Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A study of enhancement strategies for artificial intelligence-based automated design tools in complex design tasks

Yuan Ning[1], Kiesu Kim[1,✉]

[1] *Department of Industrial Design, Silla University, Busan, 46958, South Korea*

ABSTRACT

With the important achievements of AI team in the application of diffusion model, automated generation demonstrates its stability, realism and accuracy in text-guided image generation design. In this paper, the automated design generation based on the diffusion model is divided into two processes: forward diffusion and reverse generation. The stable diffusion model is constructed from three parts: self-encoder, U-Net network structure, and text encoder, and the ControlNet control architecture is used to realize the control of the diffusion model to learn a specific task. The model is pre-trained using a combination of three functions: perceptual loss, adversarial loss, and cyclic consistency loss. The LoRA algorithm is added to the U-Net layer of the Stable diffusion model to realize the design task function enhancement of the stable diffusion model. The application effect of the improved Stable diffusion model is analyzed through comparative experiments. The experimental results show that the CLIP Score interval of this paper's model is between 20 and 32.5, while the LPIPS is between 0.1 and 0.6, and the kernel density centroid is (24.85, 0.4), which means that the proposed method in this paper meets the design requirements with higher fidelity while accomplishing the image design task. In the design images generated by automation, in terms of visual logic rationality, the structural proportions, line contours, light and shadow and picture hierarchical relationships of the generated images perform well, with the evaluation results of 3.258, 3.564, 3.058, 3.615, respectively, which indicates that the design images generated by automation using the model have a strong richness.

Keywords: Stable diffusion model, ControlNet control, LoRA algorithm, CLIP Score, automated design

1. Introduction

Artificial intelligence is a technology and methodology that simulates and imitates human intelligence. It can learn and analyze large amounts of data in order to autonomously perform

✉ Corresponding author.
E-mail address: 17663989830@163.com (K. Kim).
Received 15 January 2024; Revised 10 March 2024; Accepted 30 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-277

specific tasks or solve problems. Artificial intelligence can simulate human's ability of perception, learning, decision-making and language [1-4]. For example, technologies such as natural language processing, image recognition and machine learning are application areas of AI. And automated design tools are usually based on AI algorithms and technologies, providing a series of functions and features [5-7], in order to help designers accomplish various design tasks. Currently the main automated design tools include, simulation software, database and knowledge base, CAD software, etc. These tools can improve the efficiency and accuracy of the work while reducing human input [8-10].

In complex design tasks, automated design tools with artificial intelligence provide designers with new ideas and tools. Through automated design methods, designers can accomplish design tasks more efficiently and obtain better design solutions [11-13]. Automated design tools can assist designers in various design activities and provide the required data and knowledge. However, despite the advantages of AI and automation tools in handling big data and improving efficiency and accuracy, the role of humans is still indispensable [14-16]. Only through human-machine cooperation can the advantages of AI and automation tools be better utilized to maximize work efficiency and accuracy [17-18].

In this paper, with the direction of artificial intelligence technology in automated design tools, the diffusion model is divided into two processes, forward diffusion and reverse generation, using a specific diffusion mechanism, and a stable diffusion model is composed from three parts, namely, variational self-encoder, U-Net network structure, and text encoder. The three loss functions of associative perceptual loss, antagonistic loss, and cyclic consistency loss are used to pre-train the model to improve the quality and diversity of design image generation. On this basis, the fine-tuning network LoRA model is added to ensure the original samples while actively generating new sample types. The two-dimensional convolutional layer is used to realize the feature extraction of the input image, and after the fusion of the results of the convolutional layer, the down-sampling layer and the up-sampling layer to form the final output features, the parameters are saved in the LoRA model to achieve the purpose of image style migration. Compare the model of this paper with advanced image design techniques, analyze the design results and fidelity, and evaluate the application effect of the model.

2. Application of artificial intelligence aids in automated design

2.1. Direction of application:

1) The application of AI in graphic recognition has revolutionized graphic design [19]. Image recognition technology can help designers quickly and accurately find the required materials and resources, greatly improving the efficiency of design. For example, the use of AI technology can quickly identify the theme and color in the picture, helping designers to find inspiration and materials faster. The application of this technology not only improves the efficiency of designers, but also brings them more creative space. Models trained on a large number of images and design data can also be utilized to generate design concept sketches, layout drafts, etc. in a variety of styles and themes, helping designers expand the boundaries of their thinking. Or by analyzing the characteristics of different art styles and design genres, it can try to integrate multiple styles to provide designers with new style-oriented inspiration.

2) The application of AI in typography design is also a highlight in the field of graphic design. Traditional typography requires designers to manually adjust fonts, spacing and layout, which is time-consuming and laborious. With the application of AI technology, designers can quickly realize automated typesetting through intelligent typesetting tools, which greatly saves time and energy.AI can also automatically generate the best layout plan according to the content, so that the design process becomes more intelligent and efficient. Text and graphics layout: According to the input

content and design needs (such as posters, brochures, etc.), automatically generate a reasonable layout of the layout, including text size, spacing, graphic position, etc., multi-page consistency layout: for a series of design (such as multi-page documents, web pages, etc.) to ensure consistency of style and layout between different pages.

3) AI can also play an important role in creative design. For example, using Generative Adversarial Network (GAN) technology, designers can train models to generate images in various artwork styles, providing more inspiration and possibilities for their designs. The application of this technology can not only help designers develop creative ideas, but also provide more options and possibilities in the design process.

4) AI is also widely used in color matching and color prediction. By analyzing a large amount of color data, AI can help designers predict future color trends and guide designers to choose color schemes that are more in line with market and user needs. At the same time, AI can also provide professional color matching suggestions for design based on color psychology and color theory to make the design more in line with aesthetic and emotional needs.

5) Intelligent Image Editing Image Repair and Enhancement: Automatically removes noise, scratches and other defects in the image to improve the clarity and quality of the image, Intelligent Keying: Quickly and accurately identifies the main body of the image and the background and separates them, which is more efficient and refined than traditional tools, Color Adjustment: Automatically analyzes and suggests color schemes suitable for a particular design theme, or adjusts the color balance and tone of an existing image. Color Adjustment.

2.2. Step-by-step process of Artificial Intelligence Involvement in design:

1) Determine creative direction and refine key themes. At the initial stage of the design process, designers will gain an in-depth understanding of key information such as project background, target audience and market trends. Through the comprehensive analysis of this information, the designer extracts keywords and themes that are closely related to the project. These keywords and themes not only provide a clear direction for subsequent creative conceptualization, but also ensure a high degree of compatibility between the design solution and the project needs.

2) Prepare a prompt vocabulary for in-depth guidance generation. With the continuous development of AI technology, especially the release of the latest AI models such as ChatGPT 4.0, the dialog function of AI drawing tools has become an important part of the design process. At this stage, designers interact deeply with the AI drawing tools and guide the tools to generate images that meet the design requirements by writing words with guiding hints.

3) Efficient automation of the design process. With the help of advanced AI technology, designers can get rid of tedious and repetitive manual operations and realize the automation and intelligence of the design process. In this process, artificial intelligence plays a central role. From the initial conceptualization to the final design product, AI can assist designers to complete a series of complex design tasks.

2.3. Automated design generation based on diffusion modeling

2.3.1. Stabilizing diffusion models:

1) Forward diffusion process. The forward process, also known as the diffusion process, utilizes a specific diffusion mechanism in which the DDPM converts the output image X_0 to Gaussian white noise $X_t \sim N(0,1)$ in T time step during model training [20].

For the original data $x_0 \sim q(x_0)$, each step adds the following Gaussian noise to the data x_{t-1} from the previous step:

$$q(X_t | X_{t-1}) = N\left(X_t; \sqrt{1-\beta_t} X_t, \beta_t I\right) \tag{1}$$

In each step of the diffusion process, a specific variance value β_t, with a value range of 0-1, is involved, and the variance value gradually increases with the gradual increase in the number of diffusion steps. When the number of diffusion steps accumulates to a certain degree, the final generated X_T will no longer retain the characteristics of the original data and will be completely transformed into random noise. Throughout the diffusion process, each step of diffusion generates a data image X_t with noise, and the output of each step depends on the state of the previous step and affects the development of the subsequent steps, which is expressed by the formula (2):

$$q\left(X_{1:T}\left|X_0\right.\right)=\prod_{t=1}^{T}q\left(X_t\left|X_{t-1}\right.\right) \tag{2}$$

The diffusion process is based on sampling the original data for any t-step x_t to obtain the distribution $x_t\sim q\left(x_t\left|x_0\right.\right)$. Definitions $\alpha_t=1-\beta_t$ and $\overline{\alpha}=\prod_{i=1}^{t}\alpha_i$ are obtained by the reparameterization trick:

$$X_t=\sqrt{\alpha}X_0+\sqrt{1-\overline{\alpha_t}}\varepsilon \tag{3}$$

where $\left\{\alpha_t\right\}_{t=1}^{T}$ is a preset hyperparameter value and $\varepsilon_{t-1}\sim N\left(0,1\right)$ is a random Gaussian noise, which can be obtained by performing the inverse parameterization trick:

$$q\left(X_t\left|X_{t-1}\right.\right)=N\left(X_t;\sqrt{\overline{\alpha_t}}X_0,\left(1-\overline{\alpha_t}\right)l\right) \tag{4}$$

By introducing $\alpha_t=1-\beta_t$, x_t can be viewed as a linear combination with the error I, and by simply controlling α_t to be as close to 0 as possible, the distribution of x_t can be obtained close to that of a random noise.

2) Reverse generation process. The forward diffusion process, is the process of applying noise to the image, then the corresponding reverse process is a process of eliminating the noise to restore the image. The DDPM model sets up a parameterized neural network θ to simulate this reverse generation process, the distribution of the reverse generation process is denoted as $p_\theta\left(x_{t-1}\left|x_t\right.\right)$, and the form of its distribution follows a Gaussian distribution. Its mean μ_θ and variance σ_θ are parameterized with x_t and t as outputs, then there are:

$$p_\theta\left(x_{t-1}\left|x_t\right.\right)=N\left(X_{t-1};\mu_\theta\left(X_t,t\right),\sigma_\theta\left(X_t,t\right)\right) \tag{5}$$

In order to simplify the training process of the neural network and optimize the computational efficiency, the variance σ_θ is set to a constant β_t, and the neural network is used to train the mean. The posterior probability $q\left(X_{t-1}\left|X_0,X_0\right.\right)$ is calculated based on the X_t and X_0 of the t moments, and then obtained using the Bayesian formula:

$$q\left(X_{t-1}\left|X_t,X_0\right.\right)=q\left(X_t\left|X_{t-1},X_0\right.\right)\frac{q\left(X_{t-1}\left|X_0\right.\right)}{q\left(X_t\left|X_0\right.\right)} \tag{6}$$

This is obtained from the properties of the Gaussian distribution:

$$q\left(X_{t-1}\left|X_t,X_0\right.\right)=N\left(X_{t-1};H_t\left(X_t,X_0\right),\beta_t\right) \tag{7}$$

Among them:

$$\tilde{\mu}_t\left(X_t,X_0\right)=\frac{\sqrt{\alpha_{t-1}}\beta_t'}{1-\overline{\alpha_t}}X_0+\frac{\sqrt{\overline{\alpha}_t\left(1-\overline{\alpha}_{t-1}\right)}}{1-\overline{a_t}}X_t \tag{8}$$

$$\beta_t = \frac{1-\bar{a}_{t-1}}{1-\bar{\alpha}_t}\beta_t' \tag{9}$$

The final reduction of the noisy image x_t to the original image x_0 is given by the formula as:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}x_t - \frac{\sqrt{1-\alpha_t}}{\sqrt{\alpha_t}}\varepsilon_\theta\left(x_t, t\right) + \sigma_t \tag{10}$$

2.3.2. Selection of automated design diffusion models:

1) Stable diffusion model. The stable diffusion probabilistic model consists of three parts: the variational autoencoder (VAE), the U-Net network structure, and the text encoder [21].

(1) Variational Auto-Encoder. The variational self-encoder for stabilizing the diffusion probability model consists of two parts: encoder E and decoder D. Among them, the main function of the encoder ε is to spatially map E the image x, convert it to a tensor $z = \varepsilon(x)$ in the potential space, and transmit it to the toward U-Net network. In this process, the image space is represented as $x \in R^{H\times W\times 3}$, where H, W denotes the dimensions of the image, $z \in R^{h\times w\times c}$ denotes the potential space, and h, w, c denote the potential space dimensions. The role of the decoder, on the other hand, is to map the image tensor z in the potential space into the image space, represented as the image form $\tilde{x} = \mathrm{D}(z) = \mathrm{D}(\varepsilon(x))$. This encoding-decoding process ensures that the image information is efficiently transformed between the potential space and the image space.

(2) Unet network structure. In the diffusion model, the role of Unet is to receive “noisy” input and predict the noise to achieve denoising, and its basic structure is shown in Fig. 1. The Unet structure is also a coding-decoding structure, which consists of three parts: down-sampling, up-sampling, and skip-linking. The decoder uses the inverse convolution to upsample the data layer by layer, so that the spatial resolution of the data becomes smaller because of pooling, but the number of channels of the data gradually becomes larger because of the role of convolution, so that it can learn the high-level semantic information of the image. Eventually, the decoder maps the low-resolution feature maps into pixel-level result maps that match the original data dimensions.In order to compensate for the information that may be lost during the downsampling process in the encoding stage, the U-Net algorithm introduces a hopping mechanism between the encoder and decoder, which makes the feature maps of the corresponding positions of the encoder and the decoder able to be fused, and ensures that the decoder is able to combine the feature information of the different levels in the upsampling, so that the decoder is able to recover the spatial and edge information of the original image, which is then gradually recovered. This mechanism enables the encoder and decoder to merge the feature maps at the corresponding positions, and ensures that the decoder can combine different levels of feature information in the up-sampling process, so as to more accurately recover and improve the details of the original image.

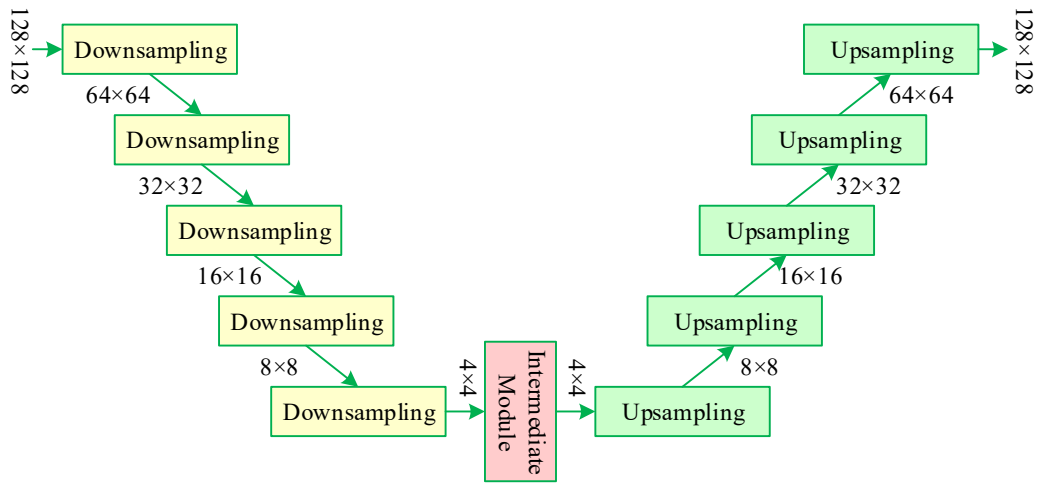


Fig. 1. Basic structure of Unet

(3) Text Encoder. Inputting additional information into Unet to enable additional control over the generated image is called conditional generation. For a given noisy image, a textual description of the desired image can be input and the purely noisy data can be used as a starting point to enable the model to denoise the noise to produce an image that can match the textual description. The textual encoder of CLIP will convert the textual description into a feature vector.

In order to obtain the relevant information of the text description, it is necessary to use a pre-trained Transformer model to compare with CLIP. CLIP's text encoder will convert the text description into a feature vector, which is used to compare the similarity with the image feature vector, so CLIP is suitable for creating useful representational information for the image from the text description. the input text cue is firstly split into words, and the words in the sentence are separated into two parts, and the words in the sentence are separated into two parts. Input text cues are firstly subdivided, converting words or phrases in a sentence into clauses, and then fed into the text encoder, which produces a 768-dimensional vector for each clause. for consistency in the input format, text cues are completed or truncated into 77 clauses, so that each text cue ends up as a tensor shaped like a 77x768 tensor for the representation of the generative condition.

2) The Role of ControlNet in Stable Diffusion Modeling. Stabilized diffusion model relies only on textual cues in generating content, however, the descriptive ability of text is limited, and many details of an image cannot be clearly described by text; therefore, a control architecture called “ControlNet” is proposed, which is an end-to-end neural network architecture that controls the diffusion model to learn the input conditions for a specific task. This is an end-to-end neural network architecture that controls the diffusion model to learn the input conditions for a particular task.

2.4. Loss function design

In the training process, an image z_0 is given and noise is gradually added to it to generate a noisy image z_t, where t denotes the number of times noise is added. As the noise is added, the image will gradually converge to a purely noisy state. Given a time step t, a textual cue c_t, and a task control condition c_f, the model is enabled to learn a network $\grave{o}_\theta$ that is capable of processing images with added noise z_t. The learning objective can be expressed as:

$$L = E_{z_0,t},c_t,c_f,\grave{o}_c,N(0,1)\left[\left\|\grave{o}-\grave{o}_\theta\left(z_0,t,c_t,c_f,\grave{o}_c\right)\right\|_2^2\right] \tag{11}$$

The loss functions used in this chapter include perceptual loss, adversarial loss, and cyclic consistency loss. The combination of the three loss functions can achieve good results in unsupervised image transformation tasks and improve the quality and diversity of generated images.

Among them, the perceptual loss is a loss function based on pre-trained deep learning models, which evaluates the quality of the generated model by comparing the difference between the generated image and the real image in the high-level feature space. This method can better capture the perceptual similarity between the generated data and the real data, making the generated image visually closer to the real image. The specific formula is as follows:

$$L_{perceptual} = \Sigma_i \left\| F_i(x) - F_i(y) \right\|_p \tag{12}$$

where x is the input image, y is the target image, F_i denotes the feature extraction function for layer i in the pre-trained neural network, Σ_i denotes the summation of feature differences across all relevant layers, and $\left\| F_i(x) - F_i(y) \right\|_p$ denotes the number of p-paradigms (usually $p=1$ or $p=2$).

The cyclic consistency loss is mainly used in models such as stabilizing diffusion to ensure that the image is restored to its initial state as much as possible when it is transformed through the generator and then transformed back to the original domain. This helps to maintain consistent information and structural stability during the transformation process and prevents the model from producing fixed outputs that are independent of the input. The specific formulas are as follows:

$$L_{cycle} = \left\| G\left(F\left(x\right)\right) - x \right\|_1 + \left\| G\left(F\left(y\right)\right) - y \right\|_1 \tag{13}$$

Adversarial loss to motivate the converted output to match the target domain. This subsection uses two adversarial discriminators $D(x)$ and $D(y)$ that aim to classify the real image of the corresponding domain with the converted image. The specific formulas are as follows:

$$L_{adv} = E\left[\log\left(D\left(x\right)\right) \right] + E\left[\log\left(1 - D\left(G\left(z\right)\right)\right) \right] \tag{14}$$

The final loss function is as follows:

$$L_{total} = \lambda_{perceptual} * L_{perceptual} + \lambda_{adv} * L_{adv} + \lambda_{cycle} * L_{cycle} \tag{15}$$

where $\lambda_{perceptual}$, λ_{adv} and λ_{cycle} are hyperparameters used to adjust the weights between different loss functions.

3. Automated design generation based on improved stable diffusion models

3.1. LoRA model construction

The stable diffusion model is a larger network model, and training the model usually requires a larger amount of data and training costs. Existing datasets of design samples are scarce and the variety of samples is small and very abstract. Therefore, it is very difficult for training a stable diffusion model that specializes in generating design samples by adding fine-tuned network models that can retain the features of the original samples as well as actively generate new kinds of samples.

The LoRA algorithm, originally used to fine-tune large language models, significantly reduces the size of the trainable parameters required for downstream tasks by decomposing the weight matrix ΔW, which needs to be updated for a large model, into the product of the trainable matrix A and matrix B by means of a low-rank decomposition. Assuming that the weight matrix of the original model is W, the updated parameters can be expressed as:

$$\Delta W = AWB \tag{16}$$

In this paper, a two-dimensional convolutional layer (Conv2d) is used to realize the feature extraction of the input image, and the main structure of the model is shown in Figure 2.

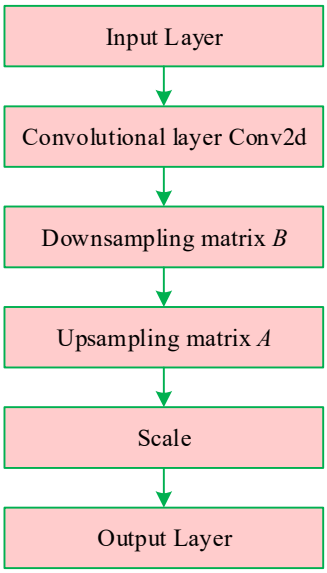


Fig. 2. The structure of LoRA algorithm

The input image is processed through the fusion of the results of the convolutional, downsampling, and upsampling layers to form the final output features, and the updated parameters are saved as a stylized LoRA model. The scaling factor Scale is used to control the update weights of matrix A. The actual fine-tuned update weights are denoted as:

$$\Delta W = AScaleWB \tag{17}$$

When the Stable Diffusion model generates images, the parameters of the stylized LoRA model are injected into the cross-attention layer in the U-Net network, so that the generated images have the features of the images in the training set of the LoRA model, and the magnitude of the expression of the features is controlled by the scaling factor, so as to achieve the purpose of image style migration.

3.2. Generate method flow

The LoRA fine-tuning network architecture, shown in Fig. 3, freezes the pre-trained model weights and injects trainable rank decomposition matrices into each layer of the Transformer architecture, greatly reducing the number of trainable parameters for the downstream task. LoRA injects trainable layers in each Transformer block because there is no need to compute the gradient for most of the model weights, greatly reducing the number of training parameters required and reduces GPU memory requirements. The quality of fine-tuning using LoRA is comparable to full-model fine-tuning and is much faster. LoRA-based fine-tuning preserves the new weights, and the generated LoRA weights can be considered as patch weights of an original pre-trained model. Therefore, the LoRA model cannot be used alone, but needs to be paired with the original model, and the two are merged to obtain the full weights.

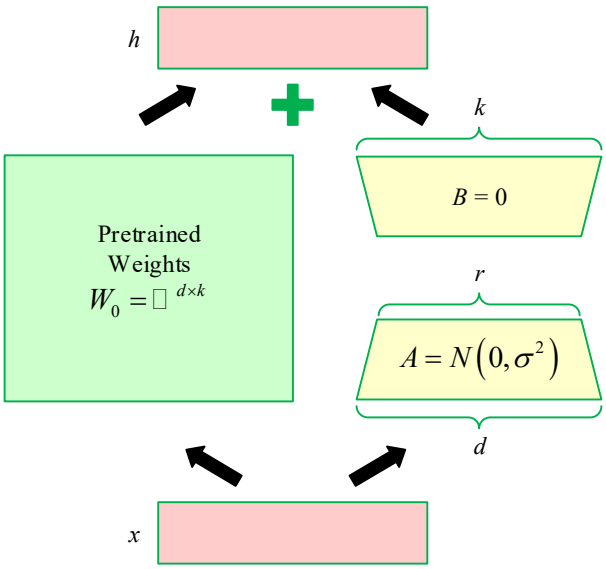


Fig. 3. LoRA Fine-tune network structure diagram

There are many dense parameter matrices during the fine-tuning training process of the large model, but the change matrix of the parameters during the fine-tuning process is not full rank, as shown in Eq. (18). The $d \times d$ -dimensional matrices are transformed by matrix determinant transformation to finally get $d \times r$ -dimensional matrices:

$$\begin{bmatrix} a & b & c & \cdots & d \\ 2a & 2b & 2c & \cdots & 2d \\ 4a & 4b & 4c & \cdots & 4d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a & 2b & 3c & \cdots & 4d \end{bmatrix} d \times d = \begin{bmatrix} a & b & c & \cdots & d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a & 2b & 3c & \cdots & 4d \end{bmatrix} d \times r \tag{18}$$

Utilizing this feature, for the pre-trained weight matrix $W_o \in \square^{d \times k}$, it can be represented by a low-rank decomposition with updates $W_o + \Delta W = W_o + BA$, $B \in \square^{d \times r}$, $A \in \square^{r \times k}$ and rank $r \square \min(d,k)$. During training, W_0 is frozen and does not receive gradient updates, while A and B are trainable parameters. Both W_0 and $\Delta W = BA$ are multiplied by the same input. For $h = W_0 x$, the forward propagation becomes:

$$h = W_0 x + WWx = W_0 x + BAx \tag{19}$$

A random Gaussian initialization is used for matrix A, and matrix B is initialized with 0, so that $\Delta W = BA$ starts with 0 at the beginning of training.

In the improved stable diffusion model, the fine-tuning task can be accomplished by adding the LoRA fine-tuning technique to the U-net's attention mechanism layer, which is the Spatial Transformer layer.

4. Analysis of the application effect of the improved stabilized diffusion model

4.1. Comparative Experiments and Analysis of Results

This paper compares with the current state-of-the-art techniques for image design, namely Imagic, DiffEdit, and InstructPix2Pix methods, and also, since the method in this paper introduces the LoRA fine-tuning method on the graph-generated map of the stabilized diffusion model (Img2Img), the graph-generated map of the standard stabilized diffusion model (Img2Img) is added to the comparison experiment for verifying whether the introduction of the LoRA fine-tuning method is

effective on the results. The 100 tasks in TedBench are categorized into four types of tasks, namely, non-rigid editing (pose conversion, shape adjustment), background replacement, subject replacement, and object addition, according to the different characteristics of the tasks, to compare the advantages and disadvantages of different methods on different tasks. And, to show the fairness, the comparison experiment does not use the two-layer U-net structure.

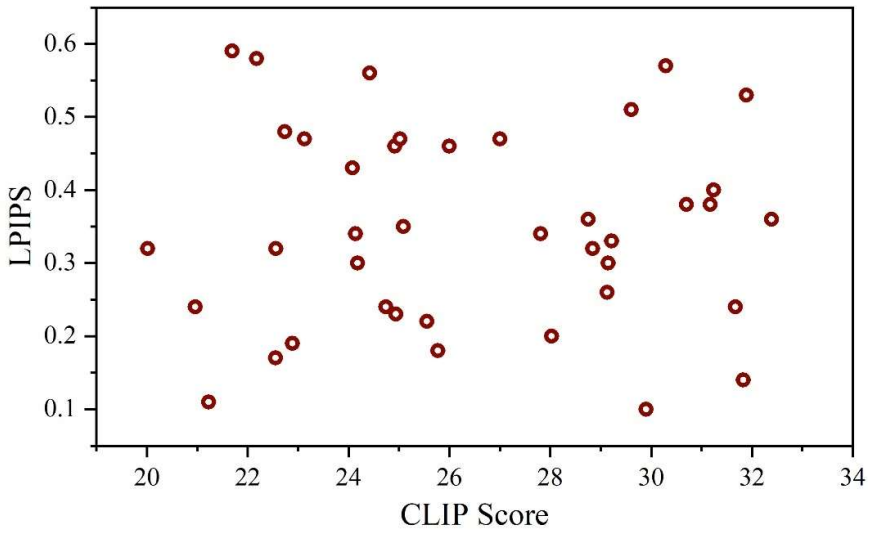
4.1.1. Comparison of design results and fidelity data. According to the practical experience, the CLIP Score of the design result is greater than 20 is regarded as the design is effective, and the five methods are effective for the design of the four types of tasks in the TedBench dataset as shown in Table 1, in which the method of this paper and Imagic are effective for most of the design tasks, with the effective rate of 98.33%, 100%, 100%, 100%, and 97%, and in the five methods The rankings are all first. The number of tasks completed by the standard Img2Img is lower than that of this paper's method, indicating that this paper's method is stronger than the standard method in terms of image design effectiveness after the introduction of LoRA fine-tuning.DiffEdit and InstructPix2Pix have difficulty in completing non-rigid editing tasks.

In terms of data comparison, because DiffEdit, InstructPix2pix can not meet the demand in many non-rigid editing tasks, so this paper and the same use of fine-tuning technology Imagic for data comparison, the TedBench in the 100 task design results of the CLIP Score and the LPIPS to find the average value of this paper's CLIP The CLIP Score of this paper is 25.9645. In the metric of LPIPS, the method of this paper shows a decrease of 39.15% compared to the standard Imagic2Img method, which indicates that the introduction of LoRA fine-tuning method is effective in improving the fidelity, meanwhile, the method of this paper shows a decrease of 34.03% compared to the leading Imagic method.

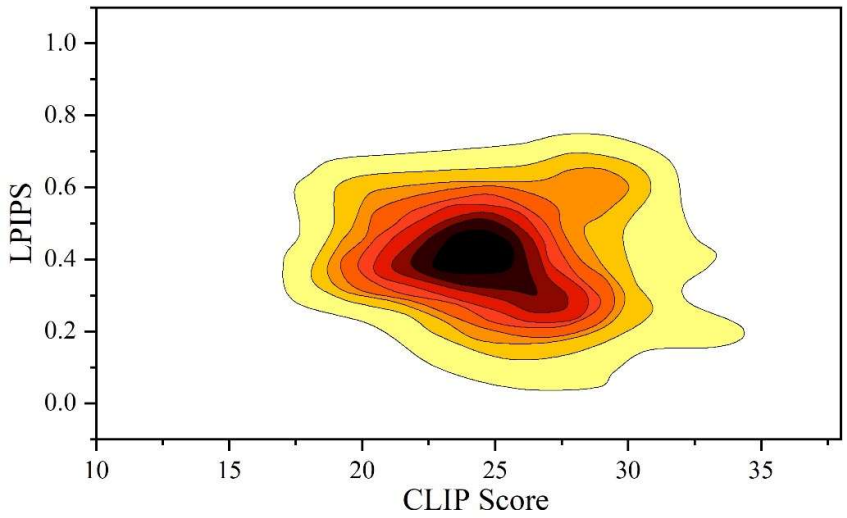
Table 1. The design is effective

Method	Nonrigid editing(60)	Object increase(21)	Subject substitution(11)	Background replacement(8)	General task(100)
This method	59	21	11	8	97
Imagic	55	20	10	8	96
Img2Img	47	15	8	6	68
DiffEdit	12	3	10	2	24
InstructPix2Pix	11	12	9	3	43
Method	CLIP Score			LPIPS	
Imagic	25.3454			0.5856	
Img2Img	23.6975			0.6348	
This method	25.9645			0.3863	

4.1.2. Fidelity analysis. Fig. 4 shows the specific fidelity analysis of this paper's method, Fig. (a) shows the scatter plot and Fig. (b) shows the kernel density estimation, in Fig. (a), it is seen that the CLIP Score interval of this paper's method ranges from 20 to 32.5, while the LPIPS ranges from 0.1 to 0.6. In Fig. (b), the kernel density centroid is (24.85, 0.4), which implies that the proposed method in this paper meets the design requirements with a higher fidelity while accomplishing the image design task.



(a)Scatter point



(b)Nuclear density estimation

Fig. 4. Comparison chart with Imagic result

4.2. Artistic evaluation

4.2.1. Subjective evaluation content and results. In this paper, the research was conducted with an online questionnaire, the total number of questions was 8, and 266 valid feedbacks were received. Among them, from the perspective of occupation, there were 163 art and design-related students or practitioners, accounting for about 61.28%, and 103 other practitioners, accounting for about 38.72%. From the point of view of the understanding of design, 56 people (21.05%) knew about it, and 210 people (78.95%) did not know about it. The 1000 automated designed images were categorized into 5 categories of simple abstract type, complex figurative type, separate, continuous and integrated groups, one category and 5 major questions. For each category, 10 representative generated comparison images were selected and put into the questionnaire, and 3 comparison images were randomly selected for each question per person when answering the questions, and the average number of times each comparison image appeared was 75 times. Respondents were asked to score and evaluate these 15 design comparison charts in groups. Respondents were asked to rate each group of designs on two dimensions and four questions. The scoring is set according to the

"Likert five-point scale method", and the Likert five-point scale is divided into five levels with corresponding scores, "1" represents negative and absolute negative, and "5" represents positive and absolute affirmation.

The purpose of this questionnaire is to compare the performance of Imagic and SD with the introduction of the improved LoRA algorithm in generating design patterns. The questionnaire was distributed to art and design professionals and non-art and design professionals to obtain the subjective evaluation of the model in generating design images from professional and non-professional groups. Table 2 shows the subjective assessment contents of the questionnaire and the mean values of individual indicators. The full score of each index is 5 points, and Table 2 shows that the scores are all above 3 points, so the design images generated by the two models perform better and can meet the basic needs of design. From the individual data, it can be seen that the improved SD has higher single mean values than Imagic in the styling (S), color (C), aesthetics (A) and application (P) indicators, while the innovation (Cr) indicator is in a flat state. Therefore, the subjective evaluation results show that both Imagic and the improved SD are able to complete the design better, but the overall performance of SD is better, with scores of 3.8265, 3.6187, and 4.0622 for stylistic features, color features, and aesthetics, respectively.SD performs better in terms of the generated appearance features, and the inherent performance and utility value of the generated designs are also superior, whereas Imagic's main advantage is the degree of innovation of the generated image, which is 0.1635 lower than that of the method in this paper.

Imagic and the improved SD were analyzed according to the respondents' different cultural and professional backgrounds in terms of the three evaluation dimensions of design appearance characteristics, internal performance and practical value. From the individual data, it can be seen that respondents with different design backgrounds believe that Imagic's strength lies in the creativity (Cr) of generating design images, which received the highest individual scores of 3.5486 and 3.6254, respectively. Imagic's score in the column of creativity (Cr) among respondents who do not know design is 3.6254, which is higher than that of SD's score of 3.5395, which indicates that respondents who do not know design think that Imagic is better than SD's score of 3.5395, which means that respondents who do not know design consider Imagic to be more innovative. design respondents believe that Imagic is more innovative than SD, while SD's strength lies in the aesthetics of the generated tattoos (A), which got the single highest score of 4.1659 and 4.0654 respectively, with scores of more than 4, indicating that they are very satisfied with the aesthetics of the model-generated image designs. Analyzing from the perspective of occupation, respondents from different occupational backgrounds showed a high degree of certainty about the degree of innovation (Cr) of the Imagic generated design, reaching 3.5496 and 3.8125 points respectively. Based on the data, it can be seen that the total score given by the respondents who understand the design is higher than the score given by the respondents who do not understand the design, which indicates that the respondents who understand the design are more satisfied with the design image generation experiment results. It can also be seen that respondents from other professions gave higher overall scores to the images generated by the two models than the scores given by the art design and related practitioners, which side by side confirms that designers have higher requirements for image design generation. Overall, respondents from different backgrounds agreed that the image generated by SD was better than Imagic.

Table 2. Subjective evaluation content and results

Model	Appearance characteristics		Intrinsic performance		Practical value
	Shape	Color	Artistic	Creativity	Application
Imagic	3.1785	3.3588	3.2795	3.4652	3.3796
This method	3.8265	3.6187	4.0622	3.6287	4.1542

Mode	Respondents	Appearance characteristics		Intrinsic performance		Practical value
		Shape	Color	Artistic	Creativity	Application
Imagic	Understand design	3.3155	3.4635	3.3152	3.5486	3.3456
	Don't know design	3.1254	3.3315	3.2485	3.6254	3.3485
This method	Understand design	3.9548	3.8452	4.1325	3.8154	4.0685
	Don't know design	3.7654	3.5564	4.0641	3.5395	3.9875

Model	Respondents	Appearance characteristics		Intrinsic performance		Practical value
		Shape	Color	Artistic	Creativity	Application
Imagic	Art man	3.0778	3.2965	3.1654	3.5496	3.2458
	Others	3.3466	3.4625	3.4598	3.8125	3.6284
This method	Art man	3.7685	3.5987	4.0654	3.6185	3.9485
	Others	3.9275	3.7453	4.1659	3.7935	4.0564

4.2.2. Comparative analysis of subjective evaluations of seven categories of generative designs. Table 3 shows a comparative analysis of the subjective evaluation of the seven categories of generated designs. Image design is rooted in “order” and develops, changes and continues in this paradigm. This paper compares the generation results of the two models in seven different paradigms to explore the organic integration of design and artificial intelligence. The single highest score in the table is SD's aesthetics (A) 4.0625 points, indicating that the aesthetics of SD-generated images are highly recognized by art designers, and SD's application value (P) 3.9722 points, which scores only second to the aesthetics, so that SD-generated designs are also highly evaluated in terms of their application value. Imagic's single highest score for each group is the innovation degree (Cr) 3.5425 points , indicating that its innovative ability is best recognized, and the single lowest score is Imagic's styling features (S) 3.0248 points, so Imagic's styling features in the comprehensive and color groups are evaluated as the lowest.

Overall, the scores in Table 3 are all above 3, so the art-related personnel are basically satisfied with the design images generated by both models, which can satisfy the basic design needs. Imagic has certain advantages in the generation of single line pattern, individual pattern and continuous pattern, with flexibility and harmony. However, in the overall modeling and details of the expression of the more rough, fuzzy deformation and other phenomena, poor clarity, background artifacts affect the aesthetics of the screen. Although SD also has the phenomenon of artifacts, the model is more capable of capturing the complex image structure and detail information, and the generated image has a clear structure, accurate content, and a high degree of match with the original in terms of geometric lines and pattern elements. It is obviously better in terms of tonal unity, overall beauty and attractiveness of colors.

Table 3. Comparative analysis of subjective evaluation of seven types of generated patterns

Group	Generating method	Fractional category	Shape	Color	Artistic	Creativity	Application
Complex representational type	Imagic	Mean	3.1254	3.2756	3.1645	3.5425	3.2154
	This model	Mean	3.7145	3.5184	4.0625	3.6125	3.9722
Simple abstract	Imagic	Mean	3.5865	3.3487	3.1678	3.5426	3.2385
	This model	Mean	3.7485	3.6853	4.0635	3.6157	3.9752
Individual design	Imagic	Mean	3.1625	3.3755	3.1685	3.5425	3.2142
	This model	Mean	3.6578	3.5648	4.0365	3.6187	3.9755
Continuous design	Imagic	Mean	3.1624	3.2498	3.1654	3.5452	3.2154
	This model	Mean	3.7298	3.5475	3.6154	3.6187	3.9752
Monochrome	Imagic	Mean	3.1248	3.4652	3.1685	3.5412	3.2387
	This model	Mean	3.7468	3.5756	4.0364	3.6165	3.9755
Color	Imagic	Mean	3.0425	3.1566	3.1659	3.5469	3.2365
	This model	Mean	3.7156	3.5425	4.0355	3.6189	3.9755
Synthesize	Imagic	Mean	3.0248	3.2348	3.1675	3.5496	3.2655
	This model	Mean	3.8522	3.6248	4.0358	3.6169	3.9725

4.3. Examination of the effect of image generation

4.3.1. Test methods. The study used the questionnaire method to examine the generation quality of the sample images.

Ten industry experts were invited to conduct a questionnaire survey on image generation quality evaluation of the generated 1000 sample images respectively, and 230 questionnaires were issued, with 230 valid questionnaires and 100% effective recovery rate.

Sampling 20 of the images, statistical analysis, respectively, the 20 generated images of the total evaluation of the single score weighted calculation, the weighted average score of the questionnaire scores as the overall evaluation results, respectively, the scores of the 18 indicators for the calculation of the average score of the 20 images in a single indicator. If the value of the weighted mean score and the average score is between 4 and 5, it means that the evaluation result is excellent, between 3 and 4, it means that the evaluation result is good, between 2 and 3, it means that the evaluation result is average, and between 1 and 2, it means that the evaluation result is poor.

4.3.2. Overall evaluation results. According to the questionnaire survey results for statistics, the evaluation results of the secondary indicators in the evaluation of the quality of image generation, using the evaluation results of the secondary indicators after weighted summation to calculate the overall evaluation of a single generated image, the overall evaluation results of the 20 generated images of the specific results are shown in Figure 5.The results of the survey show that, in the 20 images generated, there are 5 images with a good average total evaluation score, respectively, image 6, image 9, image 15, image 16, image 19, with scores of 3.0621, 3.0444, 3.3186, 3.288, 3.3764. , image 15, image 16, and image 19, with scores of 3.0621, 3.0444, 3.3186, 3.288, and 3.3764, respectively. 13 images had average overall evaluation mean scores of fair, and 2 images had average overall evaluation mean scores of poor, with scores lower than 2, namely, image 18 (1.9752) and 20 (1.9024), which indicated that the overall quality of the generated images was fair.

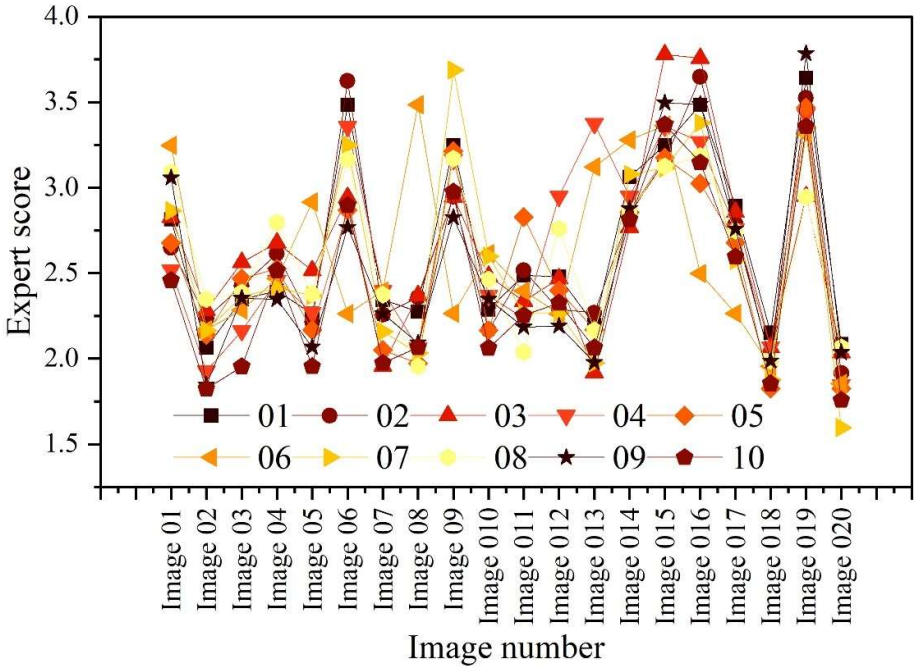


Fig. 5. Generate image quality

4.3.3. Sub-indicator evaluation results. According to the results of the questionnaire survey, the mean value of the secondary indicators is calculated, and the evaluation score of each indicator is shown in Fig. 6, in which the indicators are consistency of content description, consistency of color description, consistency of light and shadow description, consistency of style description, consistency of compositional perspective description, modeling structure and proportion, conformity of image features, reasonableness of contour lines, reasonableness of light and shadow logic, hierarchical relationship of the picture, performance of the main body modeling, picture Detailed performance, color matching effect, material texture performance, layout composition effect, stylistic features performance, visual atmosphere creation, and stylistic unity. The evaluation results of the sub-indicators show that in terms of the consistency between the generated images and the text description, the consistency between the generated effects of image light and shadow and composition and the text description is excellent, and all the sample images accurately represent the lighting effects and light and shadow relations of the outdoor sunny day, and the average value of the consistency is 3.3924. In terms of the reasonableness of the visual logic, the structural proportions of the generated images (3.258), the line outlines (In terms of visual logic rationality, the structure proportion (3.564), line outline (3.058), light and shadow (3.058) and picture hierarchy (3.615) of the generated images are all good, while the image feature conformity is average (2.515), which is inferred to be due to the fact that the semantics of image features is more interpretive and rich, while the logical constraints of other visual elements are more concise and clear in semantics, so that there is a discrepancy in the performance of the generated logics.

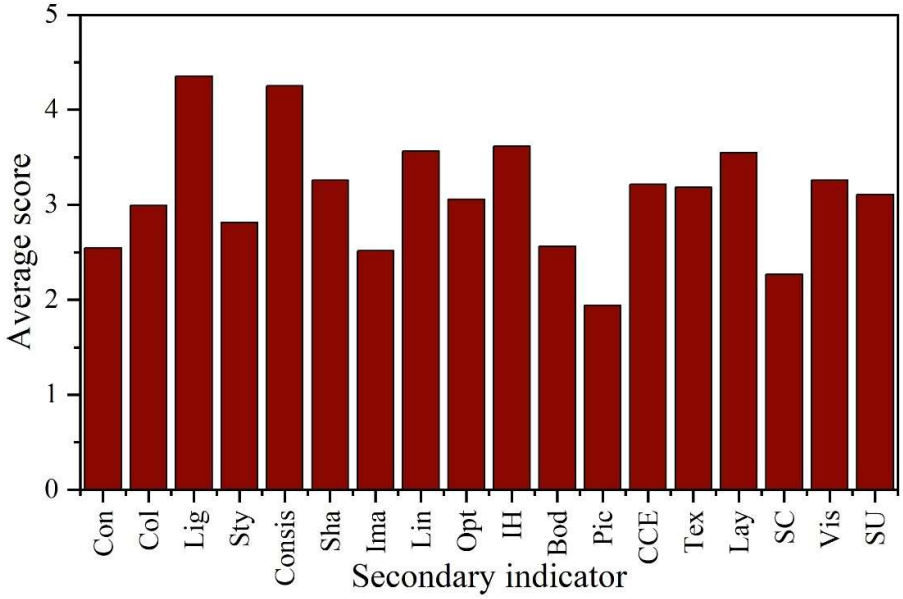


Fig. 6. The average score of the secondary index was evaluated

5. Conclusion

In this paper, with reference to the step-by-step process of AI participation in design, we propose an automated design generation method based on the diffusion model, which consists of a stabilized diffusion model from the three parts of the variational self-encoder, the U-Net network structure, and the text encoder, and designs the loss function. By adding the fine-tuned network model, the design function of the original model is realized.

1) The performance of the improved model is analyzed through comparative experiments, and among the five methods, the design efficiency of the method proposed in this paper is ranked first, with the efficiency of 98.33%, 100%, 100%, 100%, and 97%, respectively. Meanwhile, the CLIP Score is 25.9645, which is 39.15% lower than the standard Imagic2Img method in the LPIPS metrics, indicating that the introduction of LoRA fine-tuning method is effective in improving the fidelity.

2) Compared with the Imagic method, the improved SD are able to complete the design better, and the scores of the model designed in this paper are 3.8265, 3.6187, and 4.0622 in the stylistic features, color features, and aesthetics, respectively. The SD performs very well in the generated appearance features.

3) In the image generation effect test, there are five images that reach a good level, which are image 6, image 9, image 15, image 16, image 19, and the ratings are 3.0621, 3.0444, 3.3186, 3.288, 3.3764, respectively, and the proportion of above average reaches 90%, and the overall generation effect is good.

References

[1] Liu, L. , & Hu, Z. . (2022). Big data analysis technology for artificial intelligence decision-making platform construction and application. Mobile Information Systems.

[2] Sippel, A. R. , He, B. , & Krapf, N. W. . (2021). Applied artificial intelligence technology for narrative generation using an invocable analysis service and data re-organization. US11003866B1.

[3] Yang, X. . (2021). Research on the application of artificial intelligence technology in computer graphics processing. Journal of Physics: Conference Series, 2083.

[4] Chen, Y. , & Chen, Y. . (2021). Research on the application and influence of artificial intelligence

technology. Journal of Physics Conference Series, 1952(4), 042007.

[5] Aukes, D. M. , & Wood, R. J. . (2015). Popupcad: a tool for automated design, fabrication, and analysis of laminate devices. Proceedings of SPIE - The International Society for Optical Engineering, 9467, 94671B-94671B-10.

[6] Choi, W. , Jang, S. , & Kim, Ha YounLee, YuriLee, Sang-GooLee, HanbitPark, Sungchan. (2023). Developing an ai-based automated fashion design system: reflecting the work process of fashion designers. fashion and textiles, 10(1).

[7] Cunney, S. , Cunney, B. , Cunney, M. , Pedrick, K. , & Whitman, J. . (2024). SELF DRAWING TOOL FOR A COMPUTER-IMPLEMENTED AUTOMATED DESIGN, MODELING AND MANUFACTURING SYSTEM. US20240119675A1;US2024000119675A1;US2024119675A1;US2024119675.

[8] K., HEMA, CHANDRA, REDDY, J., & V., et al. (2015). Design and development of an automated tool for job shop scheduling using pso and cpso. International Journal of Mechanical & Production Engineering Research & Development.

[9] Georgiev, H. , Peeva, I. , & Ivanov, A. . (2019). Software tool for automated design of 3d cad models for mechanical engineering education. ICERI2019 Proceedings.

[10] Chen, Y. , Gao, Y. , Zhou, Y. , Chen, M. , & Ma, X. . (2019). Design of an Automated Test Tool Based on Interface Protocol. 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C). IEEE.

[11] Joshi, R. , & Kumar, N. . (2021). Artificial intelligence for autonomous molecular design: a perspective. Molecules (Basel, Switzerland), 26(22).

[12] Stump, A. R. , Carrara, A. , Srinivasan, E. , Como, C. W. , & Billi-Duran, S. M. . (2021). AI DESIGN ANALYSIS AND RECOMMENDATIONS. US20210294279A1.

[13] Ehresmann, M. , Herdrich, G. , & Fasoulas, S. . (2021). An automated system analysis and design tool for spacecrafts. CEAS Space Journal(3).

[14] Zhu, X. . (2021). Background and prospect of the application of artificial intelligence in design. Knowledge Innovation on Design and Culture.

[15] Wangoo, D. P. . (2018). Artificial Intelligence Techniques in Software Engineering for Automated Software Reuse and Design. International Conference on Computing Communication and Automation. Department of CSE & IT, Amity School of Engineering and Technology Amity University, Noida, Uttar Pradesh, India;.

[16] Dragicevic, T. , Wheeler, P. , & Blaabjerg, F. . (2018). Artificial intelligence aided automated design for reliability of power electronic systems. IEEE Transactions on Power Electronics, 34(8), 7161-7171.

[17] Liu, L. . (2022). The Frame Display of the Digital Clothing Display Design System Based on Intelligent Man-Machine Cooperation. The International Conference on Cyber Security Intelligence and Analytics. Springer, Cham.

[18] Liu, X. Y. , Yin, Y. , & Zhou, L. P. . (2010). Research on innovative product design method and support system for man-machine cooperation. Applied Mechanics & Materials, 34-35, 701-706.

[19] Jialei Jiang. (2024). When generative artificial intelligence meets multimodal composition: Rethinking the composition process through an AI-assisted design project. Computers and Composition102883-102883.

[20] Zeng Wang,Hui ru Pan,Jiang shan Li & Shi fan Niu. (2024). Exploring product rendering generation design catering to multi-emotional needs through the Superiority Chart-Entropy Weight method and Stable Diffusion model. Advanced Engineering Informatics(PC),102809-102809.

[21] Dingkun Yan,Ryogo Ito,Ryo Moriai & Suguru Saito. (2023). Two‐Step Training: Adjustable Sketch Colourization via Reference Image and Text Tag. Computer Graphics Forum(6).