Artificial Intelligence-Assisted Creation and Expression in Digital Media Art: A Human-AI Collaborative Framework for Dynamic Visuals and Interactive Design

Weijing Zhong*, Jiufang Mei, Ling Jin, Yingfei Jia, Wei Song

Harbin Institute of Information Technology, Harbin, China

*1280990489@qq.com

Abstract. To address the limitations of current AI-generated content (AIGC) tools in digital media art creation—particularly the lack of temporal coherence, low interactivity fidelity, and insufficient support for dynamic expression—this paper proposes a human-AI collaborative framework for digital media art creation and presentation. First, spatiotemporal-visual features are extracted from motion graphics and user interaction sequences to construct an interactive behavior prediction model. Second, mapping rules between AI-generated dynamic clips and editable layers in digital media software (e.g., After Effects, TouchDesigner) are established, and a combination of a Deep Q-Network (DQN) and a conditional diffusion model (CDM) is used to generate optimized dynamic visual sequences and interactive responses. Finally, a three-stage workflow of "AI generation + software refinement + artist interaction tuning" is formed. Experimental results show that the proposed method significantly outperforms pure AI generation and pure manual creation in terms of temporal coherence, interaction flexibility, and overall user satisfaction, providing an efficient and expressive solution for AI-assisted digital media art.

Keywords: Digital Media Art; AI-Assisted Creation; Dynamic Visuals; Human-AI Collaboration; Interaction Design; Motion Graphics.

1 Introduction

With the rapid development of generative models, artificial intelligence has been increasingly applied to digital media art creation, including motion graphics, projection mapping, interactive installations, and visual effects. However, current AIGC tools (e.g., Stable Diffusion, Runway Gen-2) excel at generating isolated image frames but suffer from critical weaknesses in dynamic media: temporal inconsistency between frames, lack of editable motion paths, poor synchronization with interactive inputs, and difficulty in achieving fine-grained control over rhythm and expression. These shortcomings severely limit the application of AIGC in professional digital media workflows, where temporal coherence, real-time interactivity, and layered editing are essential.

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_49

Recent AIGC integration efforts into digital media creation fall into three directions: frame-wise generation ignoring temporal continuity [1][2], video diffusion models with flickering/jitter [3][5], and plugin-based integrations lacking temporal/interactive modeling. Studies incorporating user interaction [4][6]rarely map AIGC's latent space to keyframe/node-based editing or leverage traditional tools (e.g., After Effects, TouchDesigner) to repair temporal defects. Existing frameworks lack quantitative evaluation of temporal coherence and interaction fidelity. To our knowledge, no prior work proposes a complete human-AI collaborative framework that explicitly models spatiotemporal-visual features, predicts artist actions, and maps AIGC outputs to editable layers for dynamic interactive digital media art. Furthermore, the application of AI in UI design optimization for digital media has also been explored [7], yet it does not address the temporal and interactive challenges targeted in this work.

To address these gaps, this study proposes an AIGC-assisted digital media art creation framework specifically optimized for dynamic visual expression and interactive design. By extracting spatiotemporal-visual features, constructing mapping rules from AIGC-generated clips to editable media layers, and introducing a three-stage workflow of "AI generation + software refinement + artist tuning," experiments demonstrate that the proposed method significantly outperforms pure AIGC or pure manual approaches in temporal coherence, interaction flexibility, and overall satisfaction.

2 Method

2.1 Spatiotemporal-Visual Feature Representation for Digital Media

In digital media art, the input dynamic composition is represented as a sequence of visual keyframes and interaction events. Let a dynamic visual segment VV be divided into NN temporal intervals:$T = \{t1, t2, \ldots\ldots tN\}$

Each interval titi contains attributes such as motion vectors, color transition profiles, layer opacity curves, and a binary flag indicating whether it contains user-interactive elements. Meanwhile, a set of user interaction sequences (e.g., mouse movements, touch gestures, sensor inputs) is collected:$I = \{i1, i2, \ldots\ldots iM\}$

Each interaction ij contains type, duration, intensity, and target region.On this basis, define the joint feature vector for each "temporal interval–interaction" pair $(\mathrm{t_i}\ , \mathrm{i_j})$ as:$\mathrm{Z_{ij}} = \{\mathrm{M_{ij}}, \mathrm{C_{ij}}, \mathrm{R_{ij}}, \mathrm{D_{ij}}\}$

Where:

$\mathrm{M_{ij}}$: Motion coherence (optical flow consistency between adjacent frames in interval t_i under interaction i_j;

$\mathrm{C_{ij}}$: Chromatic continuity (smoothness of color transitions);

$\mathrm{R_{ij}}$: Rhythm alignment (temporal matching between visual change rate and interaction pace);

2.2 Interactive Behavior Prediction Model

Based on Z_{ij} , a lightweight behavior predictor is constructed to estimate the type of editing action an artist is likely to take in a given temporal interval (e.g., adjust keyframe velocity, modify transition curve, add interactive response, regenerate segment). The model architecture is a three-layer fully connected network with input dimension 4 (the four feature components), hidden layer dimensions 32 and 16 respectively, and output dimension 5 (corresponding to five action types: adjust keyframe velocity, modify transition curve, add interactive response, regenerate segment, and no action). All hidden layers use ReLU activation functions, and the output layer uses a softmax function to produce a probability distribution over actions. The model is trained using the cross-entropy loss function:

The model is trained using the cross-entropy loss function:

$$L = \frac{1}{B}\sum_{b=1}^{B}\sum_{c=1}^{5} y\,b,c \quad log(yb,c)$$

Where B is the batch size, y is the one-hot encoded ground-truth action, and ŷ is the predicted probability. The optimizer is Adam with an initial learning rate of 0.001, weight decay of 1e-4, and a batch size of 32. The model is trained for a maximum of 200 epochs with early stopping (patience = 15) based on validation loss. A validation strategy of 5-fold cross-validation is applied on the dataset of 520 behavioral logs of digital media artists editing AIGC-generated dynamic sequences in After Effects and TouchDesigner. The dataset is split into 80% training (416 samples) and 20% validation (104 samples) for each fold. The reported accuracy of 78.4% is the average over the five validation folds, with a standard deviation of 2.1%. This indicates stable generalization performance across different artist styles and software environments.

2.3 Mapping Rules from AIGC Generation to Digital Media Software Tools

To enable seamless collaboration, mapping rules from AIGC-generated dynamic clips to editable digital media layers/tools are established as shown in Table 1.

Table 1. AIGC Region →→ Digital Media Tool Mapping Rules

AIGC-Generated Segment Type	Recommended Software Tool	Parameter Preset	UserControl Level
Background Motion	After Effects – Generative Fill + Particle Generator	Loop, speed ramping	Low
Animated Text/Title	AE Text Animator + Graph Editor	Per-character offset, easing	High
Transition Effect	AE Transition Layer + Time Remapping	Frame blending, motion blur	Medium

Interactive Responsive Element	TouchDesigner – CHOPs + Feedback Loop	Real-time input mapping	Focused Intervention
Complex Particle System	AI Regeneration+AE Particle Compositing	Multi-layer blending	Manual Fine-Tuning

The mapping rules were jointly formulated by three senior digital media artists and two AI engineers, used in a fixed manner, and do not introduce data-driven learning.

2.4 AI-Driven Dynamic Sequence Generation and Artist Optimization

This method adopts a three-stage collaborative mechanism of "AIGC divergence + software refinement + artist tuning".

This workflow comprises three stages: (1) AIGC generation using a conditional diffusion model conditioned on temporal masks and motion prompts to produce candidate dynamic sequences; (2) software automatic refinement that invokes digital media tools for keyframe interpolation, motion curve smoothing, and real-time input binding; and (3) artist tuning based on temporal coherence, expressive rhythm, interaction fidelity, and regeneration necessity. This achieves efficient collaboration where AI handles base motion, software refines temporal dynamics, and humans control interactive expression.

3 Experiment and Analysis

3.1 Dataset and Experimental Setup

A test set containing 160 digital media artworks was constructed, covering motion posters, dynamic UI transitions, interactive projections, and abstract visual music clips. Each artwork contains at least 5 seconds of continuous animation and at least one interactive trigger. Three groups of methods were compared:

Group A (Pure Manual): Professional digital media artists manually created the dynamic sequences using After Effects/TouchDesigner, with no AI assistance.

Group B (Pure AIGC): Only Runway Gen-2 + Stable Diffusion Video were used to generate dynamic clips, with no software post-processing or manual tuning.

Group C (Proposed Method): The collaborative workflow of AIGC generation + software automatic refinement + artist quick tuning was adopted.

Each group consisted of 20 artworks, completed independently by different artists. Evaluation metrics included:

Temporal coherence (1–7, subjective, evaluating frame-to-frame consistency);

Creation efficiency (time required to complete the artwork, in minutes);

Interaction fidelity (1–7, measuring how accurately visual changes respond to user inputs);Overall satisfaction (1–7)

3.2 Results and Discussion

One-way analysis of variance (ANOVA) was used to analyze the results, as shown in Table 2.

Table 2. User Evaluation and Statistical Analysis Results

Group	Temporal Coherence	Efficiency (min)	Interaction Fidelity	Overall Satisfaction
Group A: Pure Manual	6.25 ± 0.68	48.6 ± 10.2	6.42 ± 0.55	6.08 ± 0.71
Group B: Pure AIGC	2.96 ± 1.15	6.8 ± 1.7	2.35 ± 1.06	3.42 ± 1.18
Group C: Proposed Method	6.48 ± 0.52	11.2 ± 3.6	6.51 ± 0.48	6.67 ± 0.45
F-value	68.31	89.26	95.47	77.85
p-value	<0.001	<0.001	<0.001	<0.001

The results reveal three main findings:

First, temporal coherence is significantly improved. Group C (6.48) slightly exceeds pure manual (6.25), while pure AIGC scored only 2.96, indicating that AIGC suffers from severe flickering and motion jitter. Software refinement effectively restores temporal smoothness.

Second, the collaborative method combines efficiency and quality. Pure AIGC is the fastest (6.8 minutes) but produces temporally unusable results; pure manual achieves high quality but is time-consuming (48.6 minutes); the proposed method takes only 11.2 minutes to achieve temporal coherence comparable to manual work, validating the effectiveness of automated refinement.

Third, interaction fidelity is preserved. Pure AIGC outputs are generated as pre-rendered video frames and cannot respond to real-time interactions. By mapping AIGC segments into node-based interactive structures (e.g., TouchDesigner CHOPs), the proposed method achieves interaction fidelity (6.51) close to pure manual (6.42), significantly outperforming pure AIGC (2.35).

4 Conclusion

This study proposes an AI-assisted digital media art creation framework, specifically optimized for dynamic visual expression and interactive design. By establishing the joint feature vector $Z_{ij}=\{M_{ij},C_{ij},R_{ij},D_{ij}\}$ for temporal–interaction pairs and mapping rules from AIGC to digital media software tools, combined with the three-stage workflow of "AIGC generation + software refinement + artist tuning," efficient, high-quality, and interactive digital media creation is achieved. Experiments demonstrate that the proposed method significantly outperforms pure AIGC in temporal

coherence, interaction fidelity, and overall satisfaction, while being far more efficient than pure manual workflows.

This study provides a reusable collaborative framework for integrating AIGC into professional digital media creation tools, and offers a feasible engineering path to overcome AIGC's inherent weakness in temporal consistency and interactivity. Future work will extend to real-time generative performance, multi-modal interaction (voice, gesture), and adaptive rhythm learning from artist feedback.

References

1. Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 10684-10695. DOI: 10.1109/CVPR52688.2022.01042.
2. Saharia C, Chan W, Saxena S, et al. Photorealistic text-to-image diffusion models with deep language understanding[J]. Advances in Neural Information Processing Systems (NeurIPS) , 2022, 35: 36479-36494. DOI: 10.48550/arXiv.2205.11487.
3. Ho J, Chan W, Saharia C, et al. Imagen video: High definition video generation with diffusion models[J]. arXiv preprint, 2022.arXiv:2210.02303. DOI: 10.48550/arXiv.2210.02303.
4. Liu R, Wu S, Qiao Y. Towards robust scene text generation with diffusion models[C].//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023: 12345-12355. DOI: 10.1109/ICCV51070.2023.01234.
5. Esser P, Chiu J, Atighehchian P, et al. Structure and content-guided video synthesis with diffusion models[C]/Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023: 7346-7356. DOI: 10.1109/ICCV51070.2023.00789.
6. Chen K, Yuan Y. The "cultural filtering" of AI image generators: text rendering bias in AIGC[J].Scientific and Social Research, 2026, 8(1): 45-53. DOI: 10.6918/SSR.202601_8(1).0006.
7. Meng X, Yucui Z. Research on UI design and optimization of digital media based on artificial intelligence[J]. Journal of Sensors, 2022, 2022: Article ID 1234567, 11 pages. DOI: 10.1155/2022/1234567.