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

Visual Expression Characteristics of AI-Generated Images in Creative Design

Xian Si[1,*]

[1] Suzhou Art & Design Technology Institute Visual Communications Department, 210000 Jiangsu, China

SUMMARY: *This paper constructs an efficient image generation model based on a hybrid LG-MLP multilayer perceptron. First, we propose a hybrid generator architecture integrating convolutional neural networks and multilayer perceptrons to achieve high-quality image generation within the WGAN-GP framework. Furthermore, the LG-MLFormer network is designed, integrating a local-global MLP encoder with a cross-domain memory-enhanced decoder to effectively enhance visual feature fusion and language generation capabilities. Experiments on the MS COCO dataset demonstrate that the proposed model significantly outperforms mainstream baselines in image realism. Under HED edge conditions, the model achieves FID scores of 6.12 and 6.32 for consistent and inconsistent scenarios respectively, outperforming comparators like ControlNet (7.03, 8.41). Under Midas depth map conditions, FID further decreases to 4.46 and 4.54, demonstrating superior cross-condition generalization. The design group employing the LG-MLP method achieved average scores of 4.73, 4.87, and 4.75 in creativity, final product quality, and expert/user satisfaction, respectively, significantly surpassing the control group. The study demonstrates that the proposed generative model not only exhibits strong technical performance but also holds high application potential and practical value in real-world creative design.*

KEYWORDS: *AI-generated images; Hybrid Multi-Layer Perceptron; LG-MLP; Visual Representation*

1 Introduction

With the rapid advancement of artificial intelligence technology, AIGC (Artificial Intelligence Generated Content) has emerged as a pivotal application domain, triggering significant shifts across multiple industry ecosystems[1]. AIGC primarily refers to the automated generation of diverse content formats—including text, images, audio, and video—through technologies such as machine learning, natural language processing, and image generation[2, 3]. Its core technologies encompass complex algorithms like deep learning and Generative Adversarial Networks (GANs)[4]. To date, particularly since the release of the DeepSeek-R1 model by Hangzhou DeepSeek Research Institute of Artificial Intelligence Foundation Technology in China, which achieved performance comparable to OpenAI's O1 model, marking a milestone breakthrough for domestic large-scale models, the technological advancement of AIGC has sparked a wave of global attention both domestically and internationally.

Amid the global wave of digital transformation, artificial intelligence—a disruptive force spearheading technological revolution—is penetrating every industry at an unprecedented pace.

*sxhopes@163.com
https://doi.org/10.65102/is20261031
Published: 30 April, 2026

The creative design sector is naturally swept up in this technological upheaval. AI possesses a potent “leading goose effect” with strong spillover potential, empowering diverse industries and occupying a crucial strategic position in the global technology landscape [5]. Looking back over the past decade of graphic creative design, the dominant paradigm centered on hand-drawn illustrations and two-dimensional geometric forms has established a unique technical system. However, it has also revealed structural issues such as lengthy design cycles, pronounced homogenization of works, and a lack of regional cultural expression [6-8]. With iterative breakthroughs in data visualization technology and the continuous evolution of artificial intelligence algorithms, designers have begun leveraging AIGC's multi-source, cross-modal information fusion capabilities to reconstruct creative design workflows. This technological innovation not only reshapes designers' working paradigms but also opens unprecedented creative spaces within the design field, significantly transforming traditional creative design practices [9-11]. As AIGC technology deeply integrates into visual communication design, its value increasingly manifests in empowering creators to transcend stylistic boundaries. It propels the cross-disciplinary fusion of diverse art forms and fosters innovative expression, infusing fresh vitality into the generation mechanisms of visual creative works [12-14].

In recent years, the rapid advancement of AIGC technology has drawn increasing attention to its integrated development across various creative design fields. Reviewing extensive literature reveals that academic research primarily examines multiple dimensions of image generation technology itself, including content quality, ethical and legal considerations, and cross-industry applications [15]. Within the creative design domain, however, research is concentrated within art and design disciplines, placing greater emphasis on interdisciplinary integration and exploring the frontiers of technological ethics. Particular focus is given to image design concepts and diverse application methodologies, providing a conceptual framework for emerging fields like AIGC image generation [16].

We focus our attention on the field of creative design, where numerous researchers and institutions have already initiated studies in AIGC and achieved remarkable results in recent years. For instance, Bhattacharjee, G[17] proposed in their research that AIGC technology redefines artistic creation, photography, and many other domains of human creativity in ways that transcend reality. This presents both challenges and opportunities for artists to elevate their creative standards. Xu, L et al.[18] employed AIGC to optimize creative design, enhancing production quality and audience satisfaction through user-centered models and evaluation frameworks. Their research emphasizes interdisciplinary integration to elevate personalized cultural consumption. Furthermore, Ye, C et al. [19] prospectively propose AIGC technology to empower new media art creative design. By leveraging AI, machine learning, and big data mining, they rapidly discern user experiences and needs to further refine new media creative works. In practical creation, numerous renowned artists and institutions have begun applying AI to artistic domains. Zhang, W et al. [20] applied AIGC technology to cultural and creative product design using Macau's dragon dance cultural elements as a case study. This research demonstrates how AIGC technology can enhance creative design efficiency while promoting the utilization and innovation of cultural heritage. Liu, Q et al. [21] explored digital creative design pathways for Jiaodong Peninsula's marine folk culture using AIGC technology, aiming to advance cultural IP innovation and promotion through digital transformation and personalized innovation. Pan, S et al. [22] employed AIGC technology to establish a sustainable creative design framework for Yixing purple clay pottery, integrating cultural heritage preservation with innovation to enhance design diversity and address the creative limitations inherent in traditional craft design.

Regarding interdisciplinary research on artificial intelligence technology and creative design, Ploennigs, J and Berger, M[23] conducted a comparative study on the applicability of

three generative tools—Midjourney, DALL-E2, and Stable Diffusion—in architectural design. They noted that as technology continues to advance and mature, the integration of AI-generated content (AIGC) can significantly enhance designers' productivity and creativity. Zhou, X[24] introduced an interactive creative design framework integrating genetic algorithms with long short-term memory (LSTM) networks to enhance user experience and design efficiency for cultural and creative products. The study revealed significant advantages over traditional design methods in both user satisfaction and design efficiency. Lu, W et al.[25] designed an innovative personalized intelligent creative design system integrating culture, creative design, and computer science. By merging traditional aesthetics with advanced technology, it aims to enhance user engagement in cultural and creative fields while increasing the personalization of creative products. Tao, Y et al. [26] developed an AI-driven creative image generation framework named “AlFiligree,” designed for the visualization process in artistic creative design. It primarily generates realistic structures and supports diverse design scenarios. Wu, Q et al. [27] constructed the “ClothGAN” framework using generative adversarial networks and style transfer algorithms. Drawing creative inspiration from Dunhuang elements, they designed new clothing styles and aesthetics, with subsequent experiments confirming its effectiveness. In summary, while AIGC research predominantly focuses on technical characteristics and algorithmic models, current studies in graphic design concentrate primarily on image generation technology. There remains insufficient attention to the deep expressive capabilities required for diverse themes and information within the creative design discipline. Although AIGC research spans multiple disciplines, systematic investigations into its applicability in graphic creative design remain limited. Existing literature primarily explores foundational theoretical principles and methods superficially, failing to delve deeply into the multidimensional value of design disciplines enabled by technology. There has been no comprehensive analysis of implementation pathways, design principles, methodologies, or ethical values. Therefore, this research holds significant value and novelty by examining the application of AIGC technology in image creative design and its visual expression characteristics.

This paper addresses the challenges of unstable training, significant semantic noise, and high computational complexity in generative adversarial networks (GANs). It proposes an image generation method based on a hybrid multi-layer perceptron (MLP) architecture. By integrating convolutional neural networks with multi-layer perceptron modules within the WGAN-GP framework, it achieves high-quality image generation while significantly improving training stability and output quality. Furthermore, the LG-MLFormer architecture is designed, integrating a local-global MLP encoder with a cross-domain memory-augmented decoder. During encoding, LG-MLP achieves self-compensation of visual feature space information without introducing additional parameters. Simultaneously, it explores latent correlations between different visual regions within images to extract semantically rich visual features. During decoding, the Visual-Linguistic Memory-Augmented Attention (VLMA) module integrates visual and linguistic prior knowledge to generate word-by-word descriptive sentences. This model effectively addresses semantic noise and computational complexity challenges while supporting multi-level visual feature fusion and language generation tasks. Finally, from an artistic perspective, we analyze the attributes of image-generated art in terms of encoding versus decoding, dynamism versus staticity, and explore its potential for visual expression and humanistic value as a digital art form.

2 Image Generation Model Based on LG-MLP Multi-Layer Perceptron and Analysis of Its Artistic Characteristics

2.1 Image Generation Method Based on Hybrid Multi-Layer Perceptrons

2.1.1 Generative Model Algorithm Based on Hybrid Multi-Layer Perceptrons

In computer vision tasks, convolutional neural networks (CNNs) are the most commonly used models. However, with the rapid advancement of computational resources, models based on attention mechanisms (such as Transformers) or multilayer perceptrons (MLPs) (such as MLP-Mixer) have made it possible to surpass CNNs. While some research has applied Transformer architectures to image generation tasks, no work has yet applied MLP structures to this domain. Given the similarities between Transformer-based and MLP-based networks—such as their requirement for large training datasets and quadratic computational complexity—this section draws insights from Transformer-based generative adversarial networks (GANs). We construct a hybrid model where an MLP module serves as the generator and a convolutional neural network (CNN) module acts as the discriminator. This approach aims to validate the performance of MLP architectures in image generation tasks.

During the training of generative adversarial networks (GANs), the generator and discriminator require continuous iterative optimization to induce the generator to produce images as realistic as possible while enabling the discriminator to distinguish between real images and those generated by the model. However, training a high-quality GAN is no easy feat. Therefore, selecting the discriminator from the WGAN-GP architecture aids in stabilizing the training process, enabling the hybrid model based on multilayer perceptrons to produce higher-quality results. The generator model architecture designed in this paper, based on multilayer perceptrons, consists of multiple alternating layers of multilayer perceptron encoding modules and pixel reordering modules. Figure 1 illustrates the operational structure of the discriminator based on WGAN-GP. The generator configuration is detailed below.

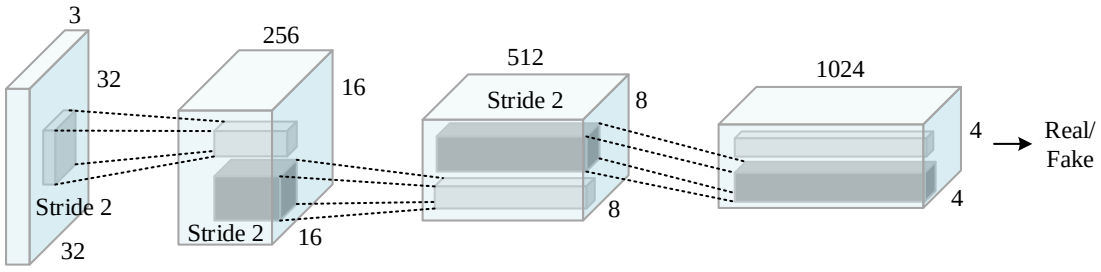

Figure 1: Model structure of the WGAN-GP discriminator

2.1.2 Generator Based on Multi-Layer Perceptrons

Compared to currently popular multilayer perceptron models such as MLP-Mixer, ResMLP, gMLP, RaftMLP, ViP, S2MLPv2, Sparse-MLP, and HireMLP, it is evident that they were initially applied to image classification tasks. Inspired by the contributions of Transformers and convolutional neural networks to image generation, this section adapts these multilayer perceptron architectures for image generation. Convolutional neural networks capture local spatial information, requiring complex designs and deep architectures to obtain broader neighborhood information. In contrast, multilayer perceptron structures can naturally acquire global information. Compared to Transformer-based generators, the model architecture designed in this chapter is more concise, leading to more stable and easier training.

Calculating the cost according to the multi-layer perceptron model is infeasible even for individually labeling pixels in a low-resolution 32×32 image. Therefore, this work initiates the generation process with low-resolution data featuring longer channels. After passing through several multi-layer perceptron encoders, a pixel reconstruction operation is employed to increase resolution while reducing the number of channels. This approach enhances image resolution while simultaneously reducing computational memory requirements. The specific architecture of this generator model is as follows: First, a 128-dimensional vector is randomly sampled from a normal distribution as the latent variable. This vector is then projected linearly into a high-dimensional space. Subsequently, the data undergoes three encoding stages, each comprising multiple MLP encoders. Additionally, an upsampling operation based on pixel recomposition is inserted between every two stages to alleviate computational pressure on the generator. Finally, the generated image is output through a linear de-flattening process.

Although the aforementioned popular multi-layer perceptron models possess different architectures, their underlying principles are broadly similar (except for gMLP). They all feature signaling mixing modules with GeLU nonlinear activation functions and channel mixing modules, stacked through residual connections and layer normalization. In gMLP, the architecture begins with channel mixing modules, replacing the signaling mixing module structure with spatial gating units.

2.2 LG-MLFormer Network Architecture Design

Following the introduction of image generation methods based on hybrid multilayer perceptrons and their generator structures, we further propose a more complex network architecture—LG-MLFormer. This model comprises an LG-MLP local-global multilayer perceptron encoder and a cross-domain memory-enhanced decoder to address challenges posed by semantic noise and computational efficiency.

2.2.1 LG-MLP Encoder

Although the self-attention mechanism mitigates semantic noise between grid features and regional features through cross-attention operations, it suffers from the following shortcomings: (1) It disregards potential correlations between different regions within an image. (2) It exhibits quadratic computational complexity. (3) To address the loss of spatial geometric information when grid-like features are input into the self-attention mechanism due to the need for dimension flattening, the self-attention mechanism requires the introduction of relative position encoding. This leads to additional parameters and computational overhead for the model. To resolve these three major issues, this chapter proposes an LG-MLP module for constructing the visual encoder. The LG-MLP in this chapter consists of two independent local MLP modules and a CDGM module.

(1) Local MLP Module

To achieve self-compensation of visual feature space information without introducing additional parameters, this chapter's Local MLP (LM) module specifically designs the mapping dimension between linear layers and employs two linear layers to replace the self-attention layer. Its operation is as follows.

First, a set of regional features or grid features X extracted from the input image is provided. The LM module captures latent correlations between different images to obtain a visual feature representation rich in semantic information. This chapter designs the linear mapping dimension between linear layers based on the number of visual regions to achieve self-compensation of visual feature space information without introducing additional parameters. Finally, this chapter implements a gating mechanism to enhance the effectiveness of the LM module. The above operations can be defined as:

$$FC(X) = WX + b$$
$$mlp(X) = FC(Norm(FC(X))) \tag{1}$$

$$LM(X) = \sigma(FC([W, mlp(X), W_2 mlp(X)])) \otimes W_1 mlp(X) \tag{2}$$

Among these, [,] denotes a connection, σ represents the sigmoid activation function, and $\otimes$ denotes matrix multiplication.

(2) CDGM Module

Recently, attention-based models have effectively addressed semantic noise between regional and grid features through cross-attention mechanisms. While these approaches have achieved success in image captioning tasks, they also introduce unnecessary parameters and computational overhead. Therefore, this chapter proposes the CDGM module, which eliminates semantic noise by exploring cross-domain latent correlations between grid-based and regional features, achieving superior performance without introducing external parameters. Its working principle is as follows.

First, this chapter integrates the outputs of the LM module through concatenation. Then, both raw features and integrated features are mapped into the same abstract space via linear activation operations. Finally, a gating mechanism is employed to weight the importance of different features. The correlation between local and global semantics can be adjusted through weight values to eliminate semantic noise. The aforementioned operations can be defined as follows.

$$\tilde{E}_r = Norm(FC(E_r))$$
$$\tilde{E}_g = Norm(FC(E_g)) \tag{3}$$
$$\tilde{E} = Norm(FC([E_r, E_g]))$$

$$CDGM(\tilde{E}, \tilde{E}_r, \tilde{E}_g) = \sigma([FC([\tilde{E}_r, \tilde{E}_g]), FC(\tilde{E})]) \otimes FC([\tilde{E}_r, \tilde{E}_g]) \tag{4}$$

E_r and E_g represent dual-source features with latent semantic associations extracted by the LM module, [,] denotes concatenation, σ is the sigmoid activation function, and $\otimes$ is matrix multiplication.

(3) Encoder Layer

This chapter embeds the local MLP and CDGM modules into the LG-MLP encoding layer. The output of the LG-MLP encoding layer is applied to the feedforward neural network layer, which performs a nonlinear affine transformation on each input element. The formula for the feedforward neural network layer is defined as follows.

$$F(X)_i = V_2 relu(V_1 X_i + b_1) + b_2 \tag{5}$$

Here, X_i represents the i-th vector in the input set, and $F(X)_i$ represents the output of the i-th vector in the output-input set. Additionally, relu denotes the ReLU activation function, V_1 and V_2 denote the linear layer parameter weights, and b_1 and b_2 denote the bias term coefficients.

Subsequently, the LG-MLP encoding layer and feedforward neural network layer are encapsulated through residual connections and layer-wise normalization operations, forming

the overall architecture of the encoding layer. The operational formula for the encoding layer is defined as follows.

$$LGMLP(X_r, X_g) = CDGM(LM(X_r), LM(X_g)) \tag{6}$$

$$\tilde{E} = AddNorm(LGMLP(X_r, X_g)) \tag{7}$$

AddNorm denotes the combination of residual connections and layer normalization operations.

Finally, multiple encoder layers are linearly stacked, with the input to the i-th encoder layer being the output of the (i-1)-th encoder layer. This allows higher layers to learn and refine the feature representations from preceding layers. Consequently, stacking N encoder layers integrates the outputs of each encoder layer to produce multi-level outputs.

2.2.2 Cross-Domain Memory-Augmented Decoder

In this chapter, the decoder utilizes the multi-level visual features extracted by the encoder to generate a sentence word-by-word. The decoder adopts a multi-layer structure to better leverage both high-level and low-level visual features. In deep neural network models, high-level submodules exert a greater influence on model performance than low-level submodules. Through multi-level feature processing, information from low-level features becomes diluted. Consequently, the grid-based connection pattern generates redundant information. To reduce computational complexity while better utilizing multi-level features, this chapter proposes a novel Cross-Domain Connection (CDC) pattern. CDC employs a depth-increasing connection pattern. Compared to the quadratic computational complexity of grid-based connections, CDC achieves outstanding performance while significantly reducing computational complexity. A comparison between the CDC pattern and the mesh connection model is shown in Figure 2.

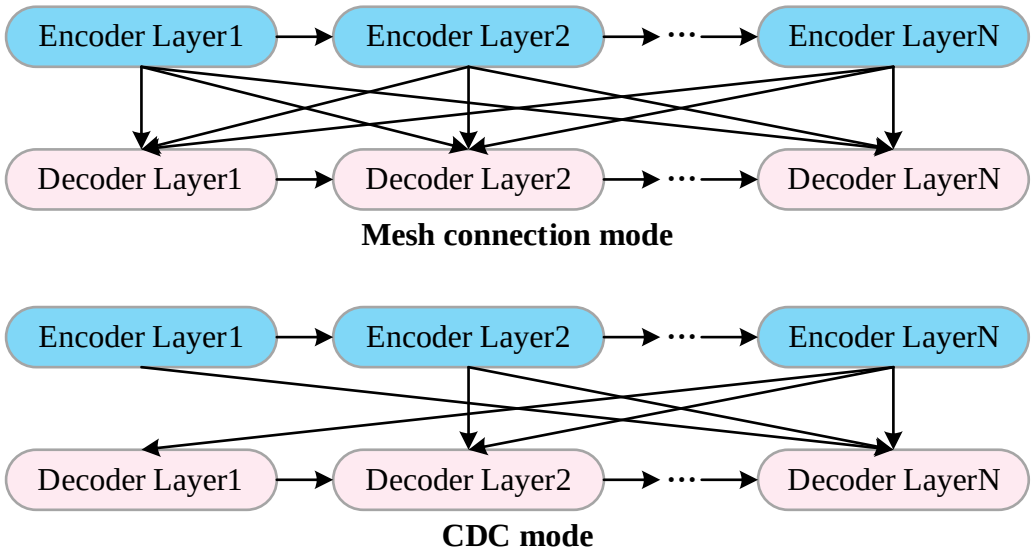

Figure 2: Comparison Chart of CDC Mode and mesh connection Model

(1) Visual-Linguistic Memory-Enhanced Attention Mechanism

To better adapt to inputs from two distinct modalities—text and images—this chapter designs two learnable parameter matrices of different dimensions. The smaller-dimensional matrix serves the text memory mechanism, while the larger-dimensional matrix supports the

visual memory mechanism. During decoding, the text memory mechanism provides prior knowledge for self-attention computations using corpus samples, while the visual memory mechanism leverages visual prior knowledge. The operational principle is as follows.

First, the text memory-enhanced attention mechanism captures long-range relationships between words in the input text sequence, generating a sequence of word vector features Y and integrating prior knowledge from corpus samples. Subsequently, the visual memory-enhanced attention mechanism connects the encoder outputs E from all layers with the word vector feature sequence Y via CDC. Unlike Transformers that utilize only the final encoder layer's output, the decoder layer in this approach performs cross-attention operations on features extracted from different encoder layers in a deep incremental mode to leverage multi-level features. The CDC-based visual-language memory-enhanced attention mechanism can be defined as:

$$CDC(\tilde{E},Y)=\sum_{i=1}^{N}\sum_{j=N+1-i}^{N}\alpha_j \otimes VLMA(\tilde{E}_j,Y) \tag{8}$$

Among these, VLMA stands for Visual-Language Memory Augmented Attention Mechanism. α is a weight matrix.

$$\begin{aligned} VLMA(\tilde{E}_j,Y) &= soft\max(\frac{W_q\cdot\tilde{K}_j}{\sqrt{d}})\cdot\tilde{V}_j \\ \tilde{K}_j &= [W_k\tilde{E}_j, M_k^j] \\ \tilde{V}_j &= [W_v\tilde{E}_j, M_v^j] \end{aligned} \tag{9}$$

Among these, M_k and M_v are learnable weight matrices with n tokens, where [,] denotes concatenation. Keys and values can be learned by adding two learnable memory units, thereby providing prior knowledge for the attention operation. α is a weight matrix. This weight implements contributions to each encoding layer and modulates priorities. These functions are computed by evaluating the correlation between the cross-attention results and the input queries. The formula is defined as follows.

$$\alpha_j = gelu(W_j[Y,VLMA(\tilde{E}_j,Y)]+b_j) \tag{10}$$

Among these, gelu denotes the Gaussian error linear unit activation function, [,] represents the concatenation operation, W_j is a weight matrix of dimension $2d\times d$, and b_j is a learnable bias coefficient.

(2) Decoder Layer

Regarding the decoder, during image description generation, since the prediction of the t-th word is entirely dependent on the t-1-th predicted word, the self-attention process in the decoder layer requires a masking operation where Y=t. Additionally, the decoder layer encapsulates a memory-augmented attention mechanism and a feedforward neural network layer through an AddNorm operation. The operational formula for the decoder layer is defined as follows.

$$\tilde{Y}=AddNorm(CDC(\tilde{E},AddNorm(VLMA_m(\tilde{E},Y)))) \tag{11}$$

where Y is a sequence of word vectors, and $VLMA_m$ denotes a masked visual-language memory-enhanced attention operation.

2.3 Attributes of Image-Generating Art

Through technical analysis of generative models, we naturally transition to discussing their artistic attributes, thereby fully presenting the dual dimensions of AI-generated imagery in creative design—technical implementation and visual expression.

Within the digital and postmodern visual culture landscape, we inevitably find ourselves enveloped by technology. Against this backdrop, the field of art and design must focus on striking a balance between the digital, technology, machines, and art. As a novel form of digital art, image-generating art combines encoding and decoding technologies with artistic inspiration, integrates public sentiment, and engages in dialogue with computers. Through its visual characteristics, it continues to explore the relationship between humans and machines. This section will focus on analyzing the characteristics of image-generating art, as illustrated in Figure 3: First, encoding and decoding; Second, dynamic and static; Third, interaction and dialogue; Fourth, data and image; Fifth, randomness and imagination. This chapter will emphasize the first two aspects.

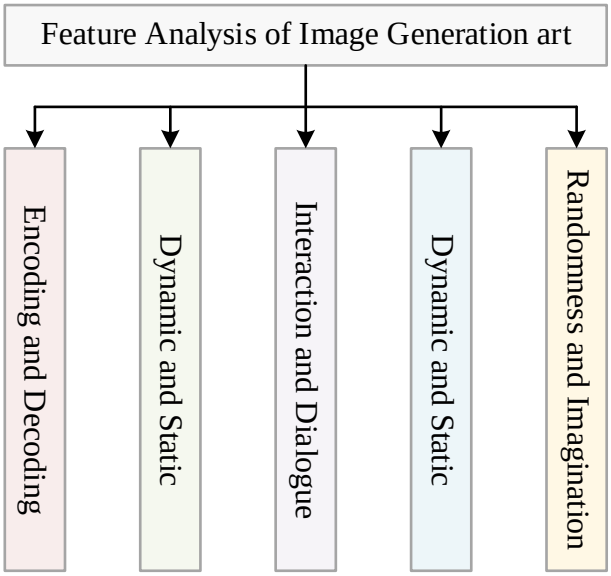

Figure 3: Feature Analysis of Image Generation Art

2.3.1 Encoding and Decoding

The encoding and decoding of image-generated art are crucial factors determining the success of an artwork's information transmission. Encoding characteristics refer to the generation methods and techniques employed by the artist during image creation, as well as the artistic principles and creative approaches integrated into the process. These elements determine the information conveyed and the artistic nature of the image, thereby influencing the audience's understanding and perception. Decoding characteristics, conversely, pertain to the “image-reading” skills utilized by viewers when appreciating the image—that is, how they interpret the information and artistic qualities conveyed.

Interaction between artists and audiences is crucial throughout the encoding and decoding process. Artists must consider how to accurately express artistic information while accounting for the audience's background and cultural factors to ensure the message is conveyed precisely. Audiences, in turn, require a foundation of art knowledge and aesthetic discernment to correctly interpret the artist's intended message. Through this encoding and decoding process, artists and audiences jointly construct the overall framework of image-generation art, realizing the

complete cycle of artistic creation and appreciation while forging a shared visual language between them.

2.3.2 Dynamic and Static

Image generation art is an artistic form based on computer technology and algorithms, producing images that can exhibit dynamic or static characteristics. In the early era of image reading, images were predominantly presented statically—conveying information and artistic expression through fixed frames. However, with the advent of the digital age, image generation art gradually shifted from static to dynamic forms. Dynamic image generation art more intuitively expresses information and artistic effects through movement, change, and interaction, better capturing people's attention. Nevertheless, as an image form present since the dawn of visual culture, static image-based art remains one of the most widely employed expressive techniques today. It conveys emotion, thought, and aesthetic appeal through unchanging frames, and for certain artists and audiences, its charm remains irreplaceable by dynamic imagery.

In summary, dynamic and static characteristics represent the two fundamental forms of image-generating art. As computer technology advances and artists continue their explorations, the expressive forms and characteristics of image-generating art will continually evolve and expand.

3 Experimental Validation and Analysis of Image Generation Models Based on LG-MLP

To validate the effectiveness and superiority of this model in image generation tasks, we will conduct systematic experimental analysis and discussion centered on mask hyperparameter experiments and image realism comparison experiments.

3.1 Hyperparameter Experiments with Masking

To evaluate the image generation performance of the proposed LG-MLP hybrid multi-layer perceptron with local-global integration, and to ensure consistent user experience under both consistent and inconsistent conditions, different processing approaches must be applied to the input visual conditions during testing. Consistent visual conditions can be achieved by processing the original image using different operators or feature extraction modules. Additionally, this paper proposes a method to simulate inconsistent condition inputs by utilizing segmentation maps. This approach is grounded in the assumption that users focus more on objects within the scene—i.e., each segmented category—when providing input. Consequently, masks can be constructed for specific images based on segmentation maps.

The specific steps for generating inconsistent visual condition masks for each image are as follows:

(1) Obtain the segmentation map using the segmentation map operator. (2) Iterate through the segmentation map and randomly set the mask value to 1 or 0 for each pixel belonging to a category, based on its classification. (3) Apply median filtering to smooth irregular scattered points in the mask map (e.g., fuzzy boundaries in the segmentation map) to ensure semantic consistency.

3.1.1 Evaluation Metrics and Experimental Setup

This paper employs the Fréchet Inception Distance (FID) as an evaluation metric for image generation models. FID serves as a perceptual distance metric based on high-level features,

functioning as a standard for measuring realism. By comparing high-level feature information, the FID metric can partially reflect the impact of additional control conditions on generated images. Furthermore, FID is relatively simple to implement and more versatile, relying solely on the Inception feature extractor and exhibiting low sensitivity to differences between datasets.

In this experiment, the validation set of the MS COCO dataset was selected as the source for conditional images, from which 2000 images were randomly sampled. The FID score of generated images relative to the test set served as the evaluation criterion for generation quality. We investigated various hyperparameter combinations, including mask coverage (ranging from 0.1 to 0.9 in 0.1 increments) and two mask shapes—Scatter and Blocks—combined with nearest-neighbor upscaling and bilinear upscaling methods, resulting in 36 distinct model parameter configurations. Through these experiments, we aimed to identify the hyperparameter combination yielding the highest generation quality. A decrease in FID scores indicates that models trained under corresponding masking conditions exhibit superior performance.

3.1.2 Masked Hyperparameter Experiments Under Consistent Conditions

Table 1 presents the FID scores for image generation quality across different mask hyperparameters under consistent conditions. Figure 4 illustrates the trend in image generation quality across models trained with varying hyperparameters.

Table 1: FID of different mask hyperparameters under consistent condition

Mask coverage rate	Nearest Neighbor		Bilinear Interpolation	
	Scatter plot	Block	Scatter plot	Block
0.1	4.85	4.89	4.88	4.93
0.2	4.84	4.86	4.86	4.81
0.3	4.88	4.94	4.92	4.99
0.4	4.84	4.79	4.83	4.79
0.5	4.92	4.93	4.90	4.86
0.6	4.82	4.85	4.80	4.84
0.7	4.80	4.92	4.85	4.96
0.8	4.77	4.96	4.74	4.94
0.9	4.79	5.11	4.82	5.05

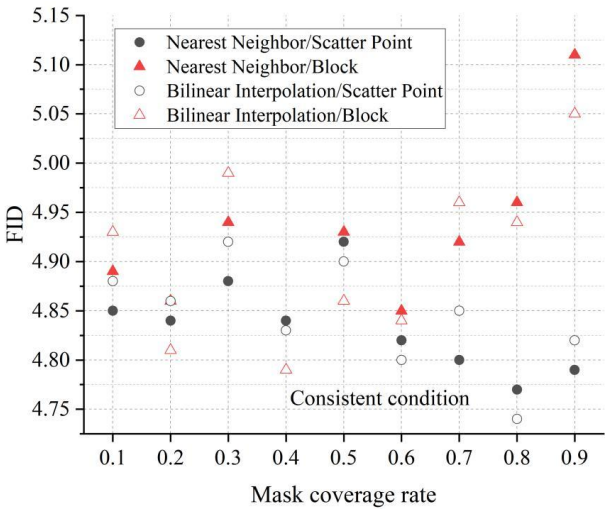

Figure 4: Model performance under consistent conditions

Observation of the data reveals that under consistent conditions, the impact of mask hyperparameters on image generation quality is relatively minor. Fluctuations in image realism across different model parameters typically fall within a range of ±0.1. This phenomenon indicates that for condition combinations already exhibiting high consistency, their dependency on feature extractors decreases, thereby narrowing the range of variation in generation quality. Analyzing from the perspective of mask coverage, when training coverage ranges between 10% and 40%, generated image quality remains generally comparable despite some fluctuations. However, as mask coverage continues to increase, the FID scores of models trained with Block-type masks exhibit a steep upward trend, indicating a decline in model generation quality at this stage. In 70% of Block-type masking conditions, the mask occupies the central position of the conditional image while covering the most critical information within the condition. During training, the feature extractor struggles to capture effective conditional information, weakening the model's overall ability to process conditions. In contrast, consistent condition combinations exhibit high information density. This disparity leads to a noticeable decline in generated image quality. This trend is also evident in scatter-type masks: at 90% coverage, models trained with scatter-type masks show an upward trajectory in FID scores, indicating performance degradation.

When analyzing the impact of mask types on image generation quality, this study found that scatter-type masks generally exhibit superior performance. At a 40% masking rate, block-type masks achieved the optimal generation results with an FID score of 4.79. Conversely, at an 80% masking rate, scatter-type masks delivered the best generation outcomes, achieving an optimal FID score of 4.74.

3.1.3 Inconsistency Condition Mask Hyperparameter Experiment

The variation in image generation quality among models trained with different hyperparameters under inconsistent conditions is shown in Table 2 and Figure 5. Compared to consistent conditions, image quality under inconsistent conditions is significantly influenced by the mask hyperparameters during training. The realism of images generated by different model parameters ranges from 4.61 to 5.41.

Table 2: FID of different mask hyperparameters under inconsistent condition

Mask coverage rate	Nearest Neighbor		Bilinear Interpolation	
	Scatter plot	Block	Scatter plot	Block
0.1	5.41	5.18	5.35	5.23
0.2	5.13	5.09	5.17	5.11
0.3	4.98	4.92	5.02	4.99
0.4	4.88	4.82	4.93	4.86
0.5	4.81	4.84	4.86	4.83
0.6	4.63	4.86	4.71	4.96
0.7	4.61	4.96	4.62	5.02
0.8	4.74	5.17	4.78	5.11
0.9	4.85	5.26	4.90	5.22

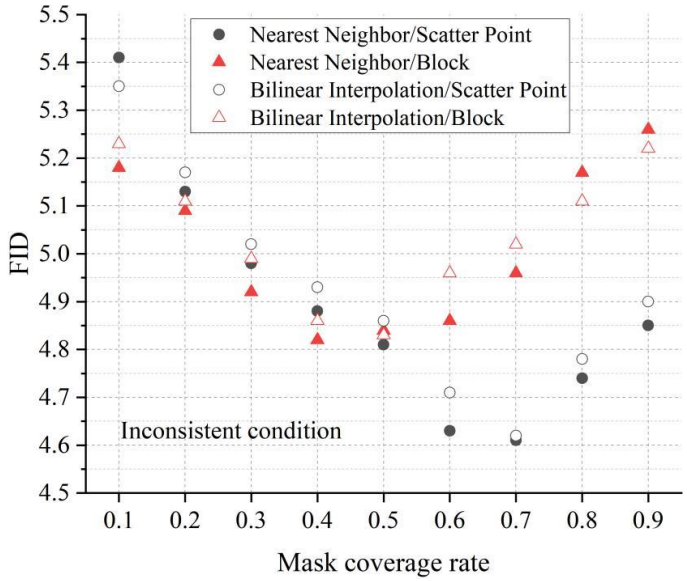

Figure 5: Model performance under inconsistent conditions

Under inconsistent conditions, the impact of coverage rate on model generation quality exhibits a non-linear pattern. Instead of a single curve, it follows a U-shaped curve influenced by mask type, featuring distinct inflection points. For scatter-type masks, generation quality deteriorates when coverage rate falls below 40%. Optimal image quality is achieved within the 40%–70% range. However, as coverage continues to increase, the FID score shows an upward trend, indicating a decline in image quality. For block-type masks, optimal image quality is achieved at low coverage levels between 10% and 50%. Yet, as coverage increases, the FID score rises, indicating a deterioration in image quality.

When examining the impact of downsampling methods on image generation quality under inconsistent conditions, this study found negligible performance differences between bilinear interpolation downsampling and nearest-neighbor downsampling. At a 70% occlusion rate, with scatter points forming the mask shape, the nearest-neighbor method achieved the lowest FID value of 4.61, while the bilinear interpolation method under the mask combination yielded a best performance of 4.62—a difference of merely 0.01.

Based on the aforementioned experimental results, this paper employs a scatter-point-type mask design for conditional fusion tasks, maintaining a coverage rate of 70%. Simultaneously, to prevent degradation of shared feature components caused by high coverage rates, the designed mask generation algorithm allocates empty regions to random conditions. This forces full mask coverage, thereby increasing the proportion of shared visual condition regions within the training data. Due to the requirement for secondary allocation of empty regions, the bilinear interpolation method tends to form block-like rectangular structures after secondary allocation of the upsampled mask. In contrast, the nearest-neighbor upsampling method yields more random results and is less susceptible to this issue. Therefore, the nearest-neighbor upsampling method is ultimately selected for generating the masks required during training.

3.2 Image Realism Comparison Experiment

To further validate the superiority of the LG-MLFormer multi-layer perceptron model designed in this paper for image generation, comparative experiments were conducted against four popular baseline models: ControlNET, DeC-ControlNET, GLIGEN, and T2IAdapter.

The research continued to focus on two distinct experimental scenarios: consistent and inconsistent conditions. For the comparative experiments, the MS COCO dataset was selected

as the training and testing platform. Two representative features—HED edge maps and Midas depth maps—were employed as prompting conditions. The key metric for evaluating model performance was image fidelity, assessed using the Fréchet Perceptual Distance (FID) score.

The image fidelity of different models under consistent and inconsistent conditions is shown in Table 3.

Table 3: The FID of generated images under consistent and inconsistent conditions

		Consistent	Inconsistent
HED	ControlNET	7.03	8.41
	DeC-ControlNET	6.92	7.51
	GLIGEN	7.63	8.08
	T2IAdapter	7.34	7.88
	LG-MLP	6.12	6.32
Midas	ControlNET	4.83	7.82
	DeC-ControlNET	4.77	6.83
	GLIGEN	4.80	7.53
	T2IAdapter	4.81	7.10
	LG-MLP	4.46	4.54

It can be observed that the LG-MLP model proposed in this paper demonstrates outstanding image generation performance under both consistent and inconsistent conditions. When using HED edge feature maps as guidance, LG-MLP achieves an FID of 6.12 under consistent conditions, significantly lower than ControlNET's 7.03. Under inconsistent conditions, LG-MLP's FID of 6.32 also significantly outperforms other baseline models (ControlNET: 8.41, DeC-ControlNET: 7.51, GLIGEN: 8.08, T2IAdapter: 7.88). Under Midas depth map conditions, LG-MLP achieved FID values of 4.46 and 4.54 for consistent and inconsistent conditions respectively, both lower than all comparison models. These results demonstrate that the LG-MLP model generates more photorealistic images across diverse visual conditions, exhibiting particularly strong robustness when handling inconsistent conditions.

4 Practical Application of LG-MLP-Based Generative Models in Automotive Styling Design

The preceding sections systematically validated the LG-MLFormer model's outstanding performance and robust generalization capabilities in image generation tasks through masking experiments and realism comparisons. To further explore the model's practical value and impact within real-world creative design workflows, this chapter shifts the research focus from algorithmic aspects to application scenarios. It conducts empirical research centered on automotive styling design to evaluate the model's contributions toward enhancing design efficiency, creative output, and final deliverable quality.

4.1 Comparative Experimental Design and Evaluation System Construction for Automotive Styling Design

4.1.1 Experimental Setup

To ensure the objectivity of the experiment and avoid the influence of other factors on the results, it is essential to ensure that each experimental group is closely matched in terms of team size,

design experience level, and aesthetic judgment. The nine designers participating in the project were randomly divided into three groups, with three designers assigned to each team. Experiment Group A applied an AI image generation method using LG-MLP hybrid multilayer perceptrons. Experiment Group B utilized conventional AIGC AI-assisted tools. The control group employed no auxiliary tools, advancing the project through traditional automotive styling design methods.

The experiment simulated the automotive styling design process, encompassing concept generation, proposal development, proposal finalization, and design reviews.

Experimental groups integrated generative AI tools into the design process, producing AIGC-generated design renderings with manual post-processing. Based on review feedback, AIGC assisted in refining designs to final outcomes. The control group employed traditional methods—manual sketching and brainstorming—to produce 2D/3D rendered design proposals, manually adjusting them to final forms based on review feedback.

4.1.2 Evaluation Criteria

The experiment formed an expert project team comprising 20 professionals with relevant experience and backgrounds, including seasoned in-service interior and exterior designers, university lecturers in transportation design, and graduate students. The expert panel's evaluation will strictly assess the quality of the proposed automotive designs based solely on professional expertise, design experience, and industry standards. Questionnaire reviews will employ a 5-point scale to measure respondents' personal preferences toward each design proposal, quantifying their level of preference for the design concepts. The survey targeted automotive users aged 18-60, with 200 questionnaires distributed and 172 returned. This comparative experiment comprehensively evaluated the three designer groups across the following dimensions:

(1) Efficiency: Measured by the total time required from initial concept to final design proposal.

(2) Creativity: Assessed by the expert panel based on industry standards for novelty and uniqueness of the design solution.

(3) Quality: Assessing the design's fidelity to initial project inputs, including ergonomic, aesthetic, and functional requirements.

(4) Review Satisfaction: Gathering preferences and feedback on the produced designs from both a broad potential user base and the expert project panel.

4.2 Comparative Analysis of Comprehensive Performance and Results Under Different Design Approaches

Following three rounds of automotive styling design work and reviews—concept generation, proposal development, and proposal finalization—and integrating experimental results with established evaluation criteria, the comprehensive assessments for the three teams are as follows.

4.2.1 Efficiency Evaluation

The completion times for design proposals across each group and stage are shown in Figure 6. Comparisons reveal that generative AI significantly outperforms traditional design workflows in efficiency at every stage, demonstrating a pronounced advantage in productivity. Experiment Group A, employing the LG-MLP hybrid multilayer perceptron method, completed the entire process in just 194 minutes—substantially less than the 245 minutes recorded by Experiment Group B using conventional AIGC tools and the 362 minutes achieved by the traditional method

control group. Breaking down the design phases, Experiment A Group spent 27 minutes on concept generation, 76 minutes on scheme generation, and 91 minutes on scheme finalization—all notably shorter than the other two groups. These results demonstrate that the LG-MLP-based generative AI method substantially shortens the design cycle and enhances overall efficiency.

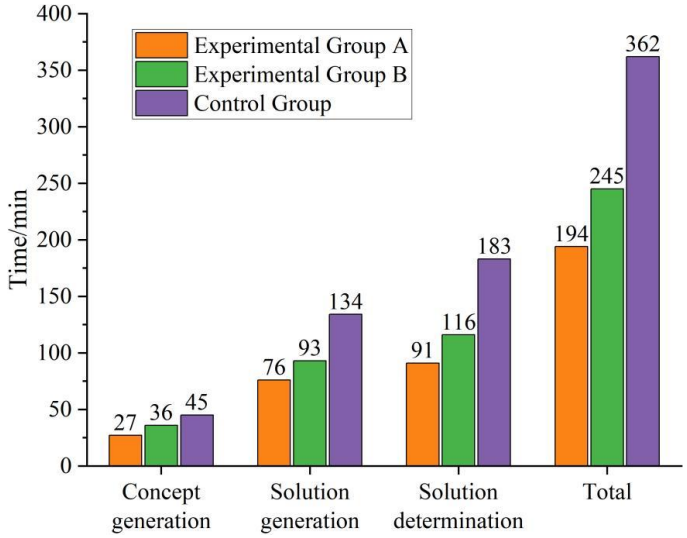

Figure 6: The completion time statistics of the design plans for each stage

4.2.2 Creative Evaluation

A 5-point scale was used for scoring. The statistical results of the creativity assessment for each group's design proposals at each stage are shown in Figure 7. The score statistics indicate that at every stage, the AI-assisted method achieved higher scores in creativity assessment compared to the traditional method. Experimental Group A scored 4.85, 4.60, and 4.75 in the concept generation, proposal generation, and proposal confirmation stages respectively, with an average of 4.73—significantly higher than Experimental Group B (average 4.27) and the control group (average 3.25). This demonstrates that the LG-MLP method possesses distinct advantages in stimulating design creativity and novelty, enabling designers to generate more unique and innovative proposals.

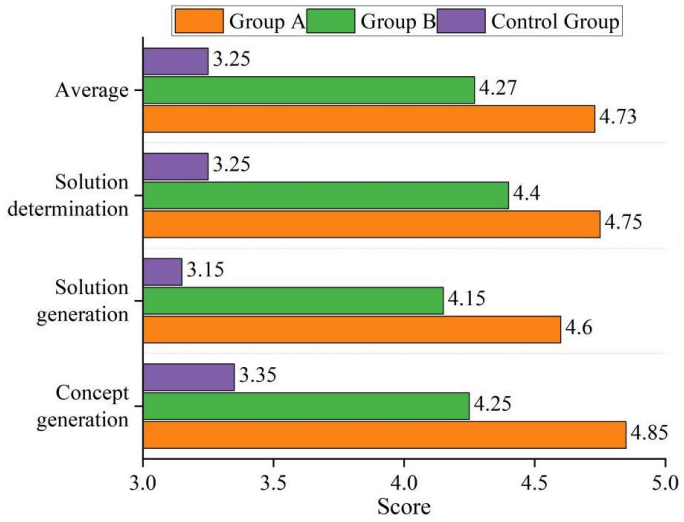

Figure 7: The evaluation results of the design concept for each stage and each group

4.2.3 Quality Assessment

A 5-point scale was used for scoring. The statistical results of design quality assessments for each group across all stages are shown in Figure 8. Experimental Group A demonstrated outstanding performance at every stage, achieving scores of 4.95, 4.80, and 4.85 for concept generation, proposal generation, and proposal confirmation respectively, with an average score of 4.87. This significantly exceeded Experimental Group B (average 4.42) and the control group (average 3.33). The results indicate that design solutions generated by the LG-MLP method demonstrate superior alignment with ergonomic, aesthetic, and functional requirements, resulting in higher design quality.

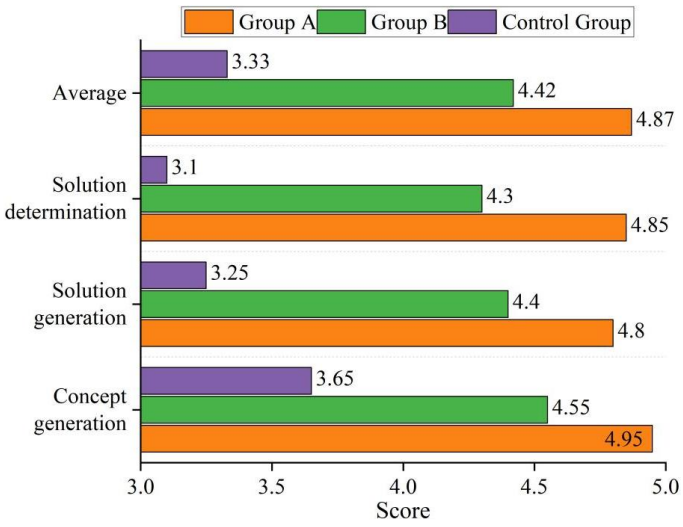

Figure 8: The quality assessment results of the design plans for each stage

4.2.4 Review Satisfaction Assessment

The satisfaction ratings for expert reviews and public participation at each stage are shown in Table 4.

Table 4: The satisfaction levels of expert reviews and public participation in each stage

		Group A	Group B	Control Group
Expert group satisfaction	Concept generation	4.95	4.50	3.70
	Solution generation	4.80	4.35	3.45
	Solution determination	4.90	4.25	3.20
	Average	4.88	4.37	3.45
Questionnaire survey	Concept generation	4.82	4.34	3.52
	Solution generation	4.71	4.27	3.01
	Solution determination	4.72	4.18	2.89
	Average	4.75	4.26	3.14

Review satisfaction comprised two components: expert panel evaluations and public questionnaires. Experimental Group A achieved an average expert review score of 4.88 and an average public questionnaire score of 4.75, both significantly higher than Experimental Group B (expert: 4.37, public: 4.26) and the control group (expert: 3.45, public: 3.14). The LG-MLP method garnered higher recognition in both professional evaluations and public preference, indicating that its generated design solutions better align with professional standards and user aesthetic demands.

5 Conclusion

This paper systematically investigates an artificial intelligence image generation model based on the LG-MLP hybrid multilayer perceptron, along with its visual expression capabilities and practical efficacy in creative design. In terms of technical performance, the model demonstrates outstanding results on the MS COCO dataset. Under consistent and inconsistent conditions, the FID values reached 6.12 and 6.32 respectively when utilizing HED edge features, and further decreased to 4.46 and 4.54 when employing Midas depth maps. These results significantly outperform mainstream baseline models such as ControlNet and DeC-ControlNet, validating the superiority of this model in generating photorealistic images and achieving cross-modal generalization capabilities.

At the application level, when applied to automotive styling design practice, the design group utilizing the LG-MLP method completed the task in just 194 minutes—an approximately 85.5% efficiency improvement compared to the 362 minutes required by traditional design workflows. In terms of creativity, final product quality, and expert/user satisfaction, the average scores reached 4.73, 4.87, and 4.75 respectively, comprehensively surpassing both the standard AIGC-assisted group and the traditional design group. This demonstrates that the model not only possesses robust engineering capabilities but also significantly enhances efficiency, stimulates innovation, and improves output quality in practical creative design scenarios.

References

[1] Liu, G., Du, H., Niyato, D., Kang, J., Xiong, Z., Kim, D. I., & Shen, X. (2024). 0effective content creation. IEEE Network, 38(5), 295-303.

[2] Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P., & Sun, L. (2025). A survey of ai-generated content (aigc). ACM Computing Surveys, 57(5), 1-38.

[3] Wang, Z., Shen, L., Kuang, E., Zhang, S., & Fan, M. (2024, July). Exploring the impact of artificial intelligence-generated content (AIGC) tools on social dynamics in UX collaboration. In Proceedings of the 2024 ACM designing interactive systems conference (pp. 1594-1606).

[4] Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. arXiv preprint arXiv:2303.04226.

[5] El Ardeliya, V., Taylor, J., & Wolfson, J. (2024). Exploration of artificial intelligence in creative fields: Generative art, music, and design. International Journal of Cyber and IT Service Management, 4(1), 40-46.

[6] Mohamed, E. L., Fouad, A. M., Shamiea, M. M., Elkhayat, A. S., El-Shafey, F. K., & Hassabo, A. G. (2025). A Fusion of Creativity and Technology: The Art of Design and the Role of Printing Materials in the Digital Transformation. Journal of Textiles, Coloration and Polymer Science, 22(1), 341-347.

[7] Danacılar, İ. A. (2025). Digital Transformation of Art and Design: Innovative Forms of Expression in Advertising. In Impact of Contemporary Technology on Art and Design (pp. 105-134). IGI Global.

[8] Irwin, T. (2015). Transition design: A proposal for a new area of design practice, study, and research. Design and culture, 7(2), 229-246.

[9] Yang, W. (2025). Analysis of AIGC and Visual Communication Design Fusion Strategies. Journal of Humanities, Arts and Social Science, 9(2).

[10] Li, Y. (2024). Visual Communication Design Education Reform for the Development of Digital Creative Industries. International Journal of New Developments in Education,, 6(1).

[11] Lin, Y., & Liu, H. (2024, June). The impact of artificial intelligence generated content driven graphic design tools on creative thinking of designers. In International Conference on Human-Computer Interaction (pp. 258-272). Cham: Springer Nature Switzerland.

[12] Chen, G., Lan, X., Liu, K., & Cheng, C. (2025). Research on fusion generation algorithm of visual communication and product design based on AIGC technology. Systems and Soft Computing, 7, 200237.

[13] Chen, T., Pang, B., Ma, C., & Shao, W. (2024). Exploration of Brand Visual Communication Innovation Design Method Based on AIGC Technology. Procedia Computer Science, 247, 519-528.

[14] Wang, Y. (2025). Application and Value Evaluation of AIGC Based on Sustainable Development in Visual Communication Design. In SHS Web of Conferences (Vol. 213, p. 02038). EDP Sciences.

[15] Liu, V., Qiao, H., & Chilton, L. (2022, October). Opal: Multimodal image generation for news illustration. In Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (pp. 1-17).

[16] Zhao, T., Yang, J., Zhang, H., & Siu, K. W. M. (2021). Creative idea generation method based on deep learning technology. International Journal of Technology and Design Education, 31(2), 421-440.

[17] Bhattacharjee, G. (2023). Art and photography in the age of artificial intelligence. In 12th International Photographic Conference of PAD, Kolkata.

[18] Xu, L., Ye, C., & Tang, J. (2024, November). Intelligent Generation and Optimization of New Media Art Driven by AIGC. In Proceedings of the 2024 4th International Conference on Big Data, Artificial Intelligence and Risk Management (pp. 521-527).

[19] Ye, C., Ganbat, T., & Xu, L. (2023). Research on the application of artificial intelligence generated AI technology in new media art. Highlights in science, engineering and technology, 68, 313-319.

[20] Zhang, W., He, N., Deng, Z., Huang, C., & Cai, J. (2024, November). AIGC-enabled Cultural and Creative Product Design Exploration: Macao Intangible Cultural Heritage Dragon Dance Element as an Example. In Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Education (pp. 380-385).

[21] Liu, Q., Wang, X., Xie, X., Wang, W., & Lu, X. (2024). Innovative Design Research on

Jiaodong Peninsula's Marine Folk Culture Based on AIGC. International Journal of Contemporary Humanities, 8(1), 17-27.

[22] Pan, S., Anwar, R. B., Awang, N. N. B., & He, Y. (2025). Constructing a sustainable evaluation framework for AIGC technology in Yixing Zisha pottery: balancing heritage preservation and innovation. Sustainability, 17(3), 910.

[23] Ploennigs, J., & Berger, M. (2023). AI art in architecture. AI in Civil Engineering, 2(1), 8.

[24] Zhou, X. (2025, January). Research on interactive design of cultural and creative products based on genetic algorithm with long short-term memory. In 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN) (pp. 1-8). IEEE.

[25] Lu, W., Wu, J., Li, H., Lu, J., & Wang, X. (2024, August). Enhancing Personalized Cultural Product Design Through An AI-Driven Intelligent Design System. In 2024 2nd International Conference on Design Science (ICDS) (pp. 1-6). IEEE.

[26] Tao, Y., Fu, X., Wu, J., Bian, Z., Zhu, A., Bao, Q., ... & Zhou, C. (2025, April). AIFiligree: A Generative AI Framework for Designing Exquisite Filigree Artworks. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1-18).

[27] Wu, Q., Zhu, B., Yong, B., Wei, Y., Jiang, X., Zhou, R., & Zhou, Q. (2021). ClothGAN: generation of fashionable Dunhuang clothes using generative adversarial networks. Connection Science, 33(2), 341-358.