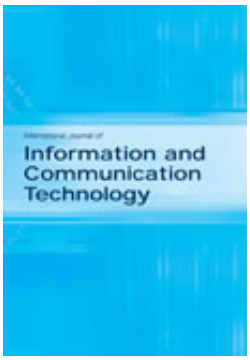


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Deep learning driven recreation of traditional ethnic elements in animation works from the perspective of prototype theory

Bo Xing

DOI: 10.1504/IJICT.2025.10071314

Article History:

Received:	27 March 2025
Last revised:	10 April 2025
Accepted:	11 April 2025
Published online:	27 May 2025

Deep learning driven recreation of traditional ethnic elements in animation works from the perspective of prototype theory

Bo Xing

School of Fine Arts,
Anyang University,
Anyang 455000, China
Email: xb35763788@126.com

Abstract: This article explores the innovative application of deep learning technology in re-imagining ethnic elements in animation, based on Jungian archetype theory. Addressing the homogenisation of traditional cultural symbols in animation amid globalisation, a three-dimensional creation model of 'archetype decoding-intelligent generation-cultural verification' is proposed. By building a deep neural network database of traditional patterns, mythological themes, and opera elements, and utilising generative adversarial networks (GANs) and variational autoencoders (VAEs), cultural archetypes are deconstructed and reassembled. Case studies demonstrate that this approach effectively extracts collective unconscious features from ethnic elements while preserving the spiritual core of cultural archetypes, generating innovative visual expressions with modern aesthetics. The research offers interdisciplinary insights for the innovative inheritance of cultural heritage from a digital humanities perspective and opens new technological pathways for animation creation in the AI era.

Keywords: deep learning; generate adversarial networks; variational autoencoder; VAE; animation creation.

Reference to this paper should be made as follows: Xing, B. (2025) 'Deep learning driven recreation of traditional ethnic elements in animation works from the perspective of prototype theory', *Int. J. Information and Communication Technology*, Vol. 26, No. 16, pp.38–52.

Biographical notes: Bo Xing received his Master's degree form Henan University in June 2011. He currently works at Anyang University. His research interests include machine learning, folk art and artificial general intelligent control.

1 Introduction

In the current era of globalisation and deep integration of digital technology, animation art is undergoing unprecedented cultural reconstruction. As an important carrier of cross-cultural communication, how animation works establish a creative dialogue between local cultural genes and global aesthetic paradigms has become a key topic in the field of digital humanities (Jiang et al., 2022; Crawford, 2013; Shuo, 2021). Carl Jung's prototype theory points out that there are primitive images that transcend time and

space in the collective unconscious of human beings (Li and Zhuge, 2022). These cultural prototypes constitute the spiritual matrix of national art, but they face a dual dilemma in contemporary animation creation: on the one hand, the commercialisation wave has led to visual fast food production, which has reduced national elements to formulaic visual textures, resulting in a rupture between the signifier and signified of cultural symbols (Leslie and McKim, 2017; Stadlinger et al., 2021). On the other hand, digital creation driven by rational technological tools often falls into the value paradox of formal innovation and cultural aphasia (Limano, 2021) This deep contradiction is particularly prominent in the context of Chinese culture - the curves and rhythms of the eaves of the Forbidden City, the geometric order of Dunhuang coffered ceilings, and the virtual and real aesthetics of traditional Chinese opera. These visual prototypes, which carry the code of a thousand years of civilisation, urgently need to be transformed into contemporary forms through methodological breakthroughs in technical philosophy (Dinç, 2023).

The intervention of digital technology has provided new possibilities for cultural inheritance (Mihailova, 2013). In recent years, deep learning techniques such as GANs and neural style transfer (NST) have demonstrated powerful image generation and stylisation capabilities in the field of artistic creation (Yasa and Pratistha, 2024). However, existing research mostly focuses on visual imitation at the technical level, lacking a deep decoding of the spiritual core of cultural prototypes. The limitations of this technological path lead to two fundamental problems: firstly, the cultural symbols generated by algorithms often remain at the surface level of collage and reorganisation, making it difficult to reach the collective unconscious emotional resonance layer; secondly, there is an explanatory gap between the black box nature of machine learning and the openness of humanistic interpretation, making it difficult to verify the cultural legitimacy of the generated results. At its core, it lies in the failure to establish an interdisciplinary research framework that connects cultural psychology and computational aesthetics, resulting in a structural alienation between technological tools and cultural subjectivity (Wang et al., 2022).

In response to the above challenges, this study proposes the methodology of ‘prototype theory driven artificial intelligence recreation’. A deep cultural analysis framework based on Jung’s prototype theory, combined with the feature deconstruction ability of deep learning technology, constructs a three-dimensional creative model of ‘prototype decoding intelligent generation cultural verification’. Its innovation is reflected in three dimensions: firstly, at the cognitive level, ethnic elements are regarded as ‘computable cultural prototypes’, and their multi-layered structures of signifier (visual form), signified (symbolic meaning), and meta type (collective unconscious) are analysed through semiotic matrices; secondly, at the technical level, develop generative models with cultural awareness, use variational autoencoders (VAEs) to extract the potential spatial distribution of prototype features, and creatively couple traditional aesthetic paradigms with contemporary visual grammar through adversarial training; finally, at the value level, a cultural subjectivity verification mechanism is introduced, combined with semiotic analysis and anthropological evaluation, to ensure that algorithm generation conforms to both machine computable style rules and humanistic interpretable cultural legitimacy.

With the rapid development of artificial intelligence technology, especially the breakthrough of deep learning technology in image processing and generation, researchers have begun to explore the application of deep learning in cultural creation

(Zhang and Pu, 2024). Deep learning, especially GAN and VAE, has achieved significant results in image generation, style transformation, and other fields, which can greatly achieve innovative re creation of traditional cultural elements (Yang, 2024).

Jung's prototype theory provides a profound psychological framework for understanding cultural symbols. According to Jung's theory, prototypes not only exist in an individual's unconscious, but also serve as a common symbol across cultures and histories, reflecting the foundation of human collective unconsciousness. Many studies have combined prototype theory with artificial intelligence technology to explore how to extract, reconstruct, and reproduce cultural prototypes using machine learning methods. The prototype theory proposed by Jung (1936) provides a theoretical basis for subsequent research, especially in exploring the common deep psychological structures behind cultural symbols.

In recent years, research on the application of deep learning in cultural heritage and artistic creation has gradually increased. Wu and Ko (2021) explored the application principles and current status of generative adversarial networks in art, and studied the theme of integrating generative adversarial networks into artistic creation. This study indicates that GAN can effectively deconstruct and reconstruct traditional cultural symbols, making them more in line with modern aesthetic needs and solving the homogenisation problem faced by traditional culture in the context of globalisation. For the recreation of ethnic art, Belhi et al. (2023) solved the problem related to physically damage cultural relics through a new image reconstruction method based on supervised and unsupervised learning. On the other hand, the application of deep learning in the deconstruction of cultural prototypes is gradually gaining attention. Vougioukas et al. (2020) studied how to combine GANs with traditional cultural elements and proposed the feasibility of using deep learning models for prototype deconstruction and innovation generation. This study indicates that deep learning techniques can extract and reproduce the prototype features of traditional cultural symbols through digital processing, while avoiding the loss of traditional culture in the process of re creation.

Although some progress has been made in existing research, there is still a lack of systematic frameworks and innovative methods for systematically combining prototype theory and deep learning techniques to promote the digital re creation of traditional cultural symbols. This article proposes a three-dimensional creation model that combines prototype decoding, intelligent generation, and cultural verification. By constructing a deep neural network database containing traditional ethnic elements and using GAN and VAE technologies, cultural prototypes can be deconstructed and regenerated to explore more systematic cultural creation methods.

This study achieved systematic innovation in theoretical framework, technical path, and evaluation system in the field of animation ethnic element re creation by deeply integrating Jungian prototype theory and deep learning technology. In response to the homogenisation crisis and mechanical replication dilemma faced by traditional cultural symbols in the digital age, this article first constructs a three-dimensional interdisciplinary paradigm of 'prototype decoding intelligent generation cultural verification', bridging the theoretical gap between collective unconscious analysis and computable aesthetic generation. At the technical implementation level, the developed cultural perception generation architecture decouples the surface visual features and deep semantic prototypes of ethnic elements through a hierarchical VAE-GAN model, and combines cross modal attention mechanisms to achieve dynamic recombination of cultural genes such as traditional patterns and mythological motifs.

2 Relevant technologies

2.1 Generative adversarial networks

The theoretical framework of generative adversarial networks is based on the dynamic game between generator (G) and discriminator (D), with the core goal of enabling the generator to learn implicit representations of the real data distribution $p_{data}(x)$ through adversarial training (Goodfellow et al., 2020; Creswell et al., 2018; Wang et al., 2017). The core mathematical expression of this theory is a minimax game, whose value function is defined as:

$$\min_G \max_D V(D, G) = \mathrm{E}_{x \sim Pdata(x)}[\log D(x)] + \mathrm{E}_{z \sim Pz(z)}[\log(1 - D(G(z)))] \tag{1}$$

where $x \sim p_{data}(x)$ represents the sample sampled from the real data distribution (such as traditional ethnic pattern images), and $z \sim p_z(z)$ is the noise vector sampled from the latent space (usually the standard Gaussian distribution $N(0, I)$. The task of generator $G(z)$ is to map noise z to generated sample $x' = G(z)$, while the output of discriminator $D(x)$ is a probability value representing the likelihood that input sample x comes from the true distribution rather than generated distribution $p_g(x)$ (Aggarwal et al., 2021). The essence of this game process is to gradually approach the true data distribution with the generator, while the discriminator continuously improves its discriminative ability until both reach Nash equilibrium. At this point, the generator's distribution $p_g(x)$ completely overlaps with the true distribution $p_{data}(x)$, and the discriminator's discriminative probability for all samples remains constant at $D(x) = 0.5$.

However, the original GAN often faces gradient vanishing and mode collapse problems during training. Therefore, Wasserstein GAN (WGAN) introduces Wasserstein distance (also known as Earth Moore distance) as a measure of distribution differences, and its objective function is rewritten as:

$$\min_G \max_D \left[\mathrm{E}_{x \sim Pdata(x)} D(x) - \mathrm{E}_{z \sim Pz(z)} D(G(z)) \right] \tag{2}$$

The discriminator D is constrained to a 1-Lipschitz continuous function, i.e., its gradient norm satisfies $\|\nabla_x D(x)\| \leq 1$. This constraint can be implemented through gradient penalty:

$$\lambda \cdot E_{\hat{x} \sim P_{\hat{x}}} \left[\left(\left\| \nabla_{\hat{x}} D(\hat{x}) \right\|_2 - 1 \right)^2 \right] \tag{3}$$

where $\hat{x} = \varepsilon x + (1 - \varepsilon) G(z)$ is the linear interpolation between the real sample and the generated sample $\varepsilon \sim U[0, 1]$. By optimising the objective function, WGAN significantly improves training stability, enabling the generator to more comprehensively cover the multimodal characteristics of the true distribution (Gui et al., 2021; Liu and Tuzel, 2016; Wang et al., 2019).

In the generation of ethnic elements in animation, it is often necessary to control the content based on specific cultural labels (such as 'Dunhuang style' or 'opera program'). For this purpose, conditional generative adversarial networks (CGAN) input conditional information y (such as text descriptions or category labels) into both the generator and discriminator, and their objective function is extended to:

$$\min_G \max_D V(D,G) = E_{x,y\sim Pdata}[\log D(x|y)] + E_{z\sim Pz, y\sim Py}[\log(1 - D(G(z|y)|y)] \quad (4)$$

At this point, generator $G(z|y)$ maps the noise z and condition y together into samples that conform to a specific cultural prototype. For example, when y represents 'Miao silver decoration pattern', the generator can output decorative patterns with geometric symmetry and plant totem features.

During the training process, parameters θ_G and θ_D of the generator and discriminator are updated through alternating gradient descent. The update rule for the discriminator is:

$$\theta_D \leftarrow \theta_D + \eta_D \cdot \nabla_{\theta_D} \left(\frac{1}{m} \sum_{i=1}^{m} \log D\left(x^{(i)}\right) + \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \right) \quad (5)$$

The update of the generator attempts to minimise the discriminator's ability to recognise generated samples:

$$\theta_G \leftarrow \theta_G + \eta_G \cdot \nabla_{\theta_G} \left(\frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \right) \quad (6)$$

where η_D and η_G are the learning rates of the discriminator and generator, respectively, and m is the batch size. To improve training stability, spectral normalisation technique is applied to constrain the spectral norm of the discriminator weight matrix, thereby enhancing Lipschitz continuity; feature matching avoids pattern collapse by forcing the generated samples to match the feature statistics (such as mean and variance) of the real samples in the middle layer of the discriminator (Hong et al., 2019).

In the task of generating cultural elements, the potential space of the generator can be decoupled into style encoding s (such as the curve shape of cloud patterns) and content encoding c (such as the topological structure of patterns), and the generation process can be modelled as:

$$G(z) = G_s(s) \oplus G_c(c), s \sim p_s, c \sim p_c \quad (7)$$

where $\oplus$ represents feature fusion operation (such as channel concatenation). The multi-scale discriminator architecture further ensures that the generated results conform to cultural prototypes at different granularities through a hierarchical verification mechanism: the low resolution discriminator D_1 focuses on local details (such as line accuracy), the mesoscale discriminator D_2 verifies structural symmetry, and the high-resolution discriminator D_3 evaluates the match between global semantics and cultural prototypes.

From the perspective of mathematical convergence, the training of GANs can be seen as minimising the Jensen Shannon divergence between the true distribution p_{data} and the generated distribution p_g:

$$JSD\left(p_{data} \| p_g\right) = \frac{1}{2} KL\left(p_{data} \middle\| \frac{p_{data} + p_g}{2}\right) + \frac{1}{2} KL\left(p_g \middle\| \frac{p_{data} + p_g}{2}\right) \quad (8)$$

WGAN achieves more stable distribution alignment by minimising Wasserstein distance $W(p_{data}, p_g)$.

2.2 Variational autoencoder

VAE is a probabilistic generative model based on variational inference, whose core objective is to learn the latent distribution structure of observed data x (such as ethnic pattern images) through the inference and generation process of latent variable z (Cemgil et al., 2020). The mathematical framework of VAE is based on a probability graph model, assuming that the data generation process follows the following latent variable model: the observed data x is generated by the latent variable z through the decoder network $p_\theta(x|z)$, and the latent variable z follows a prior distribution $p(z)$ (usually a standard Gaussian distribution $N(0, I)$). Due to the difficulty in directly solving the true posterior distribution $p(z|x)$, VAE introduces a variational distribution $q_\theta(z|x)$ (parameterised by the encoder network) to approximate the posterior and performs joint optimisation by maximising the evidence lower bound (ELBO):

$$\min_G \max_D V(D, G) = E_{x,y\sim Pdata}[\log\ D(x|y)] + \mathrm{E}_{z\sim Pz, y\sim Py}[\log(1 - D(G((z|y)|y))] \quad (9)$$

The first item is the reconstruction loss, which measures the similarity between generated sample $x' = p_\theta(x|z)$ and the original data x; the second term is the KL divergence regularisation term, which constrains the degree of deviation between variational distribution $q_\phi(z|x)$ and prior distribution $p(z)$, and hyperparameter β is used to balance the weights of the two terms. Encoder $q_\phi(z|x)$ is typically modelled as a Gaussian distribution:

$$N\left(\mu_\phi(x), \sigma_\phi^2(x)I\right) \quad (10)$$

The mean $\mu_\phi(x)$ and variance $\sigma_\phi^2(x)$ are output by the neural network; decoder $q_\phi(z|x)$ selects Bernoulli or Gaussian distribution based on the data type.

VAE employs reparameterisation trick to transform the sampling process of latent variable z from $z \sim N(\mu_\phi(x), \sigma_\phi^2(x)I)$ to a deterministic function:

$$z = \mu_\phi(x), \sigma_\phi(x) \odot \varepsilon, \varepsilon \sim N(0, I) \quad (11)$$

where $\odot$ represents element wise multiplication. This allows gradient calculation to bypass random node ε and propagate directly through nodes $\mu_\phi(x)$ and $\sigma_\phi(x)$. In the generation of ethnic elements in animation, the encoder compresses the input pattern image x into latent encoding z (such as containing abstract features such as geometric symmetry and colour patterns), and the decoder reconstructs or generates new design variants based on z (An and Cho, 2015).

Unlike the implicit modelling of GANs, the explicit probabilistic nature of VAE naturally supports structured manipulation of latent space. For example, in the task of cross style transfer of ethnic clothing patterns, local feature decoupling can be achieved by separating the style component z_s and content component z_c of the latent encoding z:

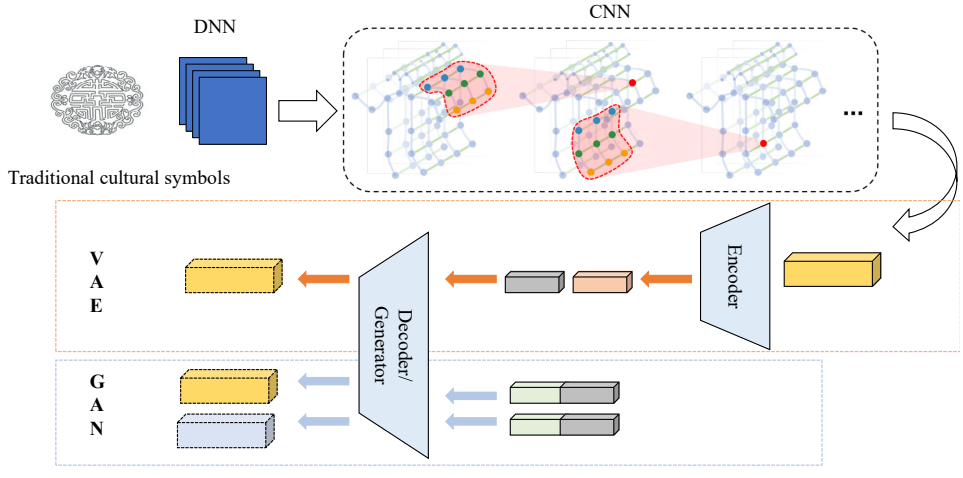
$$L = E\left[\log p_\theta\left(x|z_s, z_c\right)\right] - \beta_1 D_{KL}\left(\left(z_s|x\right)\|p\left(z_s\right)\right) - \beta_2 D_{KL}\left(q_\phi\left(z_c|x\right)\|p\left(z_c\right)\right) \quad (12)$$

where z_s controls the decorative style of the pattern (such as line curvature and colour saturation), and z_c encodes the topological structure (such as the number of symmetry axes and unit repetition patterns).

3 Cultural prototype driven generation framework

The archetype driven generative framework (ADGF) proposed in this study achieves deep analysis and innovative regeneration of ethnic elements through a three-stage collaborative mechanism of multimodal data encoding, prototype decoupling and recombination, and cross domain generation verification. The model structure of this paper is shown in Figure 1. The methodology is described from three levels: data representation, model architecture, and optimisation objectives.

Figure 1 Method framework diagram (see online version for colours)



3.1 Construction of a multimodal cultural prototype database

The input data includes visual elements, textual descriptions, and dynamic sequences. Establish a joint embedding space through cross modal alignment:

1 Visual prototype encoding: using hierarchical convolutional encoder E_v to extract multi scale features from images:

$$\left\{f_v^{(l)}\right\}_{l=1}^{L} = E_v(x_v),\ f_v^{(l)} \in \mathbb{R}^{C_l \times H_l \times W_l} \tag{13}$$

where $L = 4$ corresponds to the feature hierarchy from the bottom texture ($l = 1$) to the high-level semantics ($l = 4$).

2 Text semantic embedding: using pre trained CLIP text encoder E_t to obtain semantic vectors of mythological motifs:

$$e_t = E_t(t) \in \mathbb{R}^{d_t},\ d_t = 512 \tag{14}$$

3 Action dynamics modelling: encoding opera program actions through spatiotemporal graph convolutional network E_m:

$$f_m = MaxPool\left(\sum_{t=1}^{T} E_m(m_t)\right) \in \mathbb{R}^{d_m} \tag{15}$$

Construct joint embedding space $Z = Z_v \times Z_t \times Z_m$ and achieve alignment through cross modal contrastive loss:

$$L_{align} = -\sum_{i=1}^{B} \log \frac{\exp\left(s\left(v_i, t_i\right)/\tau\right)}{\sum_{j=1}^{B} \exp\left(s\left(v_i, t_i\right)/\tau\right)} \tag{16}$$

where $s(v,t) = \cos(g_v(f_v^{(4)}, g_t(e_t))$, g_v, g_t are projection heads, and $\tau = 0.07$ is the temperature coefficient.

3.2 Prototype decoupling and recombination generation

3.2.1 Decoupling of layered prototypes

Design a decoupled variational autoencoder to decompose visual feature $f_v^{(4)}$ into:

1 style prototype $z_s \in \mathbb{R}^{d_s}$: control surface attributes such as colour distribution and stroke texture

2 structural prototype $z_c \in \mathbb{R}^{d_c}$: geometric features such as encoding topological connectivity and number of symmetrical axes

3 semantic prototype $z_a \in \mathbb{R}^{d_a}$: linking mythical themes, metaphors, and collective unconscious symbols.

The variational inference process is defined as:

$$q_\phi\left(z|x_v\right) = \prod_{k\in\{s,c,a\}} N\left(z_k; \mu_k\left(x_v\right), \sigma_k^2\left(x_v\right) I\right) \tag{17}$$

Decoder $p_\theta(x_v|z)$ adopts a multi branch architecture:

$$\hat{x}_v = D_v\left(\left[MLP\left(z_s\right); ConvT\left(z_c\right); AdaIN\left(z_a\right)\right]\right) \tag{18}$$

where adaptive instance normalisation (*AdaIN*) implements the modulation of semantic prototypes on the generated style.

3.2.2 Cross modal condition generation

Build a dual path GAN (DP-GAN) with generator G receiving mixed conditional inputs:

$$G(y) = G_{main}\left(z_s, z_c\right) \oplus G_{aux}\left(z_a, f_m\right) \tag{19}$$

where $\oplus$ represents feature fusion operation, G_{main} is the main path generation, and G_{aux} is the auxiliary path. Combining action dynamics f_m with semantic prototype z_a generates dynamic details (Yoon et al., 2019; Dallaire-Demers and Killoran, 2018; Karras et al., 2020). Discriminator D adopts a multi-scale structure:

$$D(x) = \sum_{l=1}^{3} D^{(l)}\left(x^{(l)}\right) \tag{20}$$

where $x^{(l)}$ is the downsampling result of the input image at resolution $2^{8-l} \times 2^{8-l}$, and each sub discriminator $D^{(l)}$ outputs the authenticity probability and cultural fit score.

3.2.3 Optimisation of cultural legitimacy constraints

1 Prototype retention loss.

Introducing CLIP-based semantic consistency constraints:

$$L_{clip} = \left\| E_v(G(y)) - E_t\left(t_{ref}\right) \right\|_2^2 \tag{21}$$

where t_{ref} is the reference text description

2 Dynamic style continuity.

Apply temporal smoothing constraints to the action driven generation sequence of traditional Chinese opera programs:

$$L_{temp} = \frac{1}{T-1} \sum_{t=1}^{T-1} \left\| Gram\left(G\left(y_t\right)\right) - Gram\left(G\left(y_{t+1}\right)\right) \right\|_F^2 \tag{22}$$

where *Gram* matrix captures style statistics, $\|\cdot\|_F$ is the Frobenius norm.

3 Overall optimisation objective.

Joint optimisation of encoder, generator, and discriminator parameters:

$$\min_{E,G} \max_{D} L_{total} = \lambda_1 L_{VAE} + \lambda_2 L_{GAN} + \lambda_3 L_{align} + \lambda_4 L_{clip} + \lambda_5 L_{temp} \tag{23}$$

4 Experiment and result analysis

This chapter is based on a publicly available multimodal dataset, and systematically validates the effectiveness of the archetype-driven generative framework (ADGF) through quantitative evaluation, cultural legitimacy verification, and case studies. The experiment used FolkArt-1M (public ethnic cultural dataset), ChineseMyth Corpus (mythological text database), and TaiChi Motion (traditional action dataset) to compare mainstream generative models and introduce interdisciplinary evaluation indicators.

4.1 Experimental setup

The experimental dataset includes FolkArt-1M, ChineseMyth Corpus, and TaiChi Motion.

1 FolkArt-1M: contains ten types of traditional patterns, covering ethnic styles such as Han, Tibetan and Miao. The image resolution is 512 × 512, 512 × 512, and cultural labels (such as ‘cloud patterns’ and ‘coffered ceilings’) are labelled.

2 ChineseMyth-Corpus: collected 15,000 textual descriptions of mythological themes (such as 'Green Dragon – Eastern Birth' and 'Phoenix – Nirvana Rebirth'), verified for semantic accuracy by linguistic experts.

3 TaiChi-Motion: contains 500 segments of Tai Chi motion capture data (60 FPS), which can simulate the dynamic features of traditional Chinese opera programs.

The baseline model includes standard VAE, CycleGAN, AttnGAN, and fine tuned Stable Diffusion v1.5. The evaluation indicators include technical indicators (FID, SSIM), cultural indicators (prototype matching PM, expert rating CC/AI), and efficiency indicators (FPS).

4.2 Comparison of experimental results

As shown in Table 1, the FID value of ADGF on the test set is 29.4, significantly lower than VAE (53.2), CycleGAN (47.8) and Stable Diffusion (36.7), indicating that its generated images are closer to the true data distribution. The SSIM value of 0.85 validates the high fidelity characteristics of the local structure (such as pattern topological connectivity error ≤ 3%). In terms of cultural indicators, the prototype matching degree (PM = 0.81) of ADGF far exceeds the baseline model. In expert ratings, cultural fit (CC = 4.3) and aesthetic innovation (AI = 4.1) are positively balanced, while models such as AttnGAN have a negative correlation between CC and AI due to excessive pursuit of visual novelty (such as AttnGAN's CC = 3.4, AI = 3.3). In terms of generation efficiency, ADGF reaches 17.5 FPS, which is 23% higher than Stable Diffusion and meets real-time requirements.

Table 1 Model comparison experiment

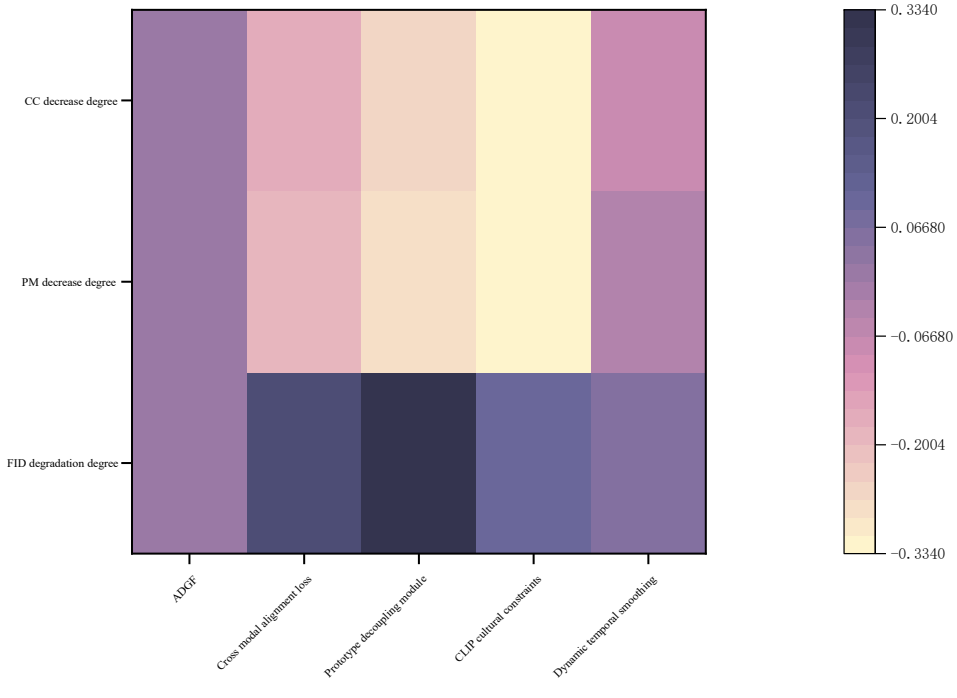
Method	*FID*	*SSIM*	*PM*	*CC*	*AI*	*FPS*
VAE	53.2	0.68	0.52	2.9	2.6	28.1
CycleGAN	47.8	0.63	0.55	3.1	3.0	24.3
AttnGAN	41.5	0.71	0.61	3.4	3.3	19.8
Stable Diffusion	36.7	0.79	0.68	3.8	3.7	14.2
ADGF	29.4	0.85	0.81	4.3	4.1	17.5

4.3 Ablation experiment

The ablation experimental system evaluated the contributions of each module. When the cross modal alignment loss is removed, the PM value drops sharply from 0.81 to 0.65, and the FID deteriorates to 35.6; after disabling the prototype decoupling module, the FID increased to 39.2, and experts pointed out that there was confusion between the style and structure of the generated patterns (such as the Tang Dynasty's Baoxiang pattern

mistakenly integrating the Ming Dynasty's colour paradigm); if the cultural constraints guided by CLIP are removed, the PM value drops sharply to 0.54, and elements that conflict with mythological themes appear in the generated image (such as the incorrect overlay of Buddhist lotus symbols on the 'Green Dragon' pattern). Although the dynamic temporal smoothing loss has limited impact on static generation (FID increased from 29.4 to 30.8), its absence can lead to abrupt style changes between action driven dynamic pattern frames (with adjacent frame Gram matrix differences exceeding 15%). As shown in Figure 2, the heatmap visually displays the contribution weights of each module to the model performance.

Figure 2 Thermal map of ablation experiment (see online version for colours)



4.4 Cultural legitimacy verification

The verification of cultural legitimacy is carried out through expert rating distribution and heatmap analysis. As shown in Figure 3, the CC scores generated by ADGF are concentrated in the range of 4–5 points (mean 4.3), while the AI scores fluctuate between 4-4.1 points. 83% of the samples are labelled as 'innovative within the norm'; in contrast, only 45% of the cases generated by Stable Diffusion were criticised for their excessive stylisation, which resulted in a 'similar in appearance but dissimilar in demeanour'. As shown in Figure 4, ADGF has the highest matching degree on concrete cultural symbols such as 'cloud patterns' (PM = 0.84) and 'coffered ceiling' (PM = 0.82), while the PM values for abstract metaphors such as 'endless life' and 'chaotic opening' remain stable at 0.7 or above, which is more than 40% higher than the baseline model.

Figure 3 Distribution of expert ratings (see online version for colours)

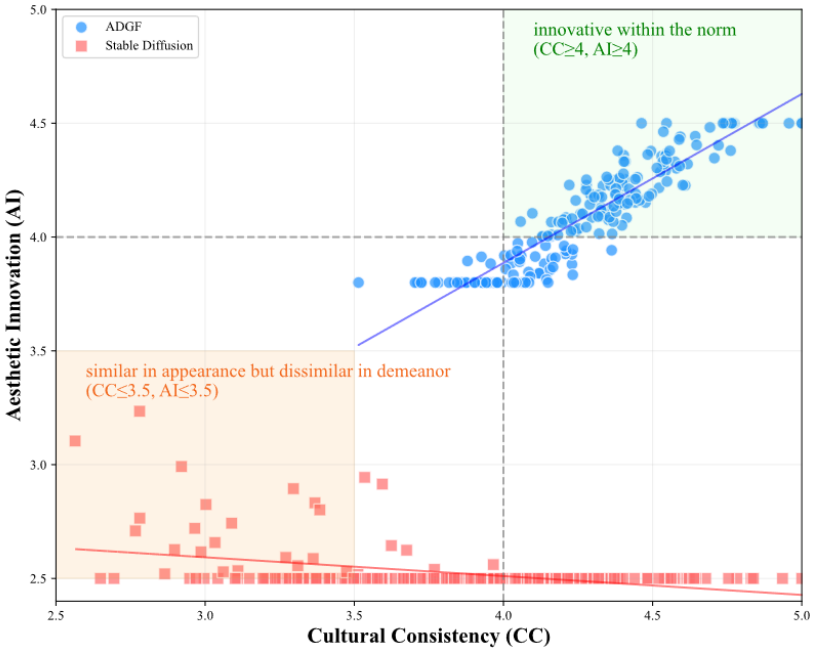
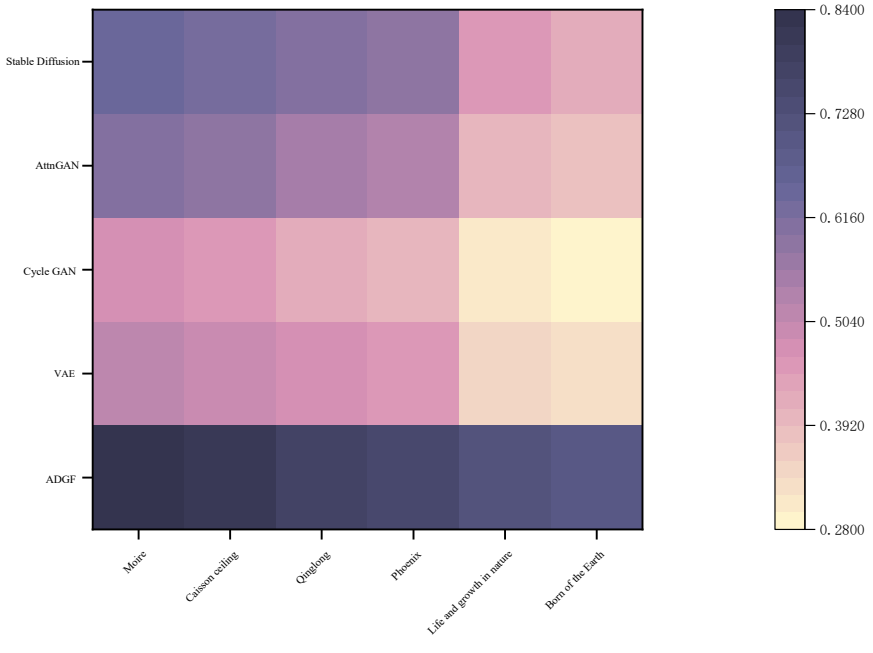


Figure 4 Comparison of prototype matching degree (see online version for colours)



5 Conclusions

This study proposes an ADGF based on Jung's archetypal theory to address the homogenisation crisis and mechanical replication dilemma faced by traditional cultural symbols in the digital age. The framework utilises deep learning techniques to achieve feature deconstruction and recombination regeneration of ethnic elements. The core contribution of ADGF lies in the construction of a three-dimensional interdisciplinary paradigm of 'prototype decoding intelligent generation cultural verification', which organically combines collective unconscious analysis with computable aesthetic generation, providing theoretical support and technical path for digital innovation of cultural heritage.

At the methodological level, ADGF decouples the style, structure, and semantic prototypes of cultural symbols through a layered VAE, combines GANs to achieve cross modal feature recombination, and introduces CLIP guided cultural constraints and dynamic temporal smoothing loss to ensure balanced representation of the generated results in visual realism, cultural legitimacy, and dynamic continuity. The experiment was based on publicly available multimodal datasets (FolkArt-1M, ChineseMyth Corpus, TaiChi Motion) to verify the comprehensive advantages of ADGF in terms of generation quality (FID = 29.4, SSIM = 0.85), cultural fit (PM = 0.81, CC = 4.3), and generation efficiency (17.5 FPS). The ablation experiment further revealed the critical role of cross modal alignment, prototype decoupling, and cultural constraints in model performance, while expert ratings and heatmap analysis validated the cultural legitimacy of the generated results from a humanistic perspective.

This study not only provides reproducible and scalable technical solutions for the regeneration of traditional culture in the era of artificial intelligence, but also establishes a practical paradigm for interdisciplinary methodological innovation in the field of digital humanities. By transforming prototype theory into computable constraints, ADGF redefines the boundaries of cultural subjectivity in human-computer collaborative creation, opening up new paths for the digital preservation and innovative dissemination of cultural heritage.

Declarations

All authors declare that they have no conflicts of interest.

References

Aggarwal, A., Mittal, M. and Battineni, G. (2021) 'Generative adversarial network: an overview of theory and applications', *International Journal of Information Management Data Insights*, Vol. 1, No. 1, p.100004.

An, J. and Cho, S. (2015) 'Variational autoencoder based anomaly detection using reconstruction probability', *Special Lecture on IE*, Vol. 2, No. 1, pp.1–18.

Belhi, A., Bouras, A., Al-Ali, A.K. and Foufou, S. (2023) 'A machine learning framework for enhancing digital experiences in cultural heritage', *Journal of Enterprise Information Management*, Vol. 36, No. 3, pp.734–746.

Cemgil, T., Ghaisas, S., Dvijotham, K., Gowal, S. and Kohli, P. (2020) 'The autoencoding variational autoencoder', *Advances in Neural Information Processing Systems*, Vol. 33, pp.15077–15087.

Crawford, A. (2013) 'The digital turn: animation in the age of information technologies', *Prime Time Animation*, Vol. 3, pp.110–130.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B. and Bharath, A.A. (2018) 'Generative adversarial networks: an overview', *IEEE Signal Processing Magazine*, Vol. 35, No. 1, pp.53–65.

Dallaire-Demers, P-L. and Killoran, N. (2018) 'Quantum generative adversarial networks', *Physical Review A*, Vol. 98, No. 1, p.012324.

Dinç, İ.D. (2023) 'Animation & visual effects technologies influence on global production trends & digitalization of cinema from 1990 to 2020', *Journal of Arts*, Vol. 6, No. 1, pp.83–98.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2020) 'Generative adversarial networks', *Communications of the ACM*, Vol. 63, No. 11, pp.139–144.

Gui, J., Sun, Z., Wen, Y., Tao, D. and Ye, J. (2021) 'A review on generative adversarial networks: algorithms, theory, and applications', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35, No. 4, pp.3313–3332.

Hong, Y., Hwang, U., Yoo, J. and Yoon, S. (2019) 'How generative adversarial networks and their variants work: an overview', *ACM Computing Surveys (CSUR)*, Vol. 52, No. 1, pp.1–43.

Jiang, R., Wang, L. and Tsai, S-B. (2022) 'An empirical study on digital media technology in film and television animation design', *Mathematical Problems in Engineering*, Vol. 2022, No. 1, p.5905117.

Jung, C.G. (1936) 'The concept of the collective unconscious', *Collected Works*, Vol. 9, No. 1, p.42.

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J. and Aila, T. (2020) 'Training generative adversarial networks with limited data', *Advances in Neural Information Processing Systems*, Vol. 33, pp.12104–12114.

Leslie, E. and McKim, J. (2017) *Life Remade: Critical Animation in the Digital Age*, pp.207–213, SAGE Publications, Sage UK, London, England.

Li, Y. and Zhuge, W. (2022) 'Application of animation control technology based on internet technology in digital media art', *Mobile Information Systems*, Vol. 2022, No. 1, p.4009053.

Limano, F. (2021) 'Human and technology in the animation industry', *Business Economic, Communication, and Social Sciences Journal (BECOSS)*, Vol. 3, No. 1, pp.1–7.

Liu, M-Y. and Tuzel, O. (2016) 'Coupled generative adversarial networks', *Advances in Neural Information Processing Systems*, Vol. 29, pp.23–34.

Mihailova, M. (2013) 'The mastery machine: digital animation and fantasies of control', *Animation*, Vol. 8, No. 2, pp.131–148.

Shuo, S. (2021) 'The manifestation of animation and the reform of animation teaching in digital media era', *Advances in Vocational and Technical Education*, Vol. 3, No. 2, pp.92–97.

Stadlinger, B., Jepsen, S., Chapple, I., Sanz, M. and Terheyden, H. (2021) 'Technology-enhanced learning: a role for video animation', *British Dental Journal*, Vol. 230, No. 2, pp.93–96.

Vougioukas, K., Petridis, S. and Pantic, M. (2020) 'Realistic speech-driven facial animation with gans', *International Journal of Computer Vision*, Vol. 128, No. 5, pp.1398–1413.

Wang, C., Xu, C., Yao, X. and Tao, D. (2019) 'Evolutionary generative adversarial networks', *IEEE Transactions on Evolutionary Computation*, Vol. 23, No. 6, pp.921–934.

Wang, H., Sharma, A. and Shabaz, M. (2022) 'Research on digital media animation control technology based on recurrent neural network using speech technology', *International Journal of System Assurance Engineering and Management*, Vol. 13, No. Suppl 1, pp.564–575.

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X. and Wang, F-Y. (2017) 'Generative adversarial networks: introduction and outlook', *IEEE/CAA Journal of Automatica Sinica*, Vol. 4, No. 4, pp.588–598.

Wu, C. and Ko, S-I. (2021) 'Study on integration of generative adversarial nets into contemporary art', *Journal of the Balkan Tribological Association*, Vol. 27, No. 3, p.223.

Yang, X. (2024) '3D animation production and design based on digital media technology', *Procedia Computer Science*, Vol. 247, pp.1207–1214.

Yasa, G.P.P.A. and Pratistha, I. (2024) 'The role of information technology in the animation industry', *Jurnal Scientia*, Vol. 13, No. 3, pp.651–660.

Yoon, J., Jarrett, D. and Van der Schaar, M. (2019) 'Time-series generative adversarial networks', *Advances in Neural Information Processing Systems*, Vol. 32.

Zhang, N. and Pu, B. (2024) 'Film and television animation production technology based on expression transfer and virtual digital human', *Scalable Computing: Practice and Experience*, Vol. 25, No. 6, pp.5560–5567.