Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Visual Transformation Mechanisms for Integrating Digital Cultural Heritage Resources into Junior High School Art Curricula

Ji Wu

DesdemonaSchwieger40@outlook.com

No.2 Junior High School, Chang'an District, Xi'an City, Shaanxi Province, Xi'an, 710100, China

Jing Zhang

School of Teacher Development, Shaanxi Normal University; Xi'an Aerospace City No.1 High School, Xi'an, 710100, China

Di Wu

No.3 Junior High School, Chang'an District, Xi'an City, Shaanxi Province, Xi'an, 710100, China

Yiming Peng

Xi'an Chang-an District No.3 Middle School, Xi'an, 710100, China

Research Article

Keywords: computer vision, multimodal encoding, computer application framework, AI-assisted evaluation, digital cultural heritage

Posted Date: August 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10609467/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Visual Transformation Mechanisms for Integrating Digital Cultural Heritage Resources into Junior High School Art Curricula

Ji Wu*

No.2 Junior High School, Chang'an District, Xi'an City, Shaanxi Province, Xi'an, 710100, China, jiwu9488@Gmail.com

Jing Zhang

School of Teacher Development, Shaanxi Normal University; Xi'an Aerospace City No.1 High School, Xi'an, 710100, China, 675408638@qq.com

Di Wu

No.3 Junior High School, Chang'an District, Xi'an City, Shaanxi Province, Xi'an, 710100, China, 452489152@qq.com

Yiming Peng

Xi'an Chang–an District No.3 Middle School, Xi'an, 710100, China, 821569149@qq.com

Abstract

When digital cultural images are integrated into junior middle school art classes, existing applications often lack standardized image preprocessing, computable feature representation, automatic task generation, and data–driven evaluation, making it difficult to connect heritage resources with reusable, repeatable classroom software workflows.To strengthen computer application support for translating cultural images into creative tasks, a modular visual transformation model integrating computer vision, rule–based matching, and intelligent application scoring was developed.Based on the digital Dunhuang murals, the characteristics of line art, color, design, composition and pose were extracted. Course tasks were generated with feature task matching, difficulty classification, and student work coding, and 160 students were tested for six weeks in four classes.The results showed that the experimental group showed more significant improvements in the extraction of visual elements, formal reorganization, and thematic expression. The correlation between AI and teacher scores was 0.86, and the mean absolute error was 3.84. The results show that this application framework improves digital heritage processing, instructional generation, evaluation, and iterative optimization.

CCS Concepts

Applied computing → Arts and humanities → Fine arts

Keywords

computer vision; multimodal encoding; computer application framework; AI–assisted evaluation; digital cultural heritage

1 Introduction

Computer application technology for digital cultural heritage has become an important intersection of computer vision, cultural computing, and intelligent educational service systems. Billi et al. utilized machine learning and deep learning to process laser scanning and drone mapping data, enabling the spatial identification and analysis of complex heritage structures [1]; Croce et al. conducted semi–automatic classification of digital heritage using the Aïoli open–source 2D/3D annotation platform, enhancing the semantic organization of heritage images and 3D data [2];Assran et al. proposed a joint embedding prediction architecture, providing a transferable representation method for visual feature learning with few examples [3]; Brooks et al. investigated instruction–based image editing models, offering technical guidance for the controlled transformation of cultural images [4]; Münster et al. further noted that artificial intelligence is driving the shift of digital heritage from storage and display toward intelligent analysis, interactive generation, and knowledge modeling [5].However, existing studies separately address heritage recognition, three–dimensional reconstruction, image generation, or visualization, leaving unresolved how cultural images can be converted into adaptive learning tasks and evaluated through a reusable computational workflow. To fill this gap, a modular framework is proposed that links standardized preprocessing, five–channel feature encoding, image–task vector matching, learner–sensitive difficulty classification, AI–assisted scoring, teacher correction, and feedback–based parameter updating. Using digital Dunhuang murals, the framework operationalizes the transformation from heritage image decomposition to task

generation, student–work assessment, and iterative optimization, thereby contributing a closed–loop computer application mechanism for junior high school art curricula.

2 Extraction of Visual Features from Digital Dunhuang Resources

First, an image–processing module standardized samples to 512x512 pixels and converted RGB data into Lab color space [6]. During the preprocessing stage, bilateral filtering was used to suppress crack noise in the murals, with parameters set to $d = 5$, $\sigma_c = 50$, and $\sigma_s = 50$;CLAHE was used to enhance low–saturation regions, with the clip limit set to 2.0 and the tile size set to 8×8. During the feature extraction stage, the Canny edge density, K–means primary color clustering, pattern repeat units, compositional center offset, and direction vectors of the flying apsaras' poses were calculated to form a visual feature vector mapped to the course task:

$$\mathbf{F}_i = \left[f_l^i, f_c^i, f_p^i, f_s^i, f_m^i \right] \tag{1}$$

where $\mathbf{F}_i$ is the visual feature vector for the i th Dunhuang image; f_l^i represents line–drawing edge density; f_c^i represents the dominant color clustering feature; f_p^i represents the pattern repetition structure; f_s^i represents compositional spatial hierarchy; f_m^i represents figure poses and the direction of flowing sashes [7]. This vector can be used as a uniform input for the subsequent course task generation and the feature comparison. As illustrated in Figure 1, the images of the flying apsaras emphasize the position of the figure, the direction of the flowing sashes, and the rhythm of the line work; the images of the coffered ceiling primarily emphasize central symmetry, continuous patterns, and repeating units; and the images of donors and narrative scenes primarily emphasize the arrangement of the figures, the spatial hierarchy, and the narrative relations within the composition.The visual information described above is not transformed into a subject for classroom copying, but is broken down into five types of transferable elements — lines, colors, patterns, compositions, and dynamic poses — before being integrated into the course task generation process.

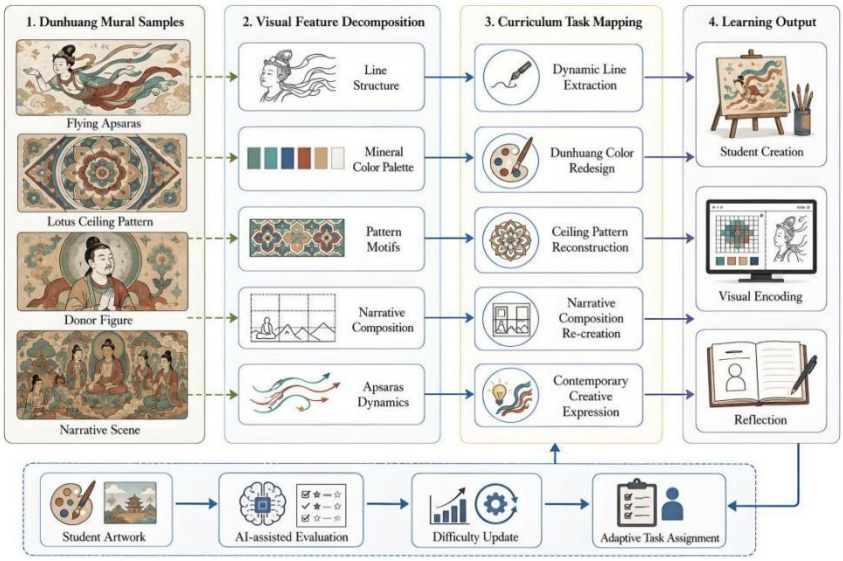


Figure 1: Schematic Diagram of the Deconstruction of Visual Characteristics in Dunhuang Murals and Their Transformation into Course Content

3 Construction and Optimization of the Visual Transformation Model

3.1 Mapping Mechanism from Cultural Images to Curriculum Tasks

Based on the feature vector of the Dunhuang image, the image can be classified into five characteristic channels: line drawing, mineral pigment, coffered ceiling, Apsara dynamics, and narrative composition.The Dunhuang murals are divided into five types of visual information: line-drawing, mineral pigments, coffered ceiling, Apsara dynamics, and narrative composition — and then mapped into a transition path from cultural images to art learning tasks by means of feature vectors and curriculum task vectors [8].The application service generates a record for each image, retrieves task vectors from structured storage, calculates similarity through the matching interface, and writes the highest-scoring result into the task candidate pool for instructional selection.The task vector is encoded based on course objectives, materials, creative steps, and students' pre-test performance, with dimensions consistent with those of the image feature vector.

Cosine similarity is used for the matching calculation:

$$M_{ij} = \frac{\mathbf{F}_i \cdot \mathbf{T}_j}{\| \mathbf{F}_i \|_2 \| \mathbf{T}_j \|_2} \quad (2)$$

where M_{ij} represents the match score between the i th Dunhuang image and the j th course task; $\mathbf{F}_i$ represents the image's visual feature vector; $\mathbf{T}_j$ represents the course task's feature vector; and $\| \cdot \|_2$ represents the L^2 norm. If $M_{ij} \geq 0.75$, the image is assigned to the corresponding task module; if $0.60 \leq M_{ij} < 0.75$, the teacher's secondary confirmation is required.

3.2 Method for Generating Student Creative Tasks

Student creative tasks are produced through a configurable rule-engine application module combining feature, learner ability, and instructional objective constraints [9]. Firstly, the system reads out the image characteristic tags from the task candidate pool, and assigns the high-line-density and high-pose orientation samples to "Extracting Dynamic Lines from Flying Apsaras"; the Dunhuang Color Redesign assigns the samples with 4-6 primary colors with a high Lab Color Difference Discrimination; and the samples with a pattern repetition rate above 0.68 and distinct center-symmetric structures are allocated to "Caisson Pattern Rearrangement." Subsequently, the system imports students' pre-test work scores, task completion rates, and visual literacy test results to generate student ability levels G_s .

The system produces three kinds of tasks for each student: Basic, Advanced, and Open. The basic tasks restrict the percentage of picture elements that can be extracted; advanced tasks add steps to formalise them; and open tasks enable media and thematic settings to be replaced [10]. The Task Generation Module uses the Top 3 Candidate Strategy, which keeps three matching tasks in mind, with the teacher choosing the last task according to the class's progress so as to avoid too narrow a classroom that results from a single algorithm's recommendation.

3.3 Optimization of Difficulty Grading for the Transformation Model

Difficulty grading uses image complexity, operation steps, opening degree, and student completion stability as constraint variables. L1 Visual Interpretation, L2 Form Extraction, L3 Form

Reorganization, and L4 Contemporary Expression.L1 is equivalent to single line, color, or pattern observation; L2 demands that portable style elements be extracted from Dunhuang images; L3 needs to rearrange lines, colors, and compositions; and L4 needs to be able to achieve personalized expression across themes and media.After each classroom assignment, the application records work-feature deviation, task omission rate, and teacher-score dispersion in the learner profile database, then invokes the feedback algorithm through the difficulty service to update model parameters, refresh task priorities, and recommend the next appropriate difficulty level .

Difficulty scoring is defined as follows:

$$D_j = \alpha C_j + \beta P_j + \gamma O_j - \lambda R_j + \mu E_j \tag{3}$$

where D_j represents the comprehensive difficulty of the task in category j ; C_j represents image complexity, obtained by normalizing edge density, the number of primary colors, and pattern repetition levels; P_j represents the number of operational steps; O_j represents task openness; R_j represents the student's historical completion rate; E_j represents deviations in work characteristics; α , β , γ , λ , and μ are weight coefficients, initially set to 0.25, 0.20, 0.20, 0.20, and 0.15, respectively. If $D_j < 0.35$ is labeled as L1, $0.35 \le D_j < 0.55$ as L2, $0.55 \le D_j < 0.75$ as L3, and $D_j \ge 0.75$ as L4.

4 AI-Assisted Evaluation and Experimental Validation

4.1 Feature Encoding of Student Works

Building on task mapping results discussed earlier, student works were collected using a standardized photography protocol. The image was compressed to 512 × 512, the white balance was fixed at D65, and the image was fed into the feature coding module after perspective correction. The application pipeline extracted Canny edge density, Lab dominant-color distribution, pattern repetition, and composition offset through operators; CLIP-ViT-B/16 generated deep embeddings, while two projection heads were fine-tuned through batch training and stored for subsequent scoring and retrieval. The learning rate was 1 × 10 − 5 and the batch size was 16.The feature vector for student

artwork is defined as:

$$\mathbf{S}_k = \left[g_l^k, g_c^k, g_p^k, g_s^k, g_m^k, g_e^k \right] \tag{4}$$

In the equation, $\mathbf{S}_k$ represents the comprehensive feature vector of the student work at position k; g_l^k represents the line structure features; g_c^k represents the color features; g_p^k represents the pattern recombination features; g_s^k represents the compositional hierarchy features; g_m^k represents the dynamic pose features; and g_e^k represents the semantic expression embedding features. Then, we compare this vector with the original Dunhuang image to calculate the degree of visual transformation.

4.2 Design of AI Evaluation Metrics

The evaluation service combines visual–feature scoring with instructor–label correction, calls ConvNeXt–Tiny and CLIP encoders through inference services, and passes fused representations through a two–layer MLP to return five–dimensional scores and diagnostic records for persistent storage. Instead of using traditional art grading rubrics, it corresponds to the task mapping chain described in Chapter 3, transforming "accuracy of extraction, effectiveness of combination, and whether the expression deviates from replication" into computable dimensions. Weights are determined by backward calibration on the basis of the consistency of the teacher and model errors; the initial weights are locked after five times cross validation, as indicated in Table 1.

Table 1: Weighting Table for AI–Assisted Student Work Evaluation Metrics

Evaluation Dimensions	Feature Source	Calculation Basis	Weight
Cultural Element Recognition	CLIP Semantic Embeddings	Similarity to Dunhuang Image Labels	0.18
Visual Element Extraction	Line, color, and pattern features	Match Rate with Original Feature Vectors	0.24
Degree of Formal Reconstruction	Composition and Pattern Structure Features	Rate of Structural Change and Stability	0.22
Clarity of Thematic	Text Labels and Image	Alignment of Work Theme with	0.16

Expression Completeness of the Work	Semantics Boundaries, Color, and Local Details	Task Objectives Image Completeness Rate vs. Proportion of Noisy Areas	0.2

The composite score is calculated as follows:

$$Q_k = \sum_{r=1}^{5} w_r q_{kr} \tag{5}$$

Where Q_k is the AI composite score for the k th work; q_{kr} is the standardized score for that work on the r th evaluation dimension; and w_r is the corresponding dimension weight, satisfying $\sum_{r=1}^{5} w_r = 1$.

In Table 1, "Visual Element Extraction" and "Formal Reconstruction Completion" weights are relatively high, suggesting that the model places more importance on the successful transformation of the characteristics of the legacy images into their own creation structures, and not just the identification of cultural symbols.

As illustrated in Figure 2, the difference between AI scores and average teacher scores is mainly in the [–5, 5] range, with an average absolute error of 3.84 and a Pearson correlation coefficient of 0.86. Errors exceeding 8 points occurred mainly in works with clear themes but weak cultural elements, because the model emphasizes feature similarity and structural completeness, whereas teachers interpret metaphor, expressive intention, emotional tension, and contextual deviation from reference images.

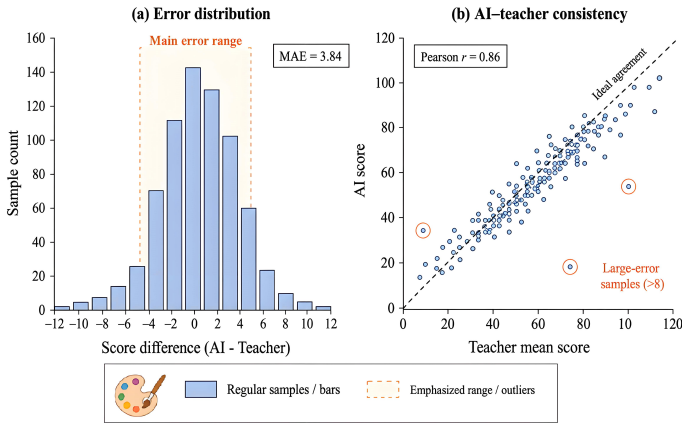


Figure 2: Distribution of Consistency Errors Between AI and Instructor Scores

4.3 Analysis of Curriculum Experiment Results

The study was carried out in four middle school classes, with 160 students, for six weeks. Two experimental classes used the integrated computer application workflow of image extraction, task mapping, AI-assisted evaluation, learner-profile storage, teacher correction, and adaptive difficulty feedback. The other two classes were the control group, and they used the image appreciation, the teacher demonstration, and the topic creation method. In order to avoid over-idealization of the data, it was possible to exclude invalid samples on the basis of occlusion, blurring or non-submission of the assigned task. Finally, the composition of the data used to analyze the model is presented in Table 2.

Table 2: Composition of the Course Experiment Dataset

Data Type	Original Count	Valid Count	Purpose
Student Work Images	320 pieces	306	Feature Encoding and Score Validation
Learning Task Sheets	480	456	Task Completion Rate and Difficulty Updates
Teacher Grading Records	960 entries	918 entries	AI Grading Label Corrections
Classroom Observation Records	18	18	Explanations for Abnormal Tasks
Student Interview Records	20 students	20 visits	Explanation of Work Expressions

Table 2 shows that the validity rates of the art images and the scoring records were 95.6% and 95.6%, respectively, to support the transfer verification in small sample conditions.The same scoring dimensions were used before and after the course, and the results of the experiment are presented in Table 3. In the experimental group, there was a significant improvement in both visual elements extraction and formal reorganization, with a 13.2 and 14.1 improvement in the control group, but this was mainly focused on the completeness of the work, suggesting that the traditional techniques helped to complete the work, but had limited impact on the conversion of the characteristics of the cultural image.

Table 3 Comparison of Course Results Between the Experimental and Control Groups

Evaluation Dimensions	Experimental Group Pre-test	Experimental Group Post-Test	Improvement	Control Group Pre-test	Control Group Post-Test	Improvement
Cultural Element Recognition	68.4	79.6	11.2	67.9	73.2	5.3
Visual Element Extraction Rate	64.7	77.9	13.2	65.1	70.4	5.3
Degree of Formal Restructuring	61.5	75.6	14.1	62.2	68.1	5.9
Completion Clarity of Topic Expression	66.8	76.2	9.4	66.5	71	4.5
Work Completeness	70.3	80.1	9.8	69.8	76.3	6.5

As shown in Fig. 3, the Dunhuang mural resources are not used as objects to be copied in the classroom, but are transformed into actual visual study cases. Apsara images correspond to dynamic lines, emphasizing the direction of flowing ribbons, body positions, and line-drawing rhythms; mineral pigments correspond to the reinterpretation of the colors, leading the students to rearrange the tone palette of their compositions with green, ochre, and earth yellow; ceiling medallion patterns correspond to pattern reconstruction, reinforcing central symmetry, repeating units, and the organization of border motifs; narrative scenes correspond to the reinterpretation of contemporary composition, prompting students to transform character relationships, spatial hierarchy, and narrative threads into individual creations. Thus, it is clear that the educational value of the course is not derived from a single case of viewing images, but from the process of deconstruction, translation, and re-creation of heritage images.

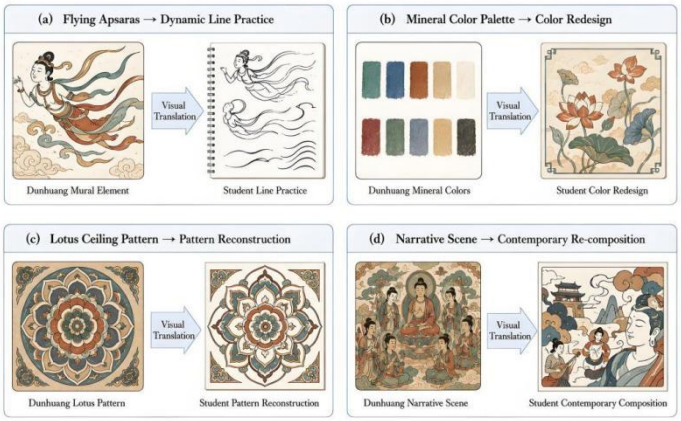


Figure 3: Teaching Case Study on the Visual Translation of Dunhuang Murals

5 Conclusion

A computer application framework for junior middle school art curricula was constructed by integrating Digital Dunhuang resources, image preprocessing, feature computing, data storage, task mapping, difficulty adaptation, and AI-assisted evaluation modules within a workflow. The framework converts mural lines, colors, patterns, compositions, and dynamic poses into structured feature vectors, enabling cultural images to be processed, stored, matched, evaluated, and iteratively transformed into reusable student creative tasks. Experimental results confirm improvements in visual-element extraction, formal reorganization, and thematic expression, while the intelligent scoring module maintains strong overall consistency with teacher evaluations. Limitations in sample scale, heritage categories, grade coverage, and computing environments indicate that application stability, transfer performance, system scalability, practical usability, and generalization still require broader cross-regional verification. Future development should expand cross-regional datasets, optimize multimodal fusion, lightweight inference, and service interfaces, and incorporate intangible heritage, museum imagery, and local resources into scalable educational software applications.

References

[1] Billi D, Croce V, Bevilacqua M G, et al. Machine learning and deep learning for the built heritage analysis: laser scanning and UAV-based surveying applications on a complex spatial grid structure[J].

Remote Sensing, 2023, 15(8): 1961.

[2] Croce V, Manuel A, Caroti G, et al. Semi-automatic classification of digital heritage on the Aïoli open-source 2D/3D annotation platform via machine learning and deep learning[J]. Journal of Cultural Heritage, 2023, 62: 187—197.

[3] Assran M, Duval Q, Misra I, et al. Self-supervised learning from images with a joint-embedding predictive architecture[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 15619—15629.

[4] Brooks T, Holynski A, Efros A A. Instructpix2pix: Learning to follow image editing instructions [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 18392—18402.

[5] Münster S, Maiwald F, di Lenardo I, et al. Artificial Intelligence for Digital Heritage Innovation: Setting Up an R&D Agenda for Europe [J]. Heritage, 2024, 7(2): 794—816.

[6] Niccolucci F, Felicetti A. Digital Twin Sensors in Cultural Heritage Ontology Applications [J]. Sensors, 2024, 24(12): 3978.

[7] Stacchio L, Garzarella S, Cascarano P, et al. DanXe: An extended artificial intelligence framework to analyze and promote dance heritage[J]. Digital Applications in Archaeology and Cultural Heritage, 2024, 33: e00343.

[8] Zabulis X, Partarakis N, Bartalesi V, et al. Multimodal Dictionaries for Traditional Craft Education[J]. Multimodal Technologies and Interaction, 2024, 8(7): 63.

[9] Kontogiorgakis E, Zidianakis E, Kontaki E, et al. Gamified VR storytelling for cultural tourism using 3D reconstructions, virtual humans, and 360 videos[J]. Technologies, 2024, 12(6): 73.

[10] Darcet T, Oquab M, Mairal J, et al. Vision Transformers Need Registers[C]//International Conference on Learning Representations. 2024, 2024: 2632—2652.