Design, Digital Painting, Mixed Media Works	1,620
Source Pairing Samples	Corresponding Units of Source Elements and Creative Parts	6,480
Text Units	Craft Descriptions, Inscriptions, Creative Explanations, Curatorial Summaries	9,740
Expert Consistency Review Samples	Used for Deep Feature Scoring Verification	2,160
Data Split	Training / Validation / Testing	3,888 / 1,296 / 1,296

In Table 1, source pictures and generated pictures do not have direct one-to-one correspondence; the things that actually get into the model training process are paired units which are composed of "source element detail-generated detail-text description." For preventing the drifting of annotation standards, this research has invited five experts who hold backgrounds in art design or intangible cultural heritage study to reach common opinion through two rounds of annotation work. The annotation contents included five kinds of information: pattern category, main color group, craft characteristics, symbolic meaning, and integration degree. The strength of fusion is divided into three grades-high, middle, low-for the after overall score checking and expert consistence research that we carry on next. The training set, verification set and test set are divided according to the 6:2:2 proportion, which ensures that works created by the same creator do not appear in different sets.

To ensure consistency between subsequent deep feature analysis and the creative logic, this study organizes the annotation system into three hierarchical levels. The first level focuses on visible structure, emphasizing the description of pattern outlines, repetitive patterns, local density, and compositional placement. The second level focuses on modes of expression, emphasizing the description of color group distribution, material texture, and craftsmanship traces. The third level focuses on semantic transfer, emphasizing the recording of the preservation, shift, and reinterpretation of original cultural meanings in the creative works.The data organization logic of the study is shown in Figure 1.

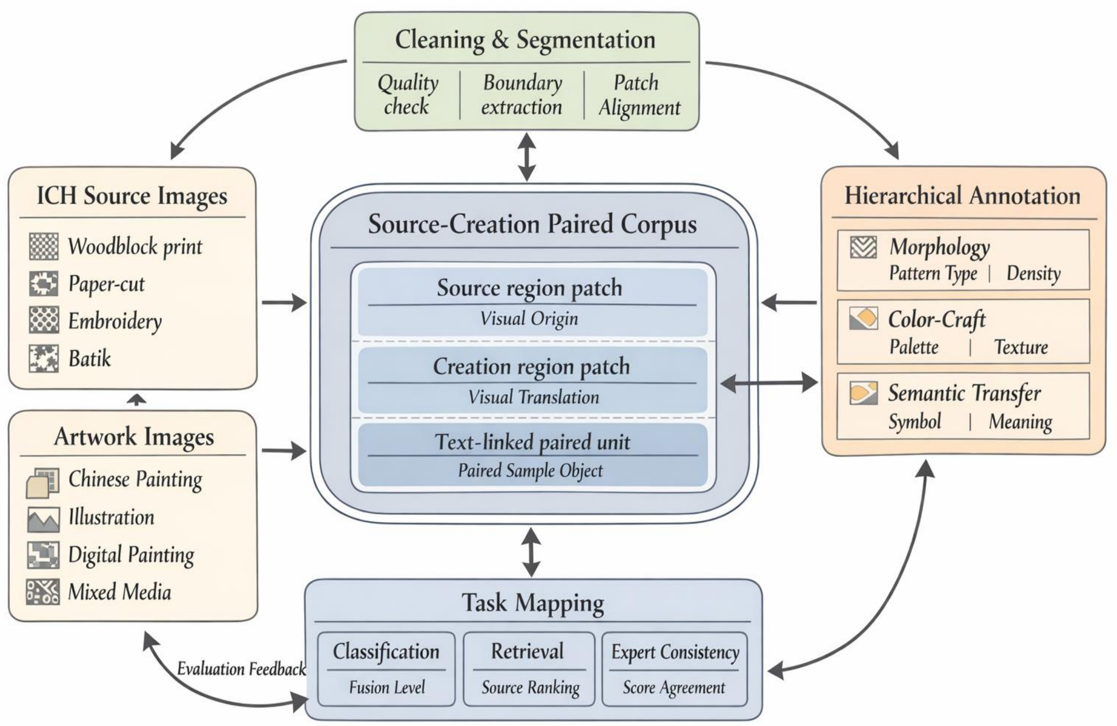


Figure 1: Diagram of the organizational mechanism for ICH source work samples.

2.2 H-DFA Hierarchical Deep Feature Analysis Framework

After having the finishing arrangement for all data, this paper furthermore builds up the Hierarchical Deep Characteristic Analysis (H-DFA) structural frame. This framework is developed step by step in three dimensions: firstly, it draws out shape-keeping information between original parts and newly-made particulars; second, it carries out calculation on the coupling degree between color and craftsmanship-related traces; and at last, it brings in textual meaning and knowledge connections to adjust the creative translation work. The results that are got in the foregoing layer are directly put into the weighted decision work of the following layer; Therefore, the optimization path of this model may be observed from its structure, hence it is not hidden inside a black box. The morphological retention score M_i is defined as shown in Equation (1).

$$M_i = \lambda_1 \cos(\mathbf{p}_i^s, \mathbf{p}_i^a) + \lambda_2 \frac{|E_i^s \cap E_i^a|}{|E_i^s \cup E_i^a|} + \lambda_3 (1 - |h_i^s - h_i^a|) \tag{1a}$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1 \tag{1b}$$

In the equation, $\mathbf{p}_i^s$ and $\mathbf{p}_i^a$ represent the pattern description vectors of the source element's local region and the created region in the th and$i$ th samples, respectively;E_i^s andE_i^a denote the sets of edge structures; and h_i^s and h_i^a denote the normalized structural entropies. This equation simultaneously incorporates global similarity, boundary consistency, and local complexity, ensuring that morphological judgment does not rely on a single visual distance. The color-craft coupling score C_i is defined as shown in Equation (2).

$$C_i = \eta_1 (1 - D_{\text{EMD}}(\mathbf{c}_i^s, \mathbf{c}_i^a)) + \eta_2 \cos(\mathbf{t}_i^s, \mathbf{t}_i^a) \tag{2a}$$

$$\eta_1 + \eta_2 = 1 \tag{2b}$$

In the equation, $\mathbf{c}_i^s$ and $\mathbf{c}_i^a$ represent the Lab color space distribution vectors of the source element and the creative region, respectively; D_{EMD} denotes the normalized Earth Mover's Distance; and $\mathbf{t}_i^s$ and $\mathbf{t}_i^a$ denote the process attribute embedding vectors. This formula incorporates both comprehensive hue similarity and consistency of process traces into the calculation, helping to distinguish between two significantly different creative approaches: "merely borrowing hues" and "preserving process syntax." The semantic transfer score S_i is shown in Equation (3).

$$S_i = \mu_1 \cos(\mathbf{w}_i^s, \mathbf{w}_i^a) + \mu_2 g_i + \mu_3 u_i \tag{3a}$$

$$\mu_1 + \mu_2 + \mu_3 = 1 \tag{3b}$$

In the equation, $\mathbf{w}_i^s$ and $\mathbf{w}_i^a$ represent the semantic vectors of the source description text and the generated description text, respectively; g_i represents the consistency of knowledge relationship paths; andu_i represents the alignment between the generated text and the local image region. The purpose of this layer is to correct samples that are "visually similar but culturally misaligned," enabling the model to distinguish between semantic retention and semantic reinterpretation. The comprehensive deep feature score D_i is shown in Equation (4).

$$D_i = \alpha M_i + \beta C_i + \gamma S_i \tag{4a}$$

$$\alpha + \beta + \gamma = 1 \tag{4b}$$

In the equation, D_i is the final deep feature fusion score, while α, β, and γ are determined through training on the validation set; in this paper, these values are set to 0.36, 0.29, and 0.35, respectively. When $D_i \geq 0.75$, the sample is classified as high fusion; when $0.55 \leq D_i < 0.75$, the sample is classified as medium fusion; and samples below 0.55 are classified as low fusion.

The innovations of H-DFA focus on three aspects. First, this paper positions the "source element-creative output" correspondence at the method's entry point, resolving the issue of separation between source and creative images found in existing research. Second, the three feature layers are not stacked in parallel; the morphological layer, color-craft layer, and semantic layer exert different weights on the final judgment, thereby preserving the fact that "visible structure" and "cultural semantics" operate in parallel in artistic creation. Finally, this paper introduces a creative adaptation weighting mechanism at the semantic level, enabling the model to identify translation forms with creative value, such as moderate modifications, cross-media reorganization, and the expansion of symbolic meaning. To demonstrate that the methodology of this study is derived from the gradual convergence of hierarchical features and creative adaptation, see Figure 2.

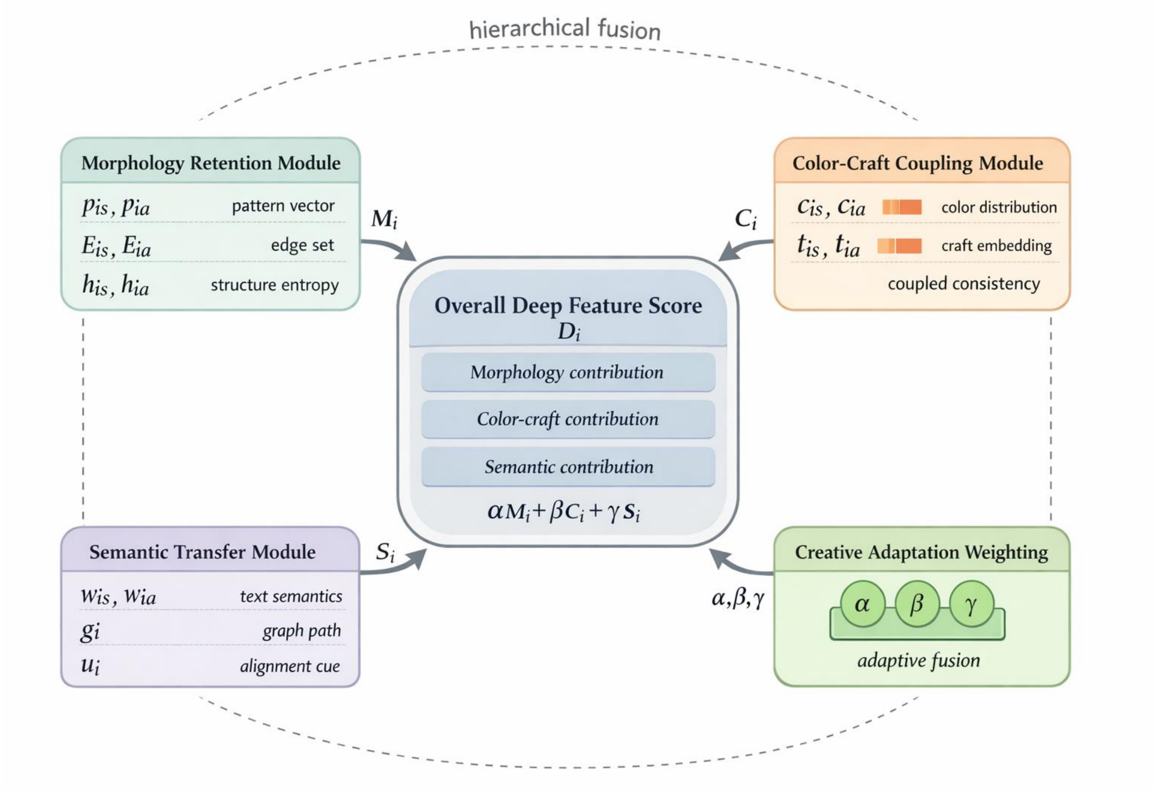


Figure 1: H-DFA Hierarchical Deep Feature Analysis and Optimization Path Diagram.

2.3 Layer-by-Layer Optimization Strategy and Experimental Evaluation Protocol

For letting the optimization part of the method be seen in experiment outcomes, this article establishes four contrast groups. The first one is the tradition visual base line HOG-SVM, which is utilized for giving a lower bound outcome under the non-semantic situation. The second one is CLIP similarity matching, which is utilized for observing the adaptability of general visual-language expressions in this task. The third one is the MICMLF model, which already has shown its text-image fusion ability in the classification of non-material cultural heritage images. The fourth kind of category is the WuMKG-guided matching, which is utilized for comparing the effectiveness of knowledge organization enhancement in the determination of source-work relationships. The H-DFA we put forward, on the opposite side, focuses on source-work matching samples and carries out fine-adjustment and assessment on this identical training set.

Experimental evaluation is based on four metrics. Accuracy is used to assess the overall correctness of fusion level classification; Macro-F1 measures balanced performance across classes; mAP@10 evaluates the ranking quality of source element retrieval results; and Spearman's ρ examines the rank correlation between model scores and expert ratings.Considering that artistic images often exhibit occlusion, repainting, and medium-related interference, this paper additionally sets up four levels of occlusion perturbation tests (0%, 10%, 20%, and 30%) and separately calculates performance metrics for six categories-Chinese ink painting, decorative painting, illustration, mural design, digital painting, and mixed media-to examine the method's generalization capability.

One united group of parameters was utilized for the training stage. The visual portion utilized ViT-B/16 to be the starting backbone, meanwhile the text portion adopted Chinese RoBERTa to be the semantic encoder. The optimization algorithm was AdamW, it has an initial study rate of 1×10^{-4}, a batch magnitude of 32, and 80 training cycles, with early stopping be triggered by the optimal outcome on the verification collection. All the methods for comparison have been operated upon the same division of data. The aim of this method is to guarantee that the Results and Discussion part can directly respond whether the method has effect, whether the optimization is correct, and whether the calculation expense can be controlled, hence not turning the discussion to individual cases that have no relation to the core research question of this paper. In order to avoid differences on the results that will be shown afterward, the comparison relations and assessment norms which are used in this paper are shown in Figure 3.

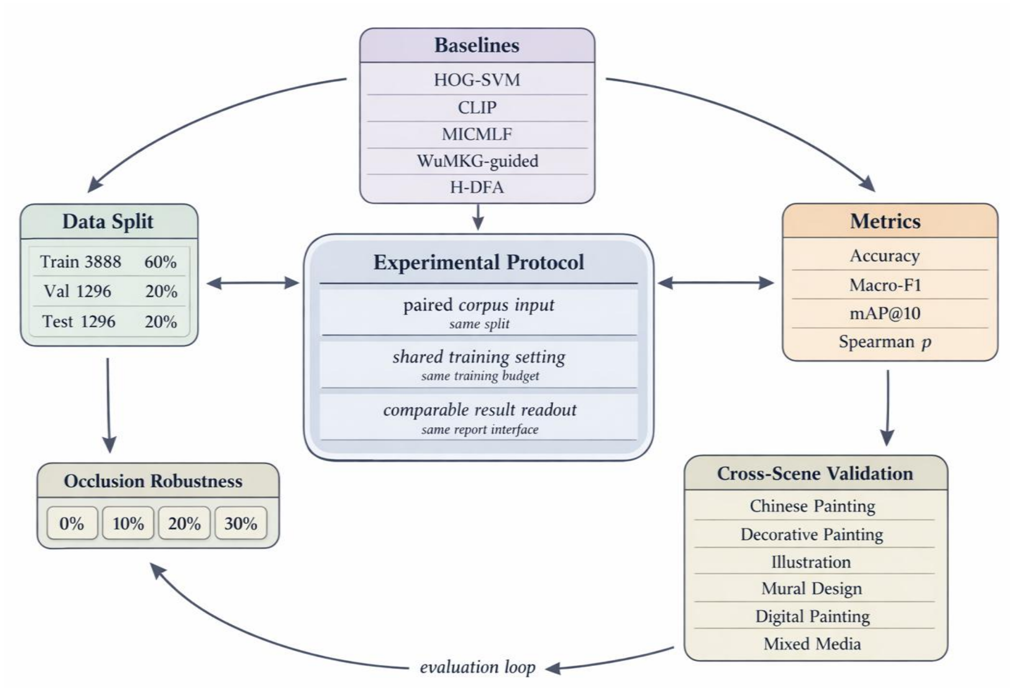


Figure 3: Experimental Protocol and Evaluation Diagram.

3 Results and Discussion

3.1 Overall Performance Validation

After we have built the source-target matching language database and done modeling for layer-by-layer characteristics, the first problem that needs to be solved is: Can H-DFA on the whole task always have better performance than the existing methods? For solving this problem, the present paper carries out comparisons among three aspects: the combination of ranking, the quality of retrieval, and the consistency of expert grading, which is displayed in Table 2 and Figure 4. Table 2 gives the whole outcomes of diverse methods upon the test set, hence Figure 4 further unifies Accuracy, Macro-F1, mAP@10, and Spearman ρ into a shared measurement to see differences in the whole property of each method. This demonstration follows the preset rule that putting figures and diagrams in front, then putting relevant outcomes after.

Table 2: Comparison of Overall Performance Across Different Methods

Method	Accuracy / %	Macro-F1 / %	mAP@10	Spearman ρ
HOG-SVM	71.4±0.8	69.8±0.9	0.621±0.011	0.580±0.021
CLIP	79.6±0.7	78.3±0.8	0.704±0.010	0.660±0.018
MICMLF	84.7±0.6	83.8±0.6	0.762±0.008	0.710±0.015
WuMKG-guided	86.3±0.5	85.2±0.5	0.781±0.007	0.740±0.014
H-DFA	90.8±0.4	89.9±0.4	0.844±0.006	0.810±0.012

Table 2 give us that H-DFA has gotten the most excellent outcomes on all four core indexes, with Accuracy, Macro-F1, mAP@10, and Spearman ρ achieve 90.8%, 89.9%, 0.844, and 0.810, respectively. When making comparison with the traditional visual baseline method HOG-SVM, these four measurement indices have obtained increases of 19.4 percentage points, 20.1 percentage points, 0.223, and 0.230, respectively; when make comparison with the common method of matching text and image which is called CLIP, they got promotion of 11.2 percentage points, 11.6 percentage points, 0.140, and 0.150, respectively; even even when we make a comparison with the stronger baseline WuMKG-guided, H-DFA therefore still obtained promotion of 4.5 percentage points, 4.7 percentage points, 0.063, and 0.070, respectively. The outcome shows that depending only on conventional texture descriptions or overall semantic similarity is still not enough for completely grasping the deep translation relationships of intangible cultural heritage elements within artistic creation. However, through putting pattern structure, color methods, and meaning shifting into one united marking frame, the model has a better ability to produce overall expressions which match the judgment made by human beings. Figure 4 further gives explanation to this difference.

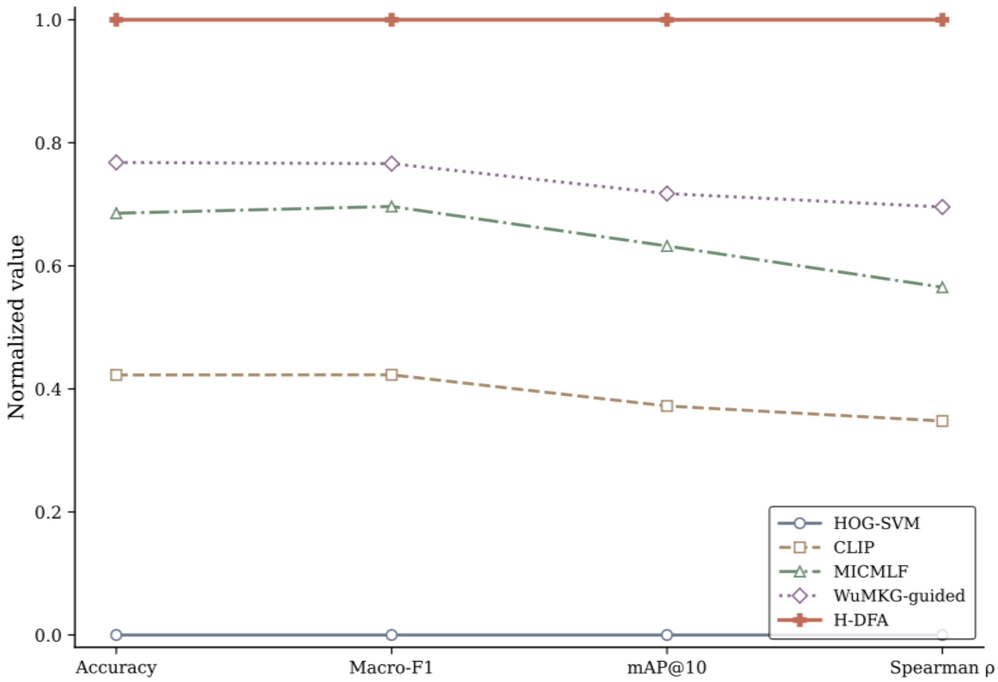


Figure 4: Multi-indicator standardized comparison curves for each method.

In Figure 4, the standardized measure curve of H-DFA is located at the most outside position, this shows that its advantages are not concentrated on a single metric, but are reflected in the simultaneous promotion in three dimensions: classification, ranking and expert consensus. Although the curves of WuMKG-guided and MICMLF have already clearly extended outwardthis shows that knowledge relationships and multimodal fusion produce a positive influence upon this task-both of them therefore still keep a steady gap with H-DFA in the aspects of Spearman's ρ and mAP@10. This indicates that only text-image fusion or only knowledge enhancement is not enough for creative translation analysis; it is still a requirement that we must explicitly bring creative adaptability and hierarchical feature relationships into the process of model construction. The CLIP curve on the whole is more flat, hence it shows its ability of general visual semantic expression; however, its maximum bound is distinctly restricted when confronting tasks in which handicraft marks, symbolic meanings, and structure protection all take effect at the same time. The HOG-SVM model possesses the smallest envelope area, this further gives indication that traditional partial visual features are unable to support deep-layer judgments for this kind of complex culture images.

In terms of the nature of the results, the advantages of H-DFA have two implications. First, the model can more accurately distinguish between high-, medium-, and low-fusion samples, indicating that it more consistently captures the boundary between "effective translation" and "superficial borrowing."Second, the model's output exhibits a higher-order correlation with expert scores, indicating that the results represent not only classification optimization in a machine-learning sense but also stronger alignment with human evaluation. This is particularly critical for research on intangible cultural heritage elements, as creative analysis requires not only the identification of visual similarity but also an understanding of whether a work maintains stable cultural significance after formal restructuring.

3.2 Layer-by-Layer Optimization, Ablation Results, and Robustness Analysis

Although the total outcomes prove that this method has good effect, thus further analysis is required for finding out which particular modules are the ones that bring this advantage. For solving this question, this paper carries out an analysis from two angles: optimization gains in each layer and anti-occlusion stability, which is displayed in Table 3 and Figure 5(a). Table 3 gives the step-by-step outcomes of the model starting from the basic visual branch to the whole H-DFA framework, and Figure 5(a) makes a comparison of the changes in Macro-F1 among different models when occlusion rates get progressively larger.

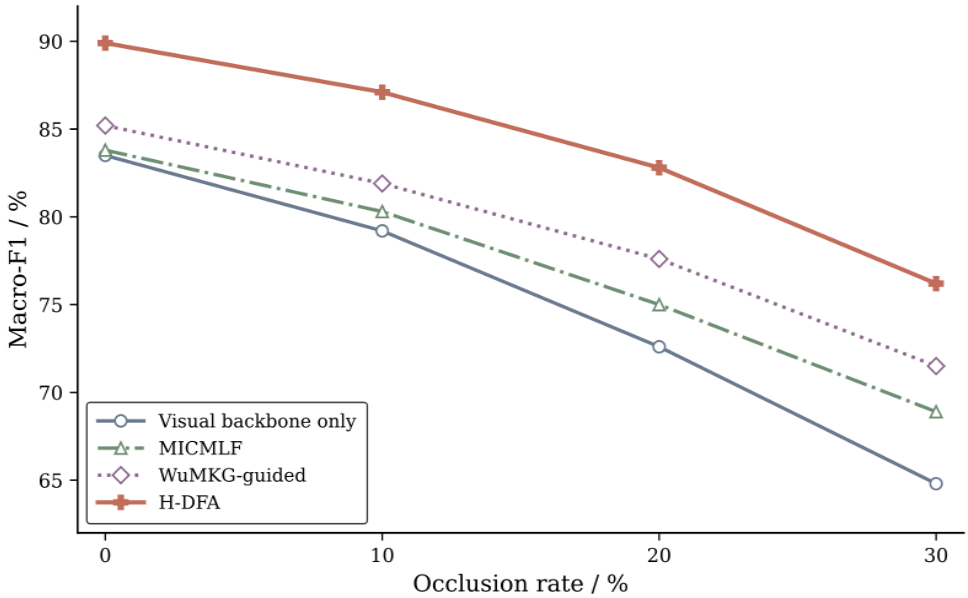
Table 3: Layer-by-Layer Optimization: Ablation and Efficiency Comparison

Variant	Accuracy / %	Macro-F1 / %	mAP@10	Spearman ρ	Latency / ms
Visual backbone only	84.2±0.6	83.5±0.7	0.751±0.009	0.700±0.017	28±1.2
+ Morphology and color decomposition	86.0±0.5	85.1±0.6	0.778±0.008	0.730±0.015	31±1.1
+ Semantic correction	88.1±0.5	87.4±0.5	0.814±0.007	0.780±0.014	35±1.0
+ Creative adaptation weighting	90.8±0.4	89.9±0.4	0.844±0.006	0.810±0.012	38±1.1

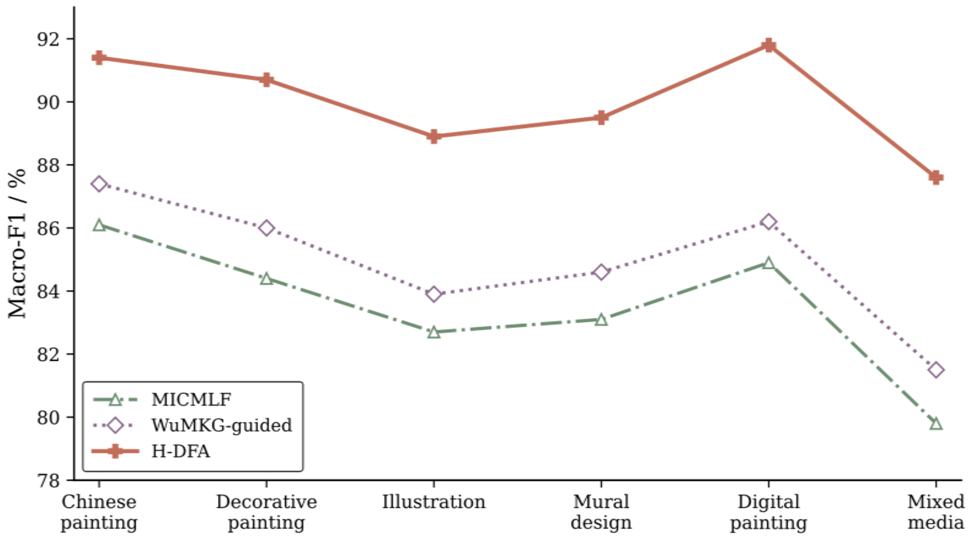
Table 3 makes clear that model performance displays an evident gradual tendency when modules are brought in. When we only use the visual foundation branch, this model already has obtained an accuracy of 84.2% and a Macro-F1 score of 83.5%, this indicates that the source-pair corpus by itself can give a comparatively stable baseline for visual recognition work. However, the response which this model gives to cultural semantics and craftsmanship traces is still limited at current stage, with mAP@10 and Spearman ρ being only 0.751 and 0.700, respectively. After we bring in the layered decomposition of patterns and colors, the Accuracy got promotion to 86.0%, Macro-F1 was raised to 85.1%, and mAP@10 and Spearman ρ were lifted to 0.778 and 0.730, each one. These outcomes tell us that through respectively building models for structural morphology and color group organization, the model can with more convenience make distinction between samples which are "visually similar but structurally different" and those which are "tonally similar yet possess different craftsmanship syntax".

After we further add semantic correction, the performance of model obtained obvious promotion again: The accuracy attained 88.1%, Macro-F1 attained 87.4%, mAP@10 promoted to 0.814, and Spearman's ρ promoted to 0.780. When we compare with the past stage, the promotion of ranking quality and expert consistency is especially obvious, hence this shows that alignment information between text descriptions, knowledge connections and image details plays a very key role in judging whether a creative work can really retain the cultural meaning of intangible cultural heritage elements. That is to say, the visual look by itself is not enough for evaluating the adaptation quality; The introduction of meaning-level information greatly enhances the model's ability of explanation.

Finally, after incorporating creative adaptation weighting, the model's four metrics further improved to 90.8%, 89.9%, 0.844, and 0.810.Compared to the "Semantic Correction" stage, Accuracy and Macro-F1 improved by 2.7 percentage points and 2.5 percentage points, respectively, while mAP@10 and Spearman's ρ both increased by 0.030. Meanwhile, the inference time per sample increased only from 35 ms to 38 ms, keeping the additional computational overhead at a low level.These results indicate that the creative adaptation weighting does not introduce an excessive computational burden, yet significantly improves the model's ability to judge samples involving "moderate deformation, medium conversion, and symbolic extension." In other words, the role of this module is not to elevate general visual similarity, but to correct the portion of translated samples most prone to misjudgment in creative analysis. To demonstrate whether the method's gains can remain stable in complex creative images, see Figure 5.



(a) Occlusion Robustness



(b) Cross-scene performance

Figure 5: Comparison of occlusion robustness and cross-scene performance.

Figure 5(a) gives further confirmation to the stability of the model. Along with the occlusion rate rising from 0% to 30%, the Macro-F1 scores of every method keep going down, but the degree of the falling differs in very big way. The visual backbone network has only its accuracy dropped from 83.5% to 64.8%, and this is a cumulative decrease of 18.7 percentage points; MICMLF drops from 83.8% to 68.9%, which is a total decrease of 14.9 percentage points; Under the guidance of WuMKG, the value drops from 85.2% to 71.5%, hence the accumulative reduction is 13.7 percentage points; H-DFA has decreased from 89.9% to 76.2%, which is a cumulative fall of 13.7 percentage points, hence it all the time keeps the highest curve position in all four occlusion levels. In Figure 5(a), H-DFA got 87.1%, 82.8%, and 76.2% under 10%, 20%, and 30% occlusion levels, respectively, this shows that the model does not depend on single local modes or lone color blocks to do classification, but instead it forms a more stable whole representation through the cooperation of level-by-level features.

This result is particularly important for artistic creation images. Actual works often involve situations such as overlaying, repainting, collage, material reflections, and partial cropping. If a model relies too heavily on locally prominent regions, its performance will rapidly degrade once key edges or color blocks are damaged.The H-DFA's decline curve is flatter, indicating that a compensatory relationship has formed among the three dimensions of morphological retention, color craftsmanship, and semantic transfer: when visual information in a specific local area is missing, other levels can still provide effective constraints.

3.3 Cross-Scene Generalization, Expert Consistency, and Practical Implications

After we have found out where the performance improvement of this method comes from, it is further needed to confirm whether H-DFA has cross-media stability, and whether its composite scores are consistent with the evaluations given by experts. For solving this problem, this article carries out analysis on performance differences in different artistic creation situations and the corresponding relations between forecast scores and human evaluation scores, which is displayed in Figures 5(b) and 6.

Figure 5(b) displays the Macro-F1 scores of the four methods across six creative scenarios. Overall, H-DFA achieved the highest results in all six scenarios-Chinese ink painting, decorative painting, illustration, mural design, digital painting, and mixed media-reaching 91.4%, 90.7%, 88.9%, 89.5%, 91.8%, and 87.6%, respectively. Compared to MICMLF, H-DFA achieved improvements of 5.3, 6.3, 6.2, 6.4, 6.9, and 7.8 percentage points across the six scenarios; compared to WuMKG-guided, the improvements were 4.0, 4.7, 5.0, 4.9, 5.6, and 6.1 percentage points, respectively.The results indicate that the advantages of H-DFA are not limited to a single medium or style, but rather it maintains stable classification performance across different creative contexts.

Looking at specific scenarios, digital painting and traditional Chinese painting achieved relatively high scores, reaching 91.8% and 91.4%, respectively. This is related to the high level of completeness in composition, color expression, and textual descriptions in these two categories of works, which allows the model to more fully extract pattern skeletons, color group orders, and semantic mapping relationships. Decorative paintings and mural designs scored slightly lower but still remained above 89%, indicating that H-DFA maintains good stability even in cases of significant formal reorganization and spatial expansion. The composite material category scored the lowest at 87.6%, yet this remains significantly higher than other methods. This is primarily due to material layering, blurred boundaries, and complex surface textures, which simultaneously interfere with morphological decomposition and the identification of craftsmanship traces. As shown in Figure 5(b), the decline in performance for other methods is more pronounced in composite materials, indicating that models lacking hierarchical constraints are more prone to misjudging source relationships when a work's surface exhibits more complex medium characteristics. Figure 6 further examines the consistency between model outputs and expert ratings. Across 216 expert-reviewed samples, H-DFA's predicted scores and human ratings are concentrated along the regression line, with a linear regression R^2 of 0.78 and a mean absolute error of 0.064.

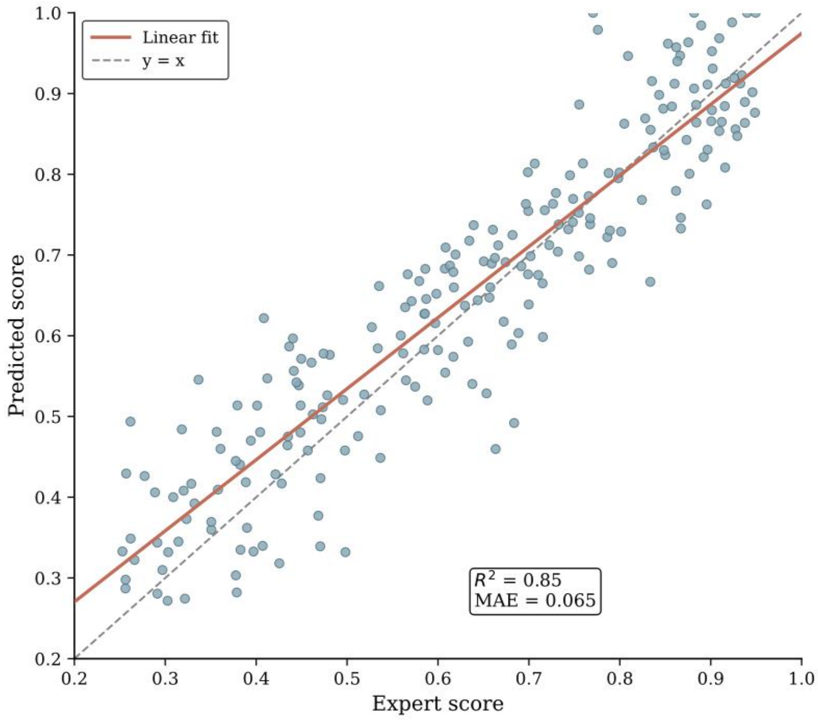


Figure 6: Scatter plot of the consistency between 's predicted scores and expert ratings.

In Figure 6, the majority of samples which have high and moderate fusion are distributed close to the diagonal line, hence this indicates the model is comparatively stable when it identifies works that "retain core visual syntax while cultural translation is finished". A little quantity of outlying points are mainly gathered in the low-fusion scope. These samples generally can be divided into three kinds: firstly, the source elements only keep partial vestiges of patterns, hence leading to not enough morphology marks; second, the coverage of materials and the transformation of medium are excessively extreme, therefore it is hard to carry out effective mapping of craft characteristics; Third, creative depictings are over-short or unclear, hence weakening the capability of the model to make distinction on the semantic transfer layer. Therefore, the errors of model mainly come from the samples which have low information density, hence not from the large-scale deviations that exist in high-fusion samples.

This has obvious meanings which can be used in practice. For digital curation work, course assessing, and creative help checking, the system most needs to accurately recognize those works which have high and medium integration, therefore these are more possible to become the core topics of teaching feedback, curation arrangement, and style file keeping. H-DFA carries out its work more stably on this group of samples, hence indicating that it not merely is fit to act as a research and analysis tool, hence also satisfies the requirements for being put into use in actual creative support situations. Combined with the single-sample deduction time of 38 ms that Table 3 displays, the model already has showed good deployable capability on traditional image search and curriculum feedback platforms. In addition, Figure 5(b) together with Figure 6 together prove that the superiority of H-DFA comes from the simultaneous satisfaction of two aspects: "cross-scenario stability" and "consistency with expert appraisals." This therefore permits it to supply more dependable technical assistance for the quantitative study of intangible cultural heritage components inside fine art creation.

4 Conclusion

This present article discusses the problem of depth translation of non-material cultural heritage (NICH) components in art creation. It has built a source and target matching language database and an H-DFA layered deep characteristic analysis frame, and thus verifies this method in four aspects: total effect, layer to layer optimization, cross situation expansion application, and expert same opinion. The outcomes show that after ICH components are put into creative creations, their keeping and rearranging cannot be completely explained only by visual likeness. Only via the combining of form structure, color and technique ability, and meaning change into one united analysis frame can the cultural continuation and creative changing that are inside the works be found by us in a more dependable way.

(1) This research has arranged a paired language corpus of ICH source works and creation works to carry out analysis. This research has put ICH source pictures, creative pictures and text descriptions together into one whole data set, and through data cleaning, cutting apart, and three-layer marking, has built matching units that are fit for classification, searching, and expert consistence evaluation. This method for arranging data makes the later analysis can go past the level of single image identification, and hence focus on the core question of "how source components are brought into creative productions."

(2) The H-DFA frame which is put forward by this paper has showed good effect on both method part and result part. The experiment outcomes display that H-DFA obtains 90.8%, 89.9%, 0.844, and 0.810 on Accuracy, Macro-F1, mAP@10, and Spearman ρ, separately, all of which are better than HOG-SVM, CLIP, MICMLF, and WuMKG-guided. Ablation experiment outcomes further show that the layered splitting of patterns and colors, meaning rectification, and creative adjustment weight all together form the main origins of effect promotion, with the meaning layer and weight layer displaying the most notable promotion in expert consistence. The results under cross-scenario conditions prove that this method can keep comparatively

stable distinguishing abilities in different media and creation contexts.

(3) The present research still possesses some restrictive shortcomings. The existing dataset mainly concentrates on ICH kinds that have obvious visual translation features, while ICH sorts that have more powerful performative, narrative and temporal components have not obtained complete inclusion; the currently existing analysis also mainly includes static pictures and written words, hence it does not have further integration of creative process paths, video record materials, and interaction data. In the future, research work may enlarge the language corpus to cover a wider scope of ICH types, and bring dynamic process information into a united framework, therefore to promote the model's capacity for explaining complicated creative behaviors and cultural spread routes.

References

[1] UNESCO. (2024). Basic texts of the 2003 Convention for the safeguarding of the intangible cultural heritage. Paris: UNESCO.

[2] UNESCO. (2024). Harnessing creativity. Paris: UNESCO.

[3] Yu, L. (2023). Digital sustainability of intangible cultural heritage: The example of the “Wu Leno” weaving technique in Suzhou, China. Sustainability, 15(12), 9803.

[4] Liu, Y., Cheng, P., & Li, J. (2023). Application interface design of Chongqing intangible cultural heritage based on deep learning. Heliyon, 9(11), e22242.

[5] Fan, T., Wang, H., & Deng, S. (2023). Intangible cultural heritage image classification with multimodal attention and hierarchical fusion. Expert Systems with Applications, 231, 120555.

[6] Banar, N., Daelemans, W., & Kestemont, M. (2023). Transfer learning for the visual arts: The multi-modal retrieval of Iconclass codes. ACM Journal on Computing and Cultural Heritage, 16(2), Article 32.

[7] Wan, J., Zhang, H., Zou, J., et al. (2024). WuMKG: A Chinese painting and calligraphy multimodal knowledge graph. Heritage Science, 12(1), 159.

[8] Xia, J., Yuizono, T., Wang, T.-Y., et al. (2024). Intangible cultural heritage course design in the digital age and its effects of interdisciplinary workshop. Education Sciences, 14(3), 325.

[9] Sasithradevi, A., Nathan, S., Chanthini, B., et al. (2024). MonuNet: A high performance deep learning network for Kolkata heritage image classification. Heritage Science, 12(1), 242.

[10] Wang, S.-F., & Chen, C.-C. (2025). Revitalizing intangible cultural heritage via derivative design: A focus on Chinese woodblock printing. PLoS ONE, 20(4), e0318807.

[11] Chen, J., Xia, H., & Yu, S. (2025). Integration of intangible cultural heritage elements into furniture design based on symbolic semantics and AHP: A case study of Qianci. BioResources, 20(2), 3714-3731.

[12] Wang, Y., Wu, G., Wang, S., et al. (2025). Visual style transfer for cultural heritage: A comprehensive review of techniques and applications. Design and Artificial Intelligence, 1(2), 100027.

[13] Yu, Q., Tao, X., & Wang, J. (2025). Sustainable design on intangible cultural heritage: Miao embroidery pattern generation and application based on diffusion models. Sustainability, 17(17), 7657.

[14] Zhang, X., Chen, D., & Qin, Y. (2025). Multimodal prototype fusion network for paper-cut image classification. npj Heritage Science, 13, 462.

[15] Zou, J., Du, Y., Liu, G., et al. (2025). Generating Chinese intangible cultural heritage images with structure and color awareness. npj Heritage Science, 13, 579.

[16] Chen, L., Jiang, P., Shen, Y., et al. (2025). PorceVis: An interactive visual analytics system for exploring the history and culture of ancient Chinese porcelain. Visual Informatics, 10(1), 100281.