ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Exploring the influence and application of AI generation technology on spatial aesthetics in environmental design

Dong Wang[1,*]

[1] Academy of Art Design, Henan Institute of Technology, Xinxiang, Henan, 453000, China

SUMMARY: *AI algorithm-driven generative technology is currently a cutting-edge field in environmental design. The study analyzes the principles and characteristics of neural networks and generative adversarial networks from the perspective of the principles of deep learning technology, and explores their applications in the field of environmental design. Subsequently, based on the generative adversarial network model, we design an environment design generation model with improved generative adversarial network. With the mechanism of moving window, the attention module in the model is improved to construct a hierarchical generative adversarial network model. Finally, the application and impact analysis of the AI generative model in environmental design are explored in the experiment. Through the case study of a historical neighborhood, in the comprehensive analysis of spatial aesthetics, the scores of "positive" scenarios are higher than those of "negative" scenarios. Meanwhile, the overall coordination is an important factor affecting the visual quality of spatial aesthetics.*

KEYWORDS: *AI; Generative Adversarial Networks; Spatial Aesthetics; Environmental Design*

1 Introduction

Since 2016, major breakthroughs in machine deep learning have fueled a research boom in artificial intelligence (AI) generation technology, which also has more applications in the design field.2022 AI generation technology represented by software tools such as ChatGPT and Midjouney has demonstrated its great potential in design creation, which triggered an epochal change in the environmental design industry, and the traditional single design The traditional single design process and model can no longer meet the growing social demand, and the technology-driven design work mode is gradually accepted and recognized by the industry, and gradually applied in practice [1-4]. Compared with the previous environmental design industry, the pursuit of "goal" drive will become less and less, and turn to the pursuit of "creation" drive. From the past purely emphasize the space function of the environmental space form to now pay more attention to the spiritual connotation of the environmental space and cultural character, environmental design more and more the pursuit of personalization, diversification, as far as possible to avoid homogenization and convergence [5, 6]. Designers also from the previous "design goals" backwards workflow to its own perception and understanding of the design site, using a specific form, function, materials, technology and other elements of the combination of the project to complete the overall program planning for program optimization and structural adjustment, and at the same time, according to its cultural value of the space and the site and aesthetic demands and other factors Comprehensive consideration to make creative solutions

*tonyccocco@163.com
https://doi.org/10.65102/is2026162
Published: 30 April, 2026

[7-9].

AI generation technology-driven design work mode change has subverted people's traditional perception of the environmental design industry, while injecting fresh blood into the development of the environmental design industry. Literature [10] utilizes Conditional Generative Adversarial Networks (GAN) to instantly detect sunlight in residential environments from images of residential areas, which is capable of determining the sunlight environment of residential design, and helps to assist in the design of light and shadow aesthetics of built environments. Literature [11] used AI generated content tools to plan nighttime tours of urban culture and tourism by generating thematic content, narrative character and plot design, light and shadow narrative scenes, and immersive videos to design light environments that meet tourists' nighttime adaptation. Literature [12] integrates AI-generated content tools and eco-interactive design, by generating eco-art content with dynamic interaction, providing a unique narrative experience for the interaction, and the user can also participate in the environment design to obtain aesthetic requirements that are more in line with the user's aesthetic requirements. Literature [13] combines machine learning and GAN to quickly obtain and generate the outdoor green space layout plan, which can design design programs belonging to different designers' styles within a reasonable range, but there are rule errors in the integration of some design elements. Literature [14] proposes an automated urban landscape layout design based on AI generation technology and emotional dictionary, which is emotionally oriented and generates a functional layout of roads and land use that meets citizens' emotions. Literature [15] used a large-scale language model to generate visual images of streets to visualize acoustic environments and proposed a soundscape-to-image diffusion model to convert soundscapes into situated visual representations for understanding and describing auditory and visual spatial perceptions to inform urban environmental design. Literature [16] focuses on acoustic features such as pitch range and speech intensity of buildings, and generates architectural images based on acoustic features under AI generation technology to form expressive, smooth and emotionally charged spatial design solutions. Literature [17] explores the application of texture in interior design by AI generation technology such as ChatGPT, which is mainly used by designers to adjust the 3D scene design scheme with the help of material and color suggestions generated based on ChatGPT to assign a more reasonable texture to the interior space. In recent years, the rapid development of AI generation technology has made the mainstream idea of environmental design shift from human-centered to human-machine hybrid, and designers and AI are not in a competitive or replacement relationship, but support each other, systematically advance side by side, and ultimately realize the close coupling of humans and machines. Therefore, it is necessary to deeply explore the influence and application of AI generation technology on spatial aesthetics in environmental design.

The article first combs through the application of deep learning in environmental design from a technical perspective, introducing the ideas of neural networks and generative adversarial networks and their applications. Then a hierarchical generative adversarial network model based on moving window is proposed, which greatly reduces the computational complexity of attention calculation by moving window and window division in generator and discriminator, and realizes the mutual communication between attention windows at the same time. Moreover, the attention mask of the moving window is added to the attention module, which realizes the batch attention calculation of the attention window under the condition that the semantic information of the overall feature map is not damaged. Subsequently, the performance test of the model is conducted in the experiment to explore the effect of the model's environment design image generation. Finally, a historical neighborhood in area A is selected for a case study to explore the impact of the method on spatial aesthetics in environmental design.

2 Environmental design generation model based on AI generation technology

2.1 Deep learning in environmental design

2.1.1 Neural Networks

The article begins with the basic principles of neural networks, on which other neural network models are based.

(1) Neuron model

Neural network is a network composed of a large number of neurons, neuron model is one of the most basic components of neural network, neuron model works by imitating the work of neurons in the human brain. Neurons in the human brain include a large number of dendrites, which are mainly used to receive information from the outside world, and the information is then processed by the nucleus of the cell and transmitted to the axon, which is connected to the dendrites of the next neuron.

MP model is the most basic neuron model of neural network, which simulates the neuron structure of human brain with adaptability. The structure contains multiple inputs and one output, and the inputs and outputs are computed in two layers. Firstly, the input a is multiplied with the weight w and then summed between the input and the neuron to get $a_1 * w_1 + a_2 w_2 + a_3 * w_3$, and then the final output is obtained through the calculation of the activation function, which is set to be the sgn function, expressed as $g(\ )$, and then the output $z = g(a_1 * w_1 + a_2 w_2 + a_3 * w_3)$. Here, the weight w is one of the most important parameters in the neural network, and the w parameter is adjusted to the best through the continuous training of the model to make the whole network can predict the data better.

The activation function here is the sgn function, which maps the input values to 0 or 1, mimicking the human neuron response, such as 1 for neuron excitation and 0 for neuron inhibition. A neural network is a network that connects multiple nerves in a certain hierarchical structure.

$$\operatorname{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1}$$

(2) Single-layer neural network - perceptron

The “Perceptron” model is a neural network consisting of two layers of neurons. It consists of an input layer and an output layer, the input layer is only responsible for input signals, not computation, and the output layer is the MP model. Based on this model structure, the perceptual machine can realize the or, and, and non-problems in logic judgment.

Its expression is:

$$y = f\left(\sum_i w_i x_i - \theta\right) \tag{2}$$

The mathematical expressions for the or, with, and without problems are as follows:

“with” $(x_1 \wedge x_2)$: let $w_1 = w_2 = 1, \theta = 2$, then $y = f(1 \cdot x_1 + 1 \cdot x_2 - 2)$, and $y = 1$ only if $x_1 = x_2 = 1$.

“or” $(x_1 \vee x_2)$: Let $w_1 = w_2 = 1, \theta = 0.5$, then $y = f(1 \cdot x_1 + 1 \cdot x_2 - 0.5)$, and $y = 1$ only

when $x_1 = 1$ or $x_2 = 1$.

"Non" $(-x_1)$: such that $w_1 = -0.6, w_2 = 0, \theta = -0.5$, then $y = f(-0.6 \cdot x_1 + 0 \cdot x_2 + 0.5)$, when $x_1 = 1$ or $x_2 = 1$, $y = 1$

where, given the training dataset, the weights w can be obtained by learning. The weight w and the threshold θ can be obtained by learning, and its learning process is relatively simple, when the input training sample is $(x, \hat{y})$, the output of the perceptual machine is y, then the weights of the perceptual machine are adjusted as follows:

$$w_i \leftarrow w_i + \Delta w_i \tag{3}$$

$$\Delta w_i = \eta (y - \hat{y}) x_i \tag{4}$$

where η is the learning rate, which is one of the hyperparameters adjusted by human beings, not learned by the model. If the perceptual machine predicts correctly when the y in the output data is the same as the given sample $\hat{y}$, the perceptual machine does not make adjustments, and if it is different, the weights are adjusted accordingly according to the degree of error.

(3) Multi-layer neural network

Although the perceptual machine has initially realized the function of learning, prediction, but due to the simple structure, only has a layer of neurons, so the learning ability is limited, can only deal with or and not this kind of linear separable problems, so it is difficult to solve the heterogeneous problem of this kind of nonlinear problems. The contingent and non-contingent problems are linearly differentiable problems, i.e., there exists a linear hyperplane to separate them.

For this problem, it is solved by two methods: activation function, and increasing the number of neural network layers.

1) Activation function

Activation function is a nonlinear function, commonly used activation functions include sgn function, Relu function, Sigmoid function. All are functions that increase the nonlinear ability of the model. When there is no increase in the activation function, the single neuron model expression is as in equation (5), and the single neuron model is shown in Figure 1. The multi-neuron model expression is as in equation (6). At this point the model is a linear function no matter how complex it is, and therefore a decision partition cannot be fully fitted to correctly separate the data in either heteroscedastic problems.

$$y = w_1 x_1 + w_2 x_2 + b \tag{5}$$

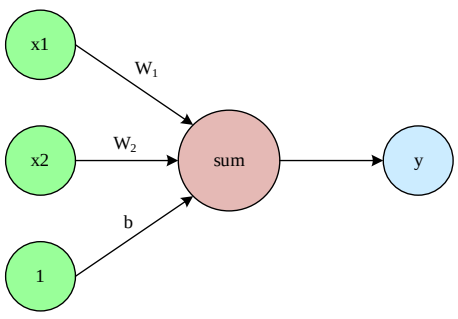


Figure 1: Single neuron model

$$\begin{aligned} y = {} & w_{2-1}\left(w_{1-11}x_1 + w_{1-21}x_2 + b_{1-1}\right) + w_{2-1}\left(w_{1-12}x_1 + w_{1-22}x_2 + b_{1-2}\right) \\ & + w_{2-3}\left(w_{1-13}x_1 + w_{1-23}x_2 + b_{1-3}\right) \end{aligned} \tag{6}$$

And taking the most commonly used sigmoid function as an example, the expression of the sigmoid function is as follows:

$$\sigma(y) = \frac{1}{1+e^{-y}} \tag{7}$$

The neural network incorporating the sigmoid function is as follows:

$$a1 = w_{1-11}x_1 + w_{1-21}x_2 + b_{1-1} \tag{8}$$

$$a2 = w_{1-12}x_1 + w_{1-22}x_2 + b_{1-2} \tag{9}$$

$$a3 = w_{1-13}x_1 + w_{1-23}x_2 + b_{1-3} \tag{10}$$

$$y = \sigma\left(w_{2-1}\sigma(a1) + w_{2-2}\sigma(a2) + w_{2-3}\sigma(a3)\right) \tag{11}$$

2) Increase the number of neural network layers

Multi-layer neural networks add hidden layers to the perceptual machine. Take a single hidden layer neural network as an example, the neural network includes an input layer, an output layer, and a hidden layer. Let the output be y, the input be x, the number of neurons in the former layer of the network is i, the number of neurons in the latter layer of the network is i, and the parameters are w.

The expression is:

$$Y = \sum_{ji} \omega_{ji}\sigma\left(\sum_{ji} \omega_{ji}\sigma\left(\sum_{ji} \omega_{ji}x_j + b_j\right) + b_j\right) \tag{12}$$

If the network contains multiple hidden layers, take a neural network with two hidden layers as an example, the neural network includes an input layer, an output layer, two hidden layers, and a total of four layers of the network. Then its expression is:

$$Y = \sum_{j}\left(\sum_{jI} \omega_{ji}\sigma\left(\sum_{ji} \omega_{ji}\sigma\left(\sum_{ji} \omega_{ji}x_j + b_j\right) + b_j\right) + b_j\right) \tag{13}$$

(4) Learning and training of neural networks Multi-layer networks have stronger learning ability and more complex model structure, so it is not enough to rely on the simple learning rules of the perceptron alone. To address this problem, the back propagation (BP) algorithm has been proposed for the training of multi-layer neural networks, which allows the model can be in the iterative process, automatically according to the output of the model and the training dataset of the error distance to adjust their own parameters.

BP algorithm is one of the most important algorithms for training neural networks, and its purpose is similar to the purpose of perceptual machine learning, i.e., through the continuous iteration of the model, calculate the error between its output and the real sample, and then according to the calculated error, iterate repeatedly to adjust the parameters of the network, so

as to make the output of the algorithm gradually close to the real data.

2.1.2 Generating Adversarial Networks

(1) Principle of Generative Adversarial Network

Generative adversarial network based on the basic idea of game theory is designed in the framework of two "generator" 'discriminator' two types of neural networks, generator to generate fake data, discriminator to determine whether the generator generated results are true, the training process of the two self-game, when the ability of the two can reach a balance, the generator generated content can be "false to true". When the ability of the two can reach a balance, the content generated by the generator can be "false to real".

(2) Application advantages of generative adversarial network

1) Two types of problems in space layout planning and design

The design of small-scale space contains two types of important problems. Its design process requires weighing various design conditions within a certain framework to arrive at diverse design solutions. Among them, the solution of the design framework and the response to the design conditions can be regarded as a kind of convergence problem. After conforming to the established framework, according to the designer's experience, digging out the hidden design conditions and arriving at a variety of design solutions can be regarded as a kind of dispersive problem.

2) Advantages of Generative Adversarial Networks in generative design and its reasons

Generative adversarial network as a deep learning technology, but its logic and deep learning is not the same, the difference between the two is also the key to the generative adversarial network "creativity", as mentioned above, the existing research in the traditional deep learning in the generation of the layout of the class of the application of the convergence problem, so this paper compares the difference between the two techniques to further clarify the applicability of GAN. Therefore, this paper compares the difference between the two techniques to further clarify the reasons why GAN is suitable for dispersive problems and fast program generation.

(3) Other Applications of Generative Adversarial Networks in Planning and Design

The input and output of Generative Adversarial Network are images, so in addition to generating designs, the application in planning design has other tasks related to figure-to-figure translation, so it is suitable for the rapid rendering of architectural, landscape, and planning plans, and the rapid generation of black-and-white line drawings to color renderings.

2.2 Environment design generation based on improved generative adversarial networks

2.2.1 Hierarchical Generative Adversarial Network Modeling

Hierarchical growth GAN models are often used to generate high-resolution images and image generation tasks with style control. When the resolution of the image is very low, the captured features are often about the overall layout of the image and contours of such a large granularity of the features, and with the gradual increase in resolution, the features that can be captured become texture, skin hair details and other small granularity of the features. According to this characteristic, the hierarchical growth of the generative adversarial network generally generates low-resolution images containing the overall layout of the image in the low stage, and then uses the network in the high stage to gradually learn the image texture details and other smaller granularity of the features in order to generate a more realistic, higher-quality images.

2.2.2 Hierarchical Generative Adversarial Network Model Based on Moving Windows

(1) Generator network structure

The overall structure of Hi-SWGAN is shown in Fig. 2. With the input phase, similar to what was done with ViT, it is required that tokens similar to those in the NLP task will be obtained. So in the generator input phase of this structure, random noise z is first fed into the mapping network, and the noise is transformed into a long sequence of $H \times W \times C$ using a mapping consisting of MLPs. This sequence is then transformed into tokens of dimension C and length $H \times W$. Each token is then combined with a learnable positional encoding as an input to the Swin Transformer block, which unlike ViT does not add the class token in the first place here.The Swin Transformer block takes the each token sequence as input and recursively computes the correspondence between each token sequence. Since the design of Swin Transformer does not change the scale of the tokens, the input and output sizes of the Transformer module remain consistent.

The one-dimensional sequence output by Swin Transformer is then up-sampled and deformed into a two-dimensional feature map, $p \in \mathbb{R}^{H \times W \times C}$. This two-dimensional feature map is fed into a sub-pixel convolutional layer (SSub-Pixel CNN), where the input low-resolution feature map can be transformed into a higher-resolution feature map through convolution and multi-channel reorganization. The original low-resolution pixel is first divided into r^2 channels of feature maps by convolution, and then the high-resolution feature maps are up-sampled using a periodic filtering method, and the magnification of the image expansion is controlled by the up-sampling factor.

The result after the Sub-Pixel CNN layer is a $p' \in \mathbb{R}^{2H \times 2W \times \frac{C}{4}}$ feature map. It is then converted again to a one-dimensional sequence and reshaped into a token sequence of length $4HW$ and embedding dimension $\frac{C}{4}$. The previous two steps are repeated until the desired resolution is reached, and then the linear flattening layer is input to scale the vector dimension to 3, resulting in an RGB image.

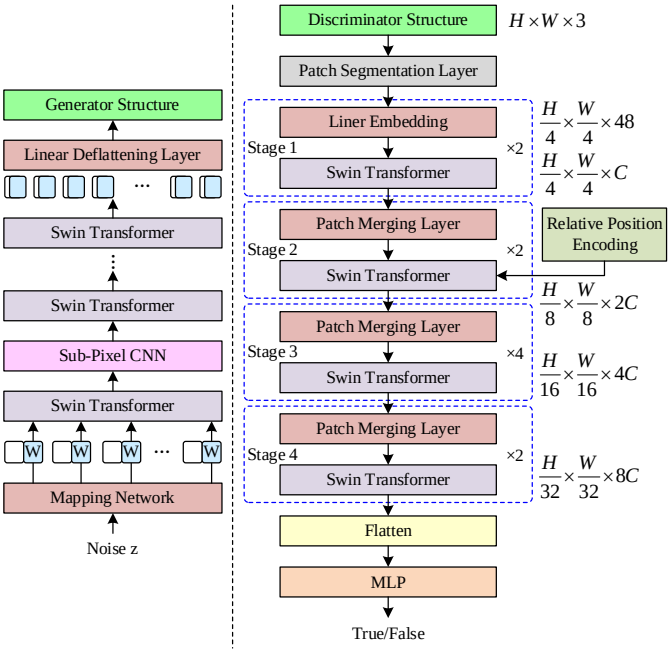


Figure 2: Hi-SWGAN overall structure

(2) Discriminator Network Structure

For the discriminator, if the resolution of the image is very high, the length of the sequence becomes huge if the pixel point is used as the basic unit, so the main goal of the discriminator design is to reduce the length of the sequence and to obtain multi-scale features as much as possible without affecting the global attention of the Transformer. Similar to ViT, in the discriminator, the input image with size $H \times W \times 3$ is first segmented by Patch, and each Patch is regarded as a token. if the size of the Patch is set to 4×4, the size of the image obtained after passing through the Patch segmentation layer is $\frac{H}{4} \times \frac{W}{4} \times 48$. Next, the dimension of the vector is transformed into a dimension acceptable to the Transformer through a linear embedding layer, which is determined by a pre-set hyperparameter C, C is usually 96, so the dimension becomes $\frac{H}{4} \times \frac{W}{4} \times C$, and the Patch is stretched at this point to a dimension for C long sequence of length HW.

If the size of the initial picture is added as $224 \times 224 \times 3$, then the length of this sequence will reach 3136, and this length of sequence is difficult for Transformer to accept, and the computational complexity of calculating the attention will be large. So here Patch merging is performed to introduce the computation of the attention window. The Swin Transformer Block in the model is based on the window to calculate the self-attention, the output of the first Swin Transformer Block is still $56 \times 56 \times 96$. If we want to obtain more scale feature information, it is necessary to construct a hierarchical Transformer, and here we need to use the operation of patch merging.

Patch Merging operation is essentially to merge neighboring small patches into a large patch to achieve the effect of a downsampled feature map. patch Merging operation flow is shown in Fig. 3, if you want to carry out twice the downsampling, every other patch is merged together, in the figure of the number of the same that is one position apart from the patch. After the merger, one tensor becomes four tensors, if the original tensor dimension is $H \times W \times C$, then the size of each tensor is now $\frac{H}{2} \times \frac{W}{2}$, and then these four tensors are spliced in the dimension of C. The size of that tensor then becomes $\frac{H}{2} \times \frac{W}{2} \times 4C$, which is equivalent to trading spatial dimensionality for a larger number of channels.

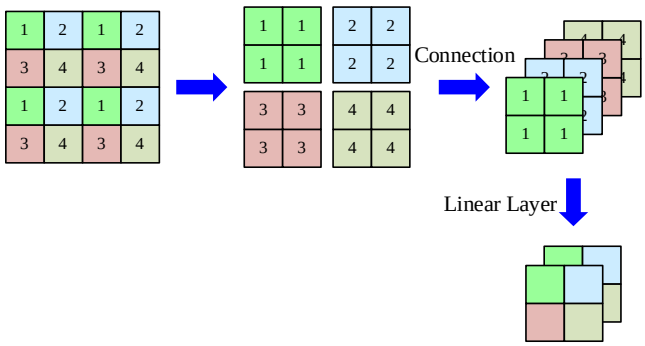


Figure 3: Patch Merging operation process

(3) Self-attention calculation based on moving window

This scheme adopts the method of moving window, first of all, the individual feature graph will be divided into windows to get a number of small windows, the attention calculation for each window, that will significantly reduce the length of the sequence, which makes the

complexity of the attention calculation is greatly reduced.

Assuming that the original size of the feature map is $h \times w \times C$, it will be divided into a number of red-bordered windows, but the smallest computational unit is still the Patch, that is, the gray-bordered squares in the figure. Each window contains $M \times M$ Pacth, i.e., the whole tensor is partitioned into $\frac{h}{M} \times \frac{w}{M}$ windows, and attention is computed in each of these $\frac{h}{M} \times \frac{w}{M}$ windows.

$$\begin{aligned} \Omega(W-MSA) &= (h/M)(w/M)\left(4M^2C^2 + 2M^4C\right) \\ &= 4hwC^2 + 2hwM^2C \end{aligned} \tag{14}$$

In multi-head self-attention, linear projection yields Q, K, V, and the input vector dimension (hw, C) passes through the linear mapping layer, and the computational complexity of this step is $3hwC^2$. The computational complexity of the dot product operation of self-attention is $2(hw)^2C$, and then the inner product with V after Mask and softmax will produce a computational complexity of hwC^2. For the division of the window for self-attention calculation, just bring in the formula, the computational complexity of each window is $4M^2C^2 + 2hwM^2C$, multiply the number of windows $\frac{h}{M} \times \frac{w}{M}$ by the computational complexity required by each window to get the computational complexity of the self-attention calculation after the division of the window. The number of windows will not be very large, so the difference between M^2 and hw can be tens or even hundreds of times, which is good for reducing memory consumption and computational complexity.

In the Transformer module will be the first window division, and then move the window operation of these two operations, a window based on the multi-head self-attention and then do another based on the moving window of the multi-head self-attention, so that the window and the window to communicate with each other, so that the combination of the two modules can be considered as a complete Swin Transformer computing unit! The Swin Transformer module is shown in Figure 4. The module's attention of the

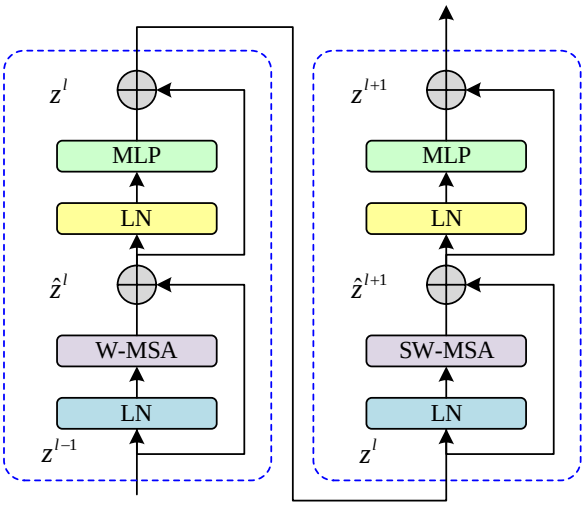


Figure 4: Swin Transformer Module

The calculation is restricted to each window as shown in equation (15):

$$Attention(Q,K,V) = Soft\max\left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{15}$$

(4) Self-attention masking based on moving windows

In order to keep the number of windows unchanged after moving and the number of patches in each window consistent, it is necessary to add the window attention mask operation in the moving window.

After the operation of moving windows, the feature map is partitioned into nine windows, and instead of calculating the self-attention on these nine windows, a cyclic shift is done first. Among the four newly partitioned windows, the relative positions of the patches in the first window are unchanged and can be used for attention calculation directly, but the other windows are spliced together with different ratios of non-neighboring contents and cannot be used for attention calculation directly. In order to correctly calculate the self-attention of each window without being interfered by the non-neighboring parts, a masking operation is needed.

2.2.3 Loss function

This scheme adopts the regularization method of gradient penalty and uses the Wasserstein distance instead of the JS scatter in the original GAN model to measure the distance between two distributions, and its superiority over the JS scatter and KL scatter is that even if the two distributions do not overlap with each other, the Wasserstein distance can still reflect their distance.

The loss function of the discriminator in the hierarchical generative adversarial network based on moving window proposed in this chapter is shown in equation (16):

$$\begin{aligned}\mathcal{L}_D = {} & \mathbb{E}_{\tilde{x}\sim p_g} D(\tilde{x}) - \mathbb{E}_{x\sim p_z(x)} D(x) \\ & + \lambda_{gp}\mathbb{E}_{\hat{x}\sim p(\hat{x})}\left[\left(\left\|\nabla_{\hat{x}} D_w(\hat{x})\right\|_2 - 1\right)^2\right]\end{aligned} \tag{16}$$

where $p_z(z)$ denotes the distribution of the real data, p_g is the distribution of the data generated by the generator, λ_{gp} denotes the weight of the penalty term, $\hat{x}$ is the random sample, $\|\cdot\|_2$ denotes the Euclidean paradigm, and ∇ is the gradient solver. $\hat{x}$ is the real image sample x and the generated image sample $\tilde{x}$ obtained by stochastic difference, and the process is shown in equation (17).

$$\hat{x} = \epsilon x + (1-\epsilon)\tilde{x} \tag{17}$$

where ϵ is a 0 to 1 uniform distribution, and $\left\|\nabla_{\hat{x}} D_w(\hat{x})\right\|_2 \neq 1$ is penalized as soon as it is possible to do so, with the penalty increasing as the distance from 1 increases, making the gradient as equal to 1 as possible.

The loss function of the generator G is shown in equation (18):

$$\mathcal{L}_G = -\mathbb{E}_{\tilde{x}\sim p_g} D(\tilde{x}) \tag{18}$$

3 Application and impact analysis of AI-generated models in environmental design

3.1 Hi-SWGAN model performance testing experiments

3.1.1 Data set establishment

In this thesis, the model is trained with the style dataset used in Hi-SWGAN and the images collected by web crawler to obtain seven styles of environmental design images, including 5225 Photo original images, 583 traditionalist style images, 1123 modernist style images, 546 minimalist style images, 433 industrial style images, 846 naturalistic style images,758 eclectic style images,936 sketch futuristic style images. Each image is 512x512 pixels in size.

3.1.2 Style discretization experiments

In order to test the utility of the discretized loss function, this thesis performs a style transformation on Photo with four datasets (traditionalist style, modernist style, minimalist style, and industrialist style) and judges the migration results on subjective and objective dimensions.

(1) Comparison of Indicators of Each Model

This thesis evaluates the models from three indexes: FID, PSNR and SSIM. 16 groups of models are distinguished by whether the objective function contains the loss function of style discretization or not, among which 8 groups of models do not contain the loss function of style discretization, and the other 8 groups contain it. 8 FID indexes, PSNR indexes and SSIM indexes of the four migration effects of the eight models when they have no style discretization are shown in Table 1. From the table, it can be seen that the model in this paper has the best image quality on the four style migrations, and the FID is between 100 and 200, which are lower than the other models.

Table 1: The four kinds of migration effects of 8 models without style

		g0d0	g1d0	g2d0	g3d0	g0d1	g1d1	g2d1	This method
Traditionalist style	FID	263.833	272.855	243.968	269.731	244.048	303.715	233.259	180.939
Modernism		215.044	185.245	208.348	218.033	237.535	272.377	229.656	176.089
Minimalist style		263.585	300.041	256.404	224.844	208.548	278.328	162.22	125.245
Industrialist style		287.346	276.02	281.098	287.249	243.731	254.532	199.763	174.017
Traditionalist style	PSNR	21.264	23.29	19.636	24.346	21.947	19.883	23.679	26.428
Modernism		15.033	21.946	18.682	18.634	22.919	13.115	23.456	26.98
Minimalist style		19.026	22.05	21.003	24.465	18.223	23.944	16.029	28.365
Industrialist style		17.53	20.025	20.218	24.234	27.496	18.934	22.204	22.681
Traditionalist style	SSIM	0.68	0.667	0.686	0.67	0.696	0.675	0.732	0.769
Modernism		0.579	0.614	0.609	0.595	0.643	0.598	0.666	0.746
Minimalist style		0.661	0.642	0.714	0.682	0.68	0.621	0.728	0.731
Industrialist style		0.541	0.509	0.619	0.51	0.623	0.587	0.619	0.686

The FID metrics, PSNR metrics and SSIM metrics of the four migration effects when the eight models have style discretization are shown in Table 2. From the table, the FID, PSNR, and SSIM scores of the g2d1 model are partly better than all other models and partly lower than this paper's model, but the difference between the two scores is not significant, and overall, the evaluation indexes of this paper's model are better than those of other models.

Table 2: The eight models have four kinds of migration effects in the discretization

		g0d0	g1d0	g2d0	g3d0	g0d1	g1d1	g2d1	This method
Traditionalist style	FID	249.793	295.318	226.294	272.297	235.528	345.169	217.72	204.56
Modernism		208.087	254.422	194.8	229.14	198.698	283.016	181.54	180.613
Minimalist style		264.102	298.837	246.171	285.952	224.459	328.926	196.929	203.369
Industrialist style		248.61	295.865	240.933	272.787	219.957	321.413	208.699	215.116
Traditionalist style	PSNR	24.569	20.892	22.623	22.72	22.599	20.202	23.549	25.882
Modernism		22.358	21.571	23.618	22.067	20.167	20.527	21.564	24.325
Minimalist style		20.55	22.691	23.405	20.371	20.993	22.22	19.309	23.533
Industrialist style		21.225	21.51	20.794	21.495	21.63	21.221	22.89	22.898
Traditionalist style	SSIM	0.693	0.696	0.745	0.655	0.675	0.631	0.771	0.787
Modernism		0.601	0.603	0.658	0.644	0.641	0.525	0.711	0.652
Minimalist style		0.659	0.669	0.724	0.658	0.713	0.641	0.763	0.751
Industrialist style		0.548	0.476	0.562	0.568	0.596	0.546	0.644	0.66

(2) Experimental training process loss

The variation of generator loss values for each model training process is shown in Fig. 5. Where the horizontal coordinate is the number of model training rounds, and the vertical coordinate is the quantized loss value. The loss value of each loss decreases steadily, the model are in stable training, compared with the loss value of other models, the loss value of the model in this paper is significantly smaller than most other models, the final loss value of the generator is in the range of 10 units.

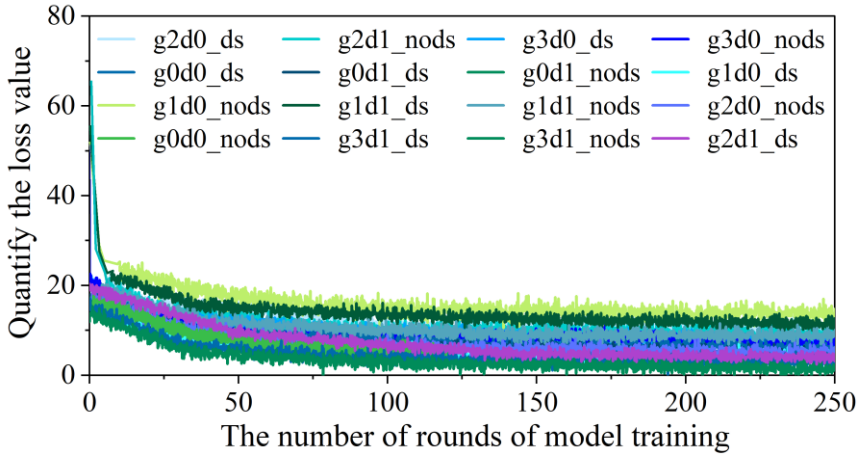


Figure 5: The variation of generator loss values during the training process

3.1.3 Comparison experiments

The evaluation results of CycleGAN model and this paper's method on the three indicators are shown in Table 3, where the smaller FID is the better, and the larger value of the remaining indicators is the better. From the table, it can be seen that the scores of CycleGAN on the three indexes are worse than that of this paper's method, and the FID index decreases from an average of 265.698 on CycleGAN to an average of 181.256, and the optimization rate reaches 31.781%. Overall the indexes of this thesis method are better than CycleGAN.

Table 3: The cyclegan model and the results of this method are evaluated

Indicators	Algorithm	Painting	Anime	Sketch	Vangogh	Mean
FID	CycleGAN	228.71	258.313	275.37	300.399	265.698
	This method	176.112	182.676	172.173	194.061	181.256
SSIM	CycleGAN	0.697	0.721	0.768	0.806	0.748
	This method	0.755	0.803	0.799	0.838	0.799
PSNR	CycleGAN	22.248	21.873	20.69	20.347	21.290
	This method	25.507	26.418	23.797	25.657	25.345

3.2 Case Studies

3.2.1 Study area

The study area selected for this paper is the historical style neighborhood in area A, which is a narrow strip of land, about 5km long, covering an area of about 225hm², and the Hi-SWGAN model is used to design the environment of this area and explore the impact of this method on spatial aesthetics.

3.2.2 Dimensions of evaluation indicators

The evaluation index of space aesthetics contains the following dimensions:

(1) Space form. Space is formed by the combination of various buildings as well as the enclosed space, and the spatial form is a comprehensive reflection of the buildings in terms of volume, modeling, combination mode, proportion, scale, etc., which is the main influencing factor of the visual quality of space.

(2) Interface quality. The interface quality of space is mainly reflected through the building façade, which can deeply portray rich spatial details through various means, including various building structural elements, surface materials and texture, decorative details, etc., and can reflect various architectural styles.

(3) Overall coordination. Each building should be regarded as part of a larger whole, forming a larger group with other surrounding buildings. Overall coherence is an important factor affecting the visual quality of the space.

(4) Historical context. Space is closely related to social life. Therefore, important historical periods, events, activities and related people will give space and buildings with special historical imprints, and these historical and cultural attributes are also important aspects that affect people's evaluation of the visual quality of space.

(5) Cultural characteristics. The built environment is affected by local culture, customs, climate, geography and other aspects, some unique architectural forms and details are reflected through continuous inheritance and evolution, which are important features reflecting regional and local culture.

3.2.3 Evaluation indicator weights

The judgment matrix and weights of the evaluation indicators are shown in Table 4, the judgment matrix has satisfactory consistency, the largest characteristic root λ_{max} =5.078, CR=0.021<0, so the same method can be used to obtain the weights of other factor layers. The weight of overall coordination is 0.381, which is more than 1/3 of the overall weight and occupies the highest proportion. Spatial form and cultural features have the second highest weight. The weights of historical background and building facade are relatively low. Therefore, it can be inferred that people are most concerned about the overall unity and consistency of the urban space composed of various spatial elements, and also show some importance to the form

and regional characteristics of buildings.

Table 4: The judgment matrix and weights of the evaluation indicators

Evaluation index	Historical background	Overall coordination	Cultural Characteristics	Interface quality	Spatial form	Indicator weight
Spatial form	1	0.361	0.424	1.405	0.903	0.108
Overall coordination	2.626	1	2.174	3.36	2.855	0.381
Cultural Characteristics	1.947	0.491	1	0.983	0.944	0.172
Interface quality	0.807	0.277	0.929	1	0.487	0.143
Spatial form	1.135	0.37	1.172	1.852	1	0.196

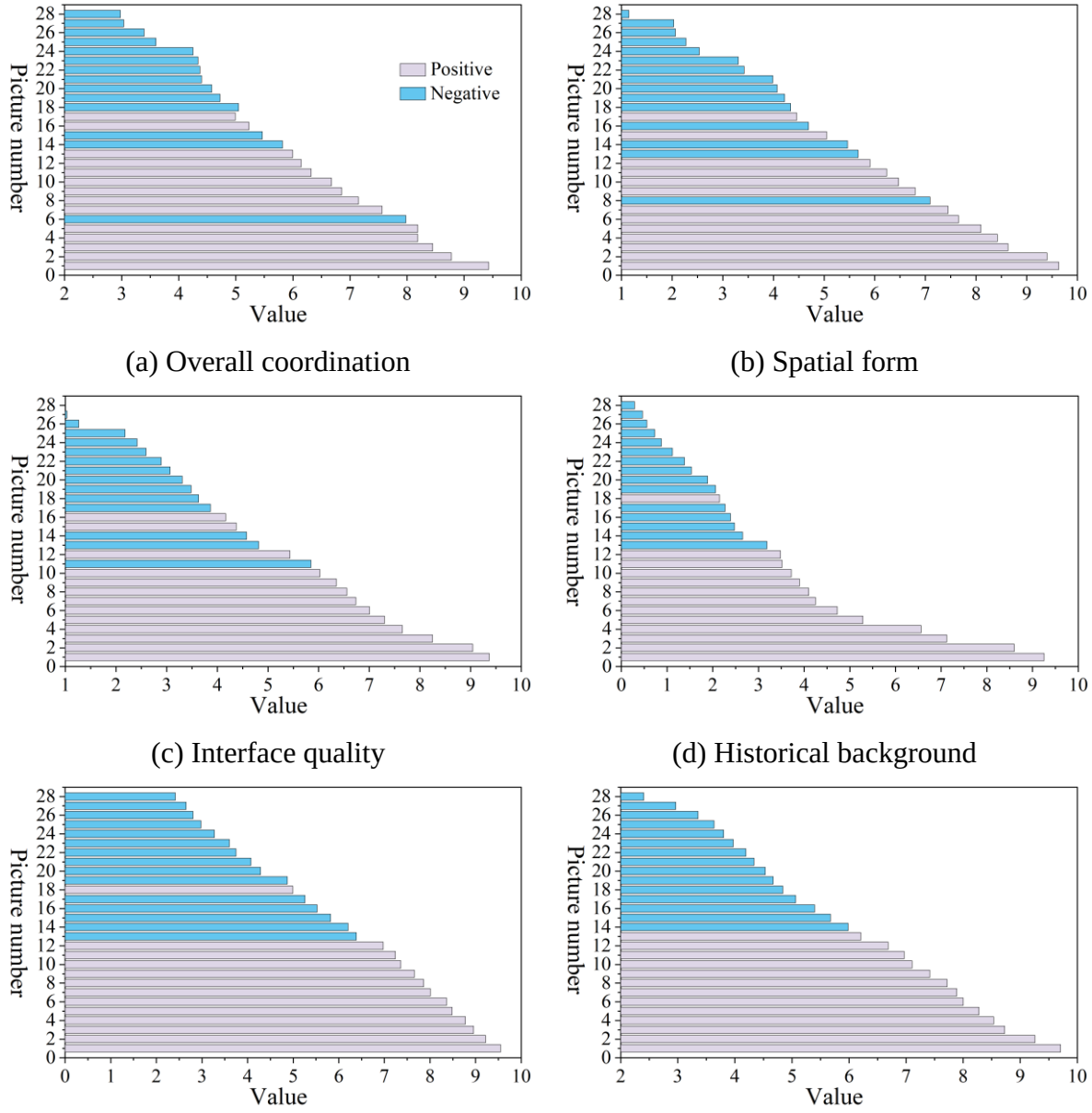
3.2.4 Comprehensive analysis of space aesthetics

In the comprehensive evaluation stage, the investigators assessed the indicators of the five dimensions in the experimental scenarios, and the measurement tool was a 10-level Likert scale, calculated the average value of each dimension, and utilized the weights of the indicators for the calculation of the weighted composite scores, and the trend of the scores of the various spatial scenarios is shown in Table 5.

Table 5: Score trends of each spatial scene

Number	Space form	Overall coordination	Interface quality	Historical background	Cultural characteristics	Spatial aesthetics score
1	6.22	5.54	7.84	8.28	3.71	6.03
2	2.13	3.03	2	0.64	1.47	1.75
3	1.12	2.73	0.78	1.04	1.46	1.7
4	1.81	3.38	0.85	1.91	1.89	2.16
5	2.09	3.9	1.61	3.17	4.04	2.99
6	1.54	4.12	0.82	0.78	2.07	2.87
7	3.98	5.36	3.64	1.75	1.91	3.66
8	5.24	6.23	5.21	2.49	2.36	4.78
9	4.71	5.38	6.22	2.18	2.95	4.35
10	2.97	6.21	2.27	2.52	2.39	3.76
11	6.42	4.55	7.56	7.4	3.66	5.43
12	4.02	4.34	3.67	1.97	2.67	3.76
13	5.99	6.59	5.82	1.98	2.32	5.46
14	3.56	4.98	4.14	4.23	3.61	4.2
15	5.59	4.81	6.03	5.49	4.24	4.38
16	3.27	4.2	3.26	3.07	3.35	3.87
17	3.07	3.13	2.71	3.23	3.87	3.14
18	3	3.72	3.61	2.93	2.95	3.26
19	2.16	3.27	2.12	1.35	1.18	2.24
20	4.62	5.81	5.98	4.12	3.87	4.43
21	2.87	3.91	3.44	1.47	2.71	3.05
22	5.86	6.39	6.94	6.03	4.24	6.35
23	2.66	3.62	3.13	0.89	1.73	2.96
24	2.54	3.58	2.42	0.72	1.82	2.37
25	5.15	6.14	5.72	3.64	3.28	4.91
26	4.8	6.91	5.69	3.84	3.38	5.83
27	4.94	4.83	5.33	3.93	3.47	4.46
28	3.42	3.6	4.31	1.57	1.07	2.63

The evaluation indexes of “positive” and “negative” scenarios are shown in Figure 6, and Figures a to f represent the overall coordination, spatial form, interface quality, historical background, cultural characteristics and spatial aesthetics scores, respectively. The results show that the indicators reflecting the objective material form, such as overall coordination, spatial form, and architectural façade, have strong differentiation boundaries, i.e., the “positive” scenarios have higher scores in these dimensions, while the “negative” scenarios have lower scores. As for historical background and cultural characteristics, which are two indicators reflecting spiritual connotation, there is no obvious boundary of differentiation between the two types of scenes, i.e., in these two dimensions, the “positive” scenes may also score very low, while the “negative” scenes may score high.



(a) Overall coordination (b) Spatial form

(c) Interface quality (d) Historical background

(e) Cultural characteristics (f) Spatial aesthetic score

Figure 6: The differentiation of evaluation indicators

The mean values of the spatial aesthetic evaluation indexes of “positive” scenes and “negative” scenes are analyzed as shown in Table 6. The results of the test show that the scores

of “positive” scenes are higher than those of “negative” scenes for these five dimensions, and there is a significant difference between the two groups of samples at the 95% confidence level ($p<0.05$).

Table 6: Mean analysis of spatial aesthetic evaluation indicators

Evaluation index	Scene type	Number of cases	Mean	Standard deviation	Homogeneity of variance test		Independent sample t-test	
					F value	Significance	T value	Significance
Overall coordination	Positive	14	7.142	0.8	0.322	0.579	6.593	0.000
	Negative	14	4.572	0.561				
Spatial form	Positive	13	7.248	1.05	2.612	0.121	8.365	0.000
	Negative	15	3.756	0.86				
Interface quality	Positive	13	6.791	1.362	0.032	0.885	6.632	0.000
	Negative	15	3.050	1.157				
Historical background	Positive	13	5.130	1.97	0.023	0.885	6.513	0.006
	Negative	15	1.593	1.121				
Cultural characteristics	Positive	13	7.959	0.808	2.065	0.172	2.136	0.049
	Negative	15	4.262	1.047				
Spatial aesthetic score	Positive	13	7.886	0.651	0.000	0.978	7.963	0.000
	Negative	15	4.324	0.648				

In summary, overall coordination is an important factor affecting the visual quality of spatial aesthetics, and most of the scenes with higher visual quality have a more regular architectural space form, the style and volume of each spatial element have a high degree of coordination and consistency, and there are generally clear landscape levels in the scenes to form a clear-cut, staggered layout form. Its spatial details are rich, full of changes, but also can achieve unity, and has a high quality of spatial enclosure and continuous interface. Secondly, people are highly sensitive to historical elements and regional cultural characteristics. It can be seen that originality is one of the most important values of historical buildings, which can arouse people's sense of regional identity and historical sentiment, and the space aesthetics in environmental design not only needs to be harmonized in appearance, but also needs to focus on the continuation of the unity of the connotation.

4 Conclusion

With the development of science and technology, image generation technology is increasingly used in environmental design. The study designs a hierarchical generative adversarial network model based on moving windows, verifies its effect in image style migration through model performance testing experiments, and explores the impact of the method on spatial aesthetics by taking a historical neighborhood as a research object. The article draws the following conclusions:

(1) In the image style generation effect experiment, compared with CycleGAN, the FID index of this model decreases from an average of 265.698 on CycleGAN to an average of 181.256, and the optimization rate reaches 31.781%. This verifies the superiority of the model in this paper.

(2) In the case study analysis of a historical district, it is found that the weight of overall coordination is the highest at 0.381, and the weight of spatial form and cultural features is the

second highest. Therefore, it is concluded that people are most concerned about the overall unity and coherence of the urban space composed of various spatial elements in environmental design.

(3) Through the comprehensive analysis of subjective evaluation of a historical neighborhood, it can be concluded that the overall coordination of AI generation technology in environmental design is an important factor affecting the visual quality of spatial aesthetics.

About the Author

Dong Wang (b. February 1983), male, Han ethnicity, native of Xinxiang, Henan Province. He holds a Master's degree and the title of Associate Professor. Currently, he serves as the Vice Dean of the School of Art and Design, Henan Institute of Technology. His main research direction is art design.

References

[1] Xu, H., Omitaomu, F., Sabri, S., Zlatanova, S., Li, X., & Song, Y. (2024). Leveraging generative AI for urban digital twins: a scoping review on the autonomous generation of urban data, scenarios, designs, and 3D city models for smart city advancement. Urban Informatics, 3(1), 29.

[2] Rao, J., & Xiong, M. (2024). Research on Environmental Architectural Design Methods Based on the AIGC Creation Method. In International Symposium on World Ecological Design (pp. 1036-1046). IOS Press.

[3] Lee, H. (2025). Designing and evaluating landscape proposals with generative AI ChatGPT-urban plaza design via landscape firm style emulation with educational potential. Journal of the Korean Institute of Landscape Architecture, 53(2), 116-132.

[4] Liu, R., & Ismail, A. I. (2025). AI Innovation in Architectural Design: Enhancing Aesthetic Experience with'Midjourney'. Journal of Advanced Research Design, 127(1), 84-95.

[5] Didichenko, M., Bulakh, I., & Kozakova, O. (2019). Spatial and Temporal Principles and Methods of the Historical Urban Environment Composition Transformations. Urban and Regional Planning, 4(4), 144-151.

[6] Lou, Y. (2019). The idea of environmental design revisited. Design Issues, 35(1), 23-35.

[7] Losonczi, A. N. N. A., Halász, B. Á. L. I. N. T., Keszei, B. A. R. B. A. R. A., Dobszai, D. Á. N. I. E. L., & Dúll, A. N. D. R. E. A. (2017). Investigating changes in space usage preferences by changing aesthetic environmental variables. In Proceedings of the 11th Space Syntax Symposium. Instituto Superior Técnico, Departamento de Engenharia Civil, Arquitetura e Georrecursos, Portugal. pp (pp. 131-1).

[8] Phd, Y. L., & Wang, W. (2024). Emotional Spaces in Environmental Design: The Integration of Aesthetics, Function, and Emotion. Cultura: International Journal of Philosophy of Culture and Axiology, 21(5).

[9] Saleh, H. A., & Alrobaee, T. R. (2024). Planning and Designing Sustainable Urban Space According to the Concept of Ecological Aesthetics. International Journal of Sustainable Development & Planning, 19(11).

[10] Hou, D. (2024). Conditional generative adversarial networks for image-based sunlight analysis of residential blocks. Energy and Buildings, 316, 114295.

[11] He, T. (2025, May). AIGC-Enabled Light and Shadow Narrative Strategies for Urban Cultural and Tourism Night Tours. In International Conference on Human-Computer Interaction (pp. 49-61). Cham: Springer Nature Switzerland.

[12] Miao, J., Zhou, Z., & Meng, F. (2025, July). Exploring Eco-Narrative Interaction through AIGC: The Creative Journey of “Plast-ocean”. In Companion Publication of the 2025 ACM Designing Interactive Systems Conference (pp. 242-245).

[13] Chen, R., Zhao, J., Yao, X., Jiang, S., He, Y., Bao, B., ... & Wang, C. (2023). Generative design of outdoor green spaces based on generative adversarial networks. Buildings, 13(4), 1083.

[14] Tang, X., & Chung, W. J. (2024). Automated urban landscape design: an AI-driven model for emotion-based layout generation and appraisal. PeerJ Computer Science, 10, e2426.

[15] Zhuang, Y., Kang, Y., Fei, T., Bian, M., & Du, Y. (2024). From hearing to seeing: Linking auditory and visual place perceptions with soundscape-to-image generative artificial intelligence. Computers, Environment and Urban Systems, 110, 102122.

[16] Softaoglu, H. (2025). Reimagining Architecture: A Semiotic Study of Sound in Ai-Generated Spatial Design. Smart Design Policies Journal, 2(1), 107-121.

[17] Gallega, R. W., & Sumi, Y. (2024). Exploring the use of generative AI for material texturing in 3D interior design spaces. Frontiers in Computer Science, 6, 1493937.