Using GLO-GAN for Realistic Lighting and Shadow Generation in Urban Art Visualization

Yue Cheng
[1]Department of Design Art, Taiyuan Institute of Technology, Taiyuan 030008, Shanxi, China
Corresponding author: Yue Cheng (e-mail: cy112011cy@163.com)

ABSTRACT To address the difficulty in balancing the physical realism of light and shadow with the representation of region-specific chromatic distribution, surface ornament regularity, façade composition patterns, and material reflectance characteristics in urban art visualization, this paper proposes a Generative Latent Optimization–Generative Adversarial Network (GLO-GAN) architecture driven by physical constraints and cultural perception. The framework refines the GLO-GAN architecture by incorporating material-aware attributes to enhance the consistency between synthesized urban scenes and their cultural context. A dual-branch generator with a shared GLO-optimized latent space is employed to jointly model physical illumination and semantic characteristics. Differentiable physical lighting constraints, including sun trajectories, ambient occlusion, and material reflectance, are integrated to ensure physically plausible rendering, while urban cultural labels provide semantic guidance for region-aware feature generation. Joint optimization combining L2 reconstruction, VGG perceptual, adversarial, and lighting consistency losses enables coordinated feature learning. A coarse-to-fine progressive upsampling strategy together with a multi-scale PatchGAN discriminator suppresses artifacts and enhances local details, whereas the StyleGAN3 backbone guarantees alias-free feature propagation and stable spatial phase alignment during illumination modulation. Considering that physically consistent light–material interaction is also fundamental to electromagnetic wave propagation analysis and computational imaging, the proposed framework provides methodological insights for intelligent electromagnetic sensing and multimodal scene perception. Experimental results demonstrate high performance, achieving a solar azimuth RMSE of 4.2°, cultural style classification accuracy of 92.7%, and an average SSIM of 0.863. The proposed method enables realistic urban art visualization while preserving stable chromatic tendencies, structural motifs, and material-associated tonal distributions, and further offers a practical computational paradigm for integrating heterogeneous material information into intelligent electromagnetic-aware visualization systems.

INDEX TERMS GLO-GAN, physical lighting constraints, computational imaging, electromagnetic sensing, multimodal scene perception

I. INTRODUCTION

WITH the increasing demand for visual expression in urban art visualization, realistic lighting and shadow renderings play a key role in project presentation, public communication, and decision support. This necessity for high-fidelity visualization extends directly to integrated material design, where the industry's complex interplay of light, color, and texture offers crucial data, ensuring that digital models accurately reflect the material quality and environmental context of proposed installations [1,2]. Traditional rendering methods rely on high-precision modeling and manual parameter adjustment, resulting in low efficiency and difficulty in rapidly responding to visualization updates [3,4]. In recent years, generative adversarial networks (GANs) have made significant progress in image synthesis and are widely used for architectural and urban visual generation [5,6]. However, existing methods often focus on visual fidelity when generating light and shadow effects, ignoring the physical laws of illumination or the cultural semantic characteristics of urban space. The cultural semantic characteristics referenced in this study refer to region-associated chromatic profiles, recurring ornament geometries, façade structural proportions, and material surface reflectance tendencies that are observable and statistically stable within each annotated urban scene category. This leads to generated outputs that achieve visual similarity but lack semantic fidelity, making them unsuitable for professional design applications [7,8]. Therefore, how to simultaneously ensure

the deep coupling of physical authenticity and cultural expression in the generation process has become a core challenge that urgently needs to be overcome in the field of intelligent design generation. The light and shadow effects in urban environmental art design not only involve physical elements such as architectural form, material reflection, and sunlight trajectory, but also carry social attributes such as regional culture, historical features, and public aesthetics. Jaglarz A [9] pointed out the important role of color in shaping architecture and urban space. Lin Q [10] emphasized in the urban night scene evaluation system that the existing evaluation system focuses on objective illumination indicators, ignores environmental characteristics and subjective feelings, and lacks consideration of regional culture, economic benefits, management and maintenance. Cai P [11] pointed out that through environmental landscape design, a bridge of emotional communication can be built between the city and the viewer, which can not only show the rich local cultural characteristics, but also arouse the viewer's resonance and enhance the viewer's sense of pleasure and belonging. Liu C [12] proposed a multidimensional urban landscape design method based on nonlinear theory to solve the problem of large differences in multidimensional urban landscape design. It extracts urban landscape characteristics into quantifiable parameters to improve its quantitative analysis ability. Ruan W Q [13]'s research results showed that nighttime tourism landscapes are composed of spatial presentation, nighttime atmosphere, commodities, nighttime activities, crowd gathering, and cultural displays. Nighttime tourism landscapes directly stimulate tourists' sensory experiences and influence their emotional, functional, and flow experiences. In recent years, generative artificial intelligence has shown great potential in the field of environmental art design, and a variety of advanced models have been used to improve the intelligent level of design expression. Gan W [14] combined the GAN with a multi-objective optimization algorithm, and the similarity of the generated results was stably distributed between 0.7 and 0.8, which was able to effectively learn the design scheme of the preferred style. Lu-Yao W [15] proposed an environmental landscape art design method based on a visual neural network model, using the Swin Transformer text encoder to represent the landscape art requirements in rural construction and then feeding the resulting text feature vector into a GAN to generate the corresponding image. Bai K [16] used the neural radiation field method to enhance the three-dimensional reconstruction performance under low light conditions. Guo X [17] constructed an image style transfer model for landscape design, using a multi-scale discriminator, and introduced a content feature mapping module and a transfer learning method. Its peak signal-to-noise ratio and similarity evaluation indexes were 17.42 and 0.91, respectively, and the image distortion degree of the model was small. Gao D [18] proposed a hybrid generative architecture that integrated StyleGAN2 and a diffusion model, and combined it with a transfer learning strategy to optimize the generalization ability of the model in small sample scenarios, achieving high-fidelity image generation and feature independence control. Zhao W [19] proposed a multi-attention mechanism based on GAN for image cartoonization, which enabled the generated images to exhibit obvious cartoon-style features and effectively improved their visual quality. Zhao H [20] used a multimodal feature fusion method based on gated attention and enhanced attention to emphasize key information, suppress non-critical details, and enhance the interaction between modalities. Existing methods generally separate physical constraints and cultural perception and lack a unified modeling paradigm. This paper proposes a GLO-GAN method driven by physical constraints and cultural perception to generate high-fidelity realistic light and shadow effects in urban environment art design. First, a learnable latent vector is initialized based on the input design sketch, and the latent space is optimized end-to-end to improve structural fidelity. Second, a decoupled dual-path architecture is constructed in the generator. On the one hand, a differentiable physical lighting module is embedded. This module combines sun trajectory, material reflection, and depth normal information to explicitly model realistic lighting and shadows, ensuring the physical plausibility of light and shadow. The differentiability of this module ensures that the lighting error can be accurately optimized to convergence through backpropagation, providing an interpretable physical prior for the generation process. Furthermore, a cultural semantics-aware mechanism is introduced to map cultural labels into embedding vectors. Adaptive Instance Normalization (AdaIN) style modulation and a culturally guided attention mechanism are used to accurately express regional styles. Finally, a multi-scale PatchGAN discriminator is used for adversarial training, and an interactive lighting control interface is designed to support dynamic adjustment of time, weather, and light source type. This method innovatively integrates differentiable rendering and cultural embedding mechanisms into the generation process. The model foundation relies on StyleGAN3, whose alias-free design reduces spatial distortion in high-frequency regions and sustains consistent cultural and illumination responses. Extending this capability to the material domain, the integration of specific industry data allows the model to capture material-specific emotional nuances. Combined with the GLO latent optimization strategy, this ensures both visual and emotional resemblance while maintaining efficient generation, significantly enhancing the usability of realistic renderings in professional visualization tasks.

II. ALGORITHM DESIGN

A. OVERALL ARCHITECTURE DESIGN

The GLO-GAN architecture, driven by physical constraints and cultural perception, introduced in this paper, consists of three components: a generator, a discriminator, and a learnable latent space, as shown in Figure 1. The generator employs a decoupled dual-pathway structure, constructing separate physical illumination branches and cultural se-

mantic branches. The generator adopts StyleGAN3 as the backbone and uses a decoupled dual-pathway structure with physical illumination and cultural semantic branches. StyleGAN3 provides an alias-free feature transformation mechanism that stabilizes spatial correlations across scales and supports consistent modulation of structural and cultural attributes. The discriminator employs a multiscale PatchGAN, discriminating local image realism at three levels: 64 × 64, 128 × 128, and 256 × 256. The learnable latent vector corresponding to the input sketch is optimized end-to-end to ensure structural fidelity. This architecture enables the parallel infusion and coordinated generation of physical realism and cultural consistency, providing a high-fidelity, interpretable, and realistic light and shadow generation path for urban art visualization.

Figure 1 illustrates the overall pipeline of the GLO-GAN, a dual-driven approach driven by physical constraints and cultural perception: the input sketch is initialized into a learnable latent vector, which is then structurally encoded and fused with the physical illumination branch and the cultural semantic branch in a shared latent space to drive the progressive decoder to generate a high-resolution image. The generated results are fed into a multi-scale PatchGAN discriminator for authenticity verification. Simultaneously, the latent vector is back-updated via differentiable optimization to ensure structural fidelity and generation stability. The final output is a realistic light and shadow rendering of an urban environment art design that combines physical authenticity with cultural consistency.

B. LEARNABLE LATENT SPACE INITIALIZATION AND OPTIMIZATION

This paper employs a generative latent optimization strategy to construct a learnable latent space [21,22]. For each input sketch $s \in \mathbb{R}^{H\times W\times C}$, a separate 512-dimensional learnable latent vector $z_s \in \mathbb{R}^{512}$ is assigned, initialized to samples from a standard normal distribution $z_s^{(0)} \sim N(0, I)$. This latent vector is not generated by the encoder, but serves as an optimizable parameter during model training and is updated synchronously with the neural network weights. During training, the generator network parameters are first fixed, and the latent vector corresponding to each sketch is independently optimized. The optimization goal is to minimize the composite loss function between the generated image and the corresponding real-world lighting effect as follows:

$$\mathcal{L}_{\text{latent}}(z_s) = \lambda_{\text{pix}}\|G(z_s) - x\|_2^2 + \lambda_{\text{perc}} \sum_{l\in L} \frac{1}{C_l H_l W_l} \|\phi_l(G(z_s)) - \phi_l(x)\|_F^2 \tag{1}$$

The first term in formula (1) is the pixel-level L2 reconstruction loss, which ensures the geometric consistency of the generated image in color and structure. The second term is the perceptual loss. The VGG16 network pre-trained on ImageNet is used as a fixed feature extractor to extract the feature map of the l layer [23,24]. Here, C_l, H_l, and W_l denote the number of channels and spatial dimensions of the feature map at layer l, and $\|\cdot\|_F$ represents the Frobenius norm. The latent vector is iteratively updated by gradient descent as follows:

$$z_s^{(t+1)} = z_s^{(t)} - \eta \nabla_{z_s} \mathcal{L}_{\text{latent}}(z_s^{(t)}) \tag{2}$$

In formula (2), η is the learning rate. The converged z_s^* latent code is channel-concatenated with the cultural semantic embedding vector e_c to form a joint latent representation:

$$z_{\text{fused}} = [z_s^*, e_c] \in \mathbb{R}^{576} \tag{3}$$

This latent space optimization mechanism avoids the information bottleneck in traditional encoder-decoder architectures, improving the fidelity of subtle geometric features in urban design sketches and providing a high-fidelity structural foundation for the subsequent precise injection of physical and cultural constraints. A trainable mapping module transforms the semantic embedding vector before it enters the joint latent space. The mapping module uses a two-layer linear transformation with activation to align the statistical distribution of the semantic vector with the StyleGAN3 latent domain. The transformed semantic vector is then fused with the latent code through channel-wise integration. This process stabilizes the interaction between structural information from the latent code and the semantic modulation signals, preventing distributional mismatch inside the generator. The loss function weight configuration is shown in Table 1.

TABLE 1. Loss Function Weight Configuration

Loss Term	Symbol	Weight Value
Pixel Reconstruction Loss	λ_{pix}	1
Perceptual Loss	λ_{perc}	0.1
Adversarial Loss	λ_{adv}	1
Lighting Consistency Loss	λ_{light}	0.5

C. IMPLEMENTATION OF THE PHYSICAL LIGHTING CONSTRAINT MODULE

This paper embeds a differentiable physical lighting calculation module in the generator's mid-level feature space to explicitly model the daylighting process [25,26]. This module takes the city's geographic location and the specified time of the design scenario as input. The solar declination angle δ is determined by the following formula:

$$\delta = 0.4093 \sin\left(\frac{2\pi}{365}(D - 81)\right) \tag{4}$$

The solar hour angle ω is:

$$\omega = \frac{\pi}{12}(t + \Delta t - 12), \quad \Delta t = 4(\lambda - \lambda_{\text{std}}) + E \tag{5}$$

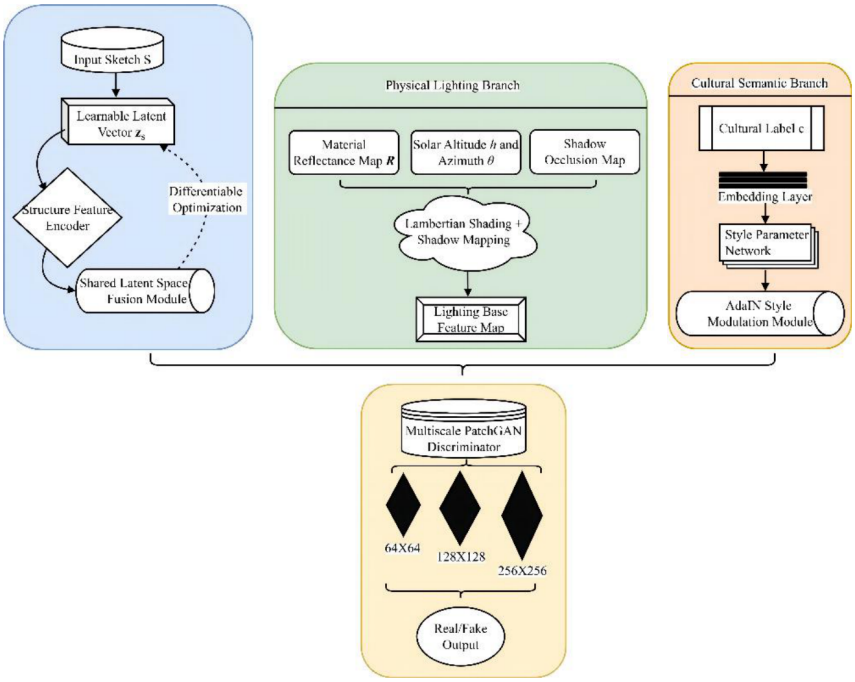

FIGURE 1. Overall Framework

In formula (5), Δt denotes the mean time difference, E is the equation-of-time correction term, and λ_{std} is the central meridian of the time zone. The solar altitude angle h and azimuth angle θ are:

$$\begin{aligned} \sin h &= \sin\phi \sin\delta + \cos\phi \cos\delta \cos\omega, \\ \cos\theta &= \frac{\sin\delta \cos\phi - \cos\delta \sin\phi \cos\omega}{\cos h}. \end{aligned} \tag{7}$$

This paper constructs the normalized incident light direction vector $\vec{d} = [\sin\theta \cos h, \cos\theta \cos h, \sin h]^T \in \mathbb{R}^3$. Based on the scene depth map and surface normal map obtained by parsing the input sketch, $\vec{n} = [n_x, n_y, n_z]/\|\vec{n}\|$ is normalized, and a shadow map $S \in [0,1]^{H\times W}$ is generated using differentiable soft shadow mapping. Light step sampling is performed along the $\vec{d}$ direction [27,28] to calculate the occlusion probability:

$$S(i,j) = \sigma\left(\alpha \sum_{k=1}^{K} \max(0, z_k(i,j) - D(i,j))\right) \tag{8}$$

In formula (8), z_k is the depth of the kth sampling point along the ray; σ is the Sigmoid function; α is the softening coefficient. The material reflectance map $R \in [0,1]^{H\times W\times 3}$ is obtained by looking up the table of the sketch semantic segmentation results, and the diffuse reflection intensity is calculated according to Lambert's cosine law:

$$I_{\text{diff}}(i,j) = R(i,j) \cdot \max(0, \vec{n}(i,j) \cdot \vec{d}) \cdot S(i,j) \tag{9}$$

The final physical lighting field is:

$$I_{\text{phys}} = I_{\text{diff}} + I_{\text{amb}} \in \mathbb{R}^{H\times W\times 3} \tag{10}$$

The illumination field is transformed into a feature dimension d through a 1×1 convolution and used as a guiding feature to perform channel splicing at the input of the 4th to 6th upsampling modules of the generator to form the illumination condition constraint:

$$F_{\text{gen}}^{(l)} = \text{Concat}\left(F_{\text{prev}}^{(l)}, \text{Upsample}(I_{\text{phys}}^{\text{feat}})\right), \quad l \in \{4,5,6\} \tag{11}$$

Through the differentiable physical layer, backpropagation can transfer illumination errors to the latent space and generator parameters, forcing the generated results to follow geometric and optical laws and avoid light and shadow misalignment or intensity anomalies that violate physical common sense.

D. CONSTRUCTION OF CULTURAL SEMANTIC PERCEPTION GUIDANCE MECHANISM

This paper constructs a cultural semantic perception guidance mechanism that transforms discrete cultural category information into a differentiable continuous style control signal and embeds it into the deep feature modulation process of the generator [29]. The structured semantic vector generated from TF-IDF and LDA is compressed to dimension 64 through a linear projection layer. This dimension is selected after evaluating the reconstruction stability of the semantic distribution and the modulation response in the generator. Lower dimensions reduce the separability of topic components, and higher dimensions introduce redundant variance that weakens the consistency of modulation during upsampling. The projection output is used as the unified semantic representation before style parameter mapping. A set of $K = 10$ typical urban cultural

labels $C = \{c_1, c_2, \ldots, c_K\}$ is predefined, and each label c_K is mapped to a learnable 64-dimensional embedding vector $e_k \in \mathbb{R}^{64}$. The compressed semantic vector and the cultural embedding share the same dimensional scale to maintain alignment in the latent intervention process, and their joint distribution determines the final modulation state of the generator, forming a cultural embedding matrix $E \in \mathbb{R}^{K\times 64}$. This embedding vector is mapped to a set of style parameters $f_{\text{style}} : \mathbb{R}^{64} \to \mathbb{R}^{2d}$ through a multi-layer fully connected network $[\gamma, \beta] = f_{\text{style}}(e_k) = \text{MLP}(e_k)$, $\gamma, \beta \in \mathbb{R}^d$. d is the channel dimension of the generator's layer 1 feature map. γ and β are used as scaling and offset parameters, respectively, and injected into the AdaIN module [30]. The mapped semantic vector also passes through a latent-space alignment layer that adjusts its statistical scale and correlation structure before interacting with the StyleGAN3 latent code. This alignment preserves the stability of feature modulation during upsampling. The cultural label assigned to each scene corresponds to a set of measurable attributes extracted from annotated datasets. Each label associates with a stable chromatic distribution derived from the normalized hue histogram of representative façades, a texture periodicity index computed from the dominant frequency of surface ornament gradients, a façade composition ratio obtained from the statistical proportion between vertical and horizontal structural segments, and a material reflectance descriptor derived from calibrated photographs. These attributes are converted into a continuous vector through a linear transformation layer and are used as the initial state of the cultural embedding. For the generator's layer 1 input feature map $X^{(l)} \in \mathbb{R}^{H\times W\times d}$, the feature response $X_i^{(l)} \in \mathbb{R}^{H\times W}$ of its channel i is instance-normalized and fused with the cultural style parameters:

$$\text{AdaIN}(X_i^{(l)}; \gamma_i, \beta_i) = \gamma_i \cdot \frac{X_i^{(l)} - \mu(X_i^{(l)})}{\sigma(X_i^{(l)})} + \beta_i \tag{12}$$

In formula (12), $\mu(\cdot)$ and $\sigma(\cdot)$ represent the mean and standard deviation within the channel, respectively:

$$\mu(X_i^{(l)}) = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} X_i^{(l)}(h, w) \tag{13}$$

$$\sigma(X_i^{(l)}) = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\left(X_i^{(l)}(h, w) - \mu\right)^2 + \epsilon} \tag{14}$$

This paper introduces a culture-aware attention mechanism at the high level of the decoder. It is assumed that the high-level feature map $F \in \mathbb{R}^{H'\times W'\times d'}$ is used to generate the query matrix $Q = W_q F$, the key matrix $K = W_k F$, and the value matrix $V = W_v F, W_q, W_k, W_v \in \mathbb{R}^{d''\times d'}$ through 1×1 convolution. Combined with the cultural embedding vector, the attention bias term is generated:

$$A = \text{Softmax}\left(\frac{QK^T}{\sqrt{d''}} + M_c\right), \quad M_c = e_k e_k^T \otimes J \in \mathbb{R}^{H'W'\times H'W'}. \tag{15}$$

In formula (15), Mc is the culturally guided attention prior matrix; J is the all-one matrix; $\otimes$ represents the Kronecker product, which is used to expand the embedding vector into the spatial attention weight. The final attention output is:

$$F_{\text{att}} = AV \tag{16}$$

Through this mechanism, the model actively enhances the responses of structural regions associated with designated cultural categories during the generation process, achieving a unified expression of both form and spirit.

E. MULTI-SCALE GENERATOR AND DISCRIMINATOR DESIGN

This paper designs a progressive multi-scale generator and discriminator architecture that supports a stable mapping from a low-dimensional latent space to high-resolution images. The generator adopts a coarse-to-fine upsampling structure, starting from a constant feature tensor of $4 \times 4 \times 512$, and then sequentially upsampling to 512×512 resolution through six levels of upsampling modules, as shown in Figure 2. Each upsampling level consists of an upsampling layer, a style convolution block, and pixel normalization. The upsampling operation uses bilinear interpolation:

$$F_{\text{up}} = \text{BilinearUp}(F_{\text{in}}) \tag{17}$$

The style convolution block maps the input feature $F_{\text{in}} \in \mathbb{R}^{H\times W\times c_{\text{in}}}$ to the intermediate number of channels through a 1×1 convolution, then injects the style parameters γ and β generated by the latent vector and cultural embedding through the AdaIN layer, and finally outputs it through a 3×3 convolution. All feature maps are pixel-normalized before nonlinear activation:

$$F_{\text{norm}}(i, j, k) = \frac{F(i, j, k)}{\sqrt{\frac{1}{C}\sum_{k=1}^{C} F(i, j, k)^2 + \epsilon}} \tag{18}$$

The discriminator uses a multi-scale PatchGAN structure to discriminate the local authenticity of the image in parallel at three spatial scales [31,32]. The input image is downsampled to obtain multi-scale versions and fed into three weighted discriminant sub-networks. Each sub-network consists of four convolutional layers and ultimately outputs an Ns×Ns-dimensional true/false probability map. Adversarial training uses a least squares generative adversarial loss to improve training stability [33,34]. The total discriminator loss is:

$$\mathcal{L}_D = \frac{1}{2}\sum_{s=1}^{3}\mathbb{E}_{x\sim p_{\text{data}}}\left[(D_s(x_s) - 1)^2\right] + \mathbb{E}_{z\sim p_z}\left[(D_s(G(z))_s)^2\right] \tag{19}$$

The generator adversarial loss is:

$$\mathcal{L}_G^{\text{adv}} = \frac{1}{2}\sum_{s=1}^{3}\mathbb{E}_{z\sim p_z}\left[(D_s(G(z))_s - 1)^2\right] \tag{20}$$

Lighting consistency loss Llight is introduced to enforce physical constraints:

$$\mathcal{L}_{\text{light}} = \lambda_{\text{light}} \cdot \|G(z) \odot M_{\text{illu}} - I_{\text{phys}}\|_1 \tag{21}$$

In formula (21), M_{illu} is the mask of the light-sensitive area; $\odot$ is the element-by-element product; $\lambda_{\text{light}} = 0.5$. The final optimization goal of the generator is:

$$\mathcal{L}_G = \mathcal{L}_G^{\text{adv}} + \lambda_{\text{perc}}\mathcal{L}_{\text{perc}} + \lambda_{\text{recon}}\mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{light}} \tag{22}$$

This multi-scale adversarial architecture effectively improves the ability to realistically model local light and shadow transitions, material boundaries, and cultural details in complex urban scenes.

Figure 2 illustrates the network structure of the proposed progressive generator: starting from a constant 4 × 4 × 512 feature tensor, it is gradually upsampled to a resolution of 512 × 512 through six levels of upsampling and style convolution blocks. Each level includes bilinear upsampling, 1 × 1 convolution, AdaIN style modulation, and 3 × 3 convolution. Pixel normalization is performed before nonlinear activation to ensure training stability. This structure uses a coarse-to-fine generation mechanism to achieve high-fidelity, style-controlled generation of realistic urban environment lighting effects while maintaining the integrity of the geometric structure. StyleGAN3 maintains continuous signal behavior in the upsampling stages, which prevents feature drift and stabilizes the geometric and chromatic distributions required for illumination constraints.

F. IMPLEMENTATION OF AN INTERACTIVE LIGHTING CONTROL INTERFACE

This paper designs an adjustable lighting parameter input interface that supports dynamic adjustment of lighting conditions during the inference phase, achieving a real-time rendering response that integrates design, lighting, shadow and culture factors [35]. The interface provides three adjustable slider controls on the front end. All parameters are transmitted to the back-end generation engine in real time via Hypertext Transfer Protocol (HTTP). Time input is linearly interpolated to obtain continuous hourly values:

$$t_{\text{cont}} = t + \frac{\text{minute}}{60}, \quad t_{\text{cont}} \in \mathbb{R} \tag{23}$$

Weather type is mapped to atmospheric transmittance and ambient light enhancement coefficient, which control direct light attenuation and diffuse light intensity, respectively. The direct light item is scaled as follows:

$$I_{\text{diff}}^{\text{att}} = \tau \cdot I_{\text{diff}} \tag{24}$$

For light source type control, when natural light is used, only the sun directional light and sky diffuse light are enabled. When artificial light is used, the natural light source is turned off; the preset city light layout is activated; a warm light source is superimposed:

$$I_{\text{art}} = I_{\text{art}} \odot R \cdot c_{\text{warm}} \tag{25}$$

When mixing lights, a time-dependent transition function is introduced:

$$\begin{aligned} I_{\text{total}} &= \alpha(t) \cdot I_{\text{phys}}^{\text{att}} + (1 - \alpha(t)) \cdot I_{\text{art}}, \\ \alpha(t) &= \frac{1}{1 + \exp(k(t - t_0))}. \end{aligned} \tag{26}$$

In formula (26), $\alpha(t)$ is an S-shaped blending weight function that controls the smooth transition from natural light to artificial light, t0 is the transition midpoint, and k controls the transition steepness. The final illumination field is input into the physical illumination constraint module, which generates corresponding illumination guidance features in real time. This is then integrated with the cultural semantic branch to drive the generator to output realistic renderings under the corresponding conditions.

III. EXPERIMENTATION AND VERIFICATION

A. EXPERIMENTAL DESIGN

This paper constructs a dataset of environmental art design renderings covering four typical cities: Beijing, Suzhou, Guangzhou, and Chengdu. This dataset is based on Computer-Aided Design (CAD) drawings of typical urban landscapes, rendered using standardized visualization workflows supported by SketchUp and Lumion software. Annotations include solar azimuth and altitude, shadow boundary masks, material classification maps, and cultural category labels. The data is split into training, validation, and test sets at an 8:1:1 ratio. Training utilizes random cropping, horizontal flipping, and light and color dithering. The experimental setup is shown in Table 2.

TABLE 2. Experimental Setup

Item	Configuration
Cities	Beijing, Suzhou, Guangzhou, Chengdu
Data Set	Training Set (80%), Validation Set (10%), Test Set (10%)
Image Resolution	512 × 512 (cropped for training)
Scene Types	Daytime, Nighttime, Twilight/Dawn
Training Platform	NVIDIA A100 Graphics Processing Unit (GPU)
Batch Size	16
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
Learning Rate	2×10^{-4}
Training Steps	100,000 iterations
Latent Vector Dimension	512-dimensional
Cultural Embedding Dimension	64-dimensional

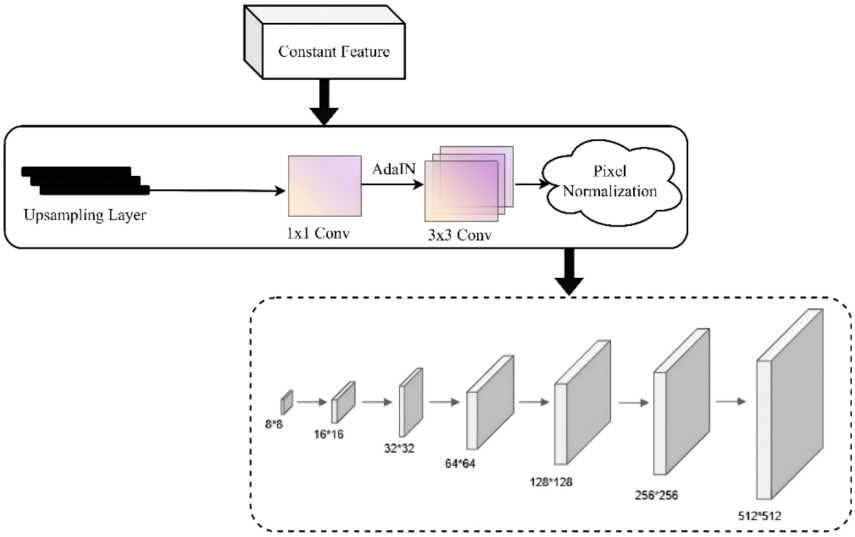

FIGURE 2. Generator Network Structure

B. PHYSICAL CONSISTENCY EVALUATION OF ILLUMINATION

To evaluate the physical consistency of illumination in generated images, this paper employs three quantitative metrics. First, the Hough transform is used to extract the principal illumination direction between the generated and ground-truth images, and the root mean square error (RMSE) of the solar azimuth angle is calculated. Ten key points are selected at the base of the building; their shadow cast lengths are measured; the relative error between the generated result and the ground-truth annotations is calculated. Finally, the Sobel operator is used to extract the brightness gradient field between the two, and the mean angle between the generated and ground-truth gradient vectors is calculated as a measure of the deviation in illumination gradient direction.

$$\mathrm{RMSE}_\theta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\theta^i_{\mathrm{gen}} - \theta^i_{\mathrm{real}}\right)^2} \tag{27}$$

$$\epsilon_{\mathrm{shadow}} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{l^i_{\mathrm{gen}} - l^i_{\mathrm{real}}}{l^i_{\mathrm{real}}}\right| \tag{28}$$

$$\Delta\phi = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\arccos\left(\frac{|\nabla l_{\mathrm{gen}} \cdot \nabla l_{\mathrm{real}}|}{\|\nabla l_{\mathrm{gen}}\|\|\nabla l_{\mathrm{real}}\|}\right) \tag{29}$$

Figure 3 shows the comparison results of the proposed method with Pix2PixHD, CycleGAN, and StyleGAN2 on three physical consistency metrics: RMSE of solar azimuth, relative error of shadow length, and illumination gradient direction deviation, in four typical urban environments (Beijing, Suzhou, Guangzhou, and Chengdu). Pix2PixHD, CycleGAN, and StyleGAN2 generally have high errors in different cities. The RMSE of the solar azimuth angle of Pix2PixHD in the four cities ranges from 17.8° to 19.1°, and that of CycleGAN is as high as 20.7° to 22.4°. The proposed method is only between 3.9° and 4.5°, with an average of 4.2°. In terms of shadow error, CycleGAN exceeds 27.8% in all cities, while the proposed method is stably controlled at 6.8% to 7.4%. The illumination gradient deviation is also reduced from 13° to 19.5° of the comparison method to 5.9° to 6.6°. The proposed method maintains extremely low and stable error levels regardless of urban building density (such as Suzhou's densely populated low-rise water towns) or diverse climate and lighting characteristics (such as Chengdu's frequent fog and overcast weather), demonstrating its strong generalization and ability to ensure physical consistency across diverse urban environments.

Pix2PixHD and CycleGAN rely on end-to-end mapping and lack explicit modeling of the physical laws of illumination, resulting in generated light and shadows that often violate geometric and optical laws. The proposed method embeds a differentiable physically-based rendering layer within the generator based on sun trajectory, material reflectivity, and depth normal maps. This ensures that the generated image's lighting direction and shadow length are consistent with the scene structure, fundamentally constraining the rationality of the lighting. At the same time, the GLO strategy optimizes the latent space, improving structural fidelity and avoiding light and shadow distortion caused by feature misalignment. This decoupled generation, which applies physical constraints prior to style modulation, enables the model to maintain high-precision reproduction of lighting parameters even in complex urban settings, thereby achieving stable and realistic light and shadow generation across diverse geographical, climatic, and architectural styles.

C. CULTURAL STYLE MATCH VERIFICATION

To assess the cultural style consistency of generated images, the generated images are fed into a pre-trained independent classifier for forward inference. The target labels used in verification are constructed from expert-curated samples that share consistent chromatic tendencies, repeatable façade

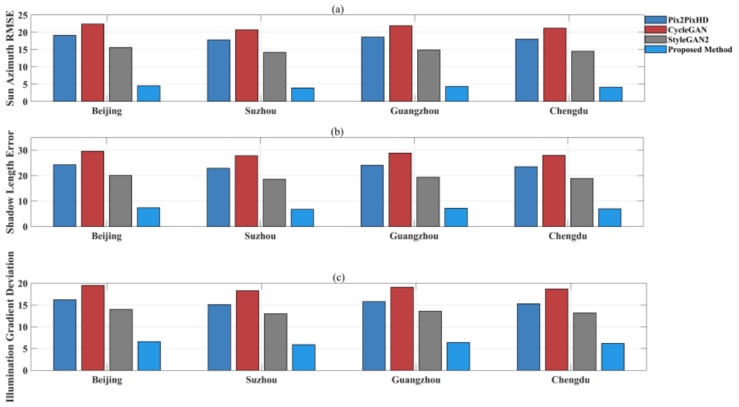

FIGURE 3. Physical Consistency Evaluation of Illumination. (a). RMSE of Solar Azimuth (°); (b). Relative Error of Shadow Length (%); (c). Illumination Gradient Direction Deviation (°)

ornament structures, and stable material reflectance patterns. These features are validated through a separate human evaluation group that confirms their perceptual coherence. The classifier predictions thus correspond to visual attributes that human viewers consistently associate with each cultural label, enabling an alignment between measurable image features and perceived cultural expression. Generated images are evaluated using a classifier trained on chromatic tendencies, facade ornament periodicity, composition ratios, and material reflectance descriptors. The resulting accuracy reflects the detectability of these measurable cultural attributes in the synthesized outputs. Across the generated samples, the cultural inputs influence the distribution of dominant hues, the spacing of ornament patterns, the proportion of façade segments, and the tonal responses of materials. These variations appear consistently when conditioning on the same cultural label, forming stable visual tendencies that correspond to the annotated attributes. At the same time, the category confusion entropy H_c is calculated, and the main style ratio is defined as the average of the maximum predicted probabilities in all generated images, which is used to measure the expression strength of the dominant style in the generated image.

$$H_c = -\sum_{c=1}^{10} p_{\text{gen}}(c) \log p_{\text{gen}}(c) \tag{30}$$

Figure 4 compares the cultural style match of the proposed method with Pix2PixHD, CycleGAN, and StyleGAN2 in four representative cities (Beijing, Suzhou, Guangzhou, and Chengdu), using three metrics: classification accuracy, class confusion entropy, and dominant style proportion. Pix2PixHD achieved only 70.2% accuracy in Beijing and further decreased to 66.8% in Suzhou. The proportion of dominant styles in any city did not exceed 73.4%, indicating a dispersion of styles. CycleGAN performed even worse, with an accuracy below 68% and a confusion entropy generally as high as 1.90. While StyleGAN2 showed some improvement, it was still lower than the proposed method overall. The proposed method showed higher concentration of target cultural attributes across scenes, indicated by an average accuracy of 92.7% and distinctly low confusion entropy. The dominant probability values remained stable, illustrating that the generated outputs retained attribute patterns associated with the intended cultural categories.

This performance breakthrough stems from the synergy between the cultural semantics-aware guidance mechanism proposed in this paper and the dual-driver architecture. Traditional methods lack explicit modeling of cultural characteristics, making the generated results susceptible to local texture interference and style drift. This method inputs city cultural labels into the generator as embedding vectors. Using AdaIN and an attention mechanism, it modulates style parameters at multiple scales, precisely activating features such as the white walls and black tiles of Jiangnan water towns and the colorful arcades of Lingnan arcades during the generation process. GLO-optimized latent spaces enhance structural fidelity, avoiding distortion of cultural elements caused by geometric misalignment. The generated results exhibit coherent chromatic structures, culturally characteristic ornament rhythms, and stable façade composition patterns under fixed cultural conditioning. These visual characteristics align with the annotated cultural attributes and remain consistent across different spatial contexts, even when adapting to different urban spatial forms and lighting conditions, truly achieving realistic light and shadow generation that captures both form and spirit.

D. OBJECTIVE EVALUATION OF IMAGE GENERATION QUALITY

To objectively evaluate the quality of generated images, the paper used Fréchet Inception Distance (FID) to measure the distribution distance between generated and real images in the Inception-V3 network feature space. The paper used Learned Perceptual Image Patch Similarity (LPIPS) as a perceptual difference metric, extracting deep features based on a pre-trained AlexNet, and calculated the average perceptual distance between generated and real images. It also calculated the structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) for each image on the entire test set to

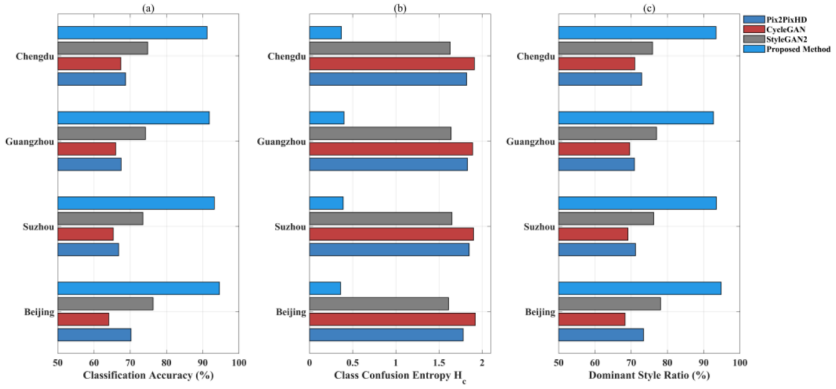

FIGURE 4. Cultural Style Match Verification. (a). Classification Accuracy; (b). Class Confusion Entropy; (c). Dominant Style Proportion

assess local structural fidelity and pixel-level reconstruction accuracy. Figure 5 shows an objective comparison of the quality of generated images of four typical cities using the proposed method, Pix2PixHD, CycleGAN, and StyleGAN2. The horizontal axis represents city categories (Beijing, Suzhou, Guangzhou, and Chengdu), while the vertical axes represent FID, LPIPS, SSIM, and PSNR. Pix2PixHD and CycleGAN achieved FIDs above 28, LPIPS exceeding 0.37, SSIM below 0.73, and PSNR below 24.5 dB in all cities, indicating significant distortion in the structural fidelity and perceptual quality of their generated images. StyleGAN2 significantly improved performance, with FID reduced to around 20, SSIM approaching 0.80, and PSNR reaching approximately 26.5 dB. The proposed method achieved top performance across all cities: FID was 15.2 in Beijing and as low as 14.6 in Suzhou. PSNR exceeded 29.0 dB in all cities, and the average SSIM reached 0.863, demonstrating its highly consistent output across diverse geographical environments and strong generalization capabilities.

The GLO-optimized latent space significantly improves the accuracy of the structural mapping from the input sketch to the output image, avoiding the blurring or misalignment of details often associated with encoding errors in traditional GANs. The physical lighting constraint module ensures that key visual cues like shadows and reflections adhere to real-world patterns, directly improving both SSIM and PSNR. The cultural semantic guidance mechanism enhances texture realism through style modulation and reduces LPIPS perceptual discrepancies. The multi-scale PatchGAN discriminator provides refined supervision in local areas, enabling the model to generate high-fidelity details even under the complex interfaces of Jiangnan water towns and the foggy lighting conditions of Chengdu, achieving stable and credible realistic light and shadow synthesis in diverse urban environments.

E. DETAIL CLARITY AND EDGE REPRESENTATION

To evaluate the detail clarity and edge expression of the generated image, the Sobel operator is applied to the generated image to extract the gradient amplitude, and its average value across the entire image is calculated as the average gradient. The image entropy is also calculated to measure the information complexity of local details:

$$H = -\sum_{k=0}^{255} p(k) \log p(k) \tag{31}$$

In formula (31), $p(k)$ is the normalized probability of the grayscale histogram. Figure 6 shows the detail clarity and edge representation of different methods in four cities. Pix2PixHD and CycleGAN generally have average gradients below 6.2 and image entropy values below 6.85, with the weakest performance in Suzhou and Chengdu, indicating blurred edges and monotonous textures in their generated images. StyleGAN2 achieved significant performance improvements, with an average gradient value of 7.20 and an entropy value of 7.02. The proposed method achieved an average gradient value of 8.63–8.71 and an entropy value of 7.31–7.38. In Suzhou, the average gradient value reached 8.71, and the entropy value reached 7.38, significantly outperforming other methods. This demonstrates its ability to maintain high sharpness and rich detail on complex interfaces, demonstrating excellent visual fidelity.

GLO optimization ensures clear building outlines and street boundaries from the initial generation stage. The physical lighting constraint module enhances the gradient response at the light-dark interface through differentiable shadow mapping, directly improving the average gradient value. The cultural semantic guidance mechanism enhances the texture density of traditional symbolic areas (such as window frames and plaques) in AdaIN modulation, significantly increasing image entropy. The multi-scale PatchGAN discriminator performs refined local edge discrimination at scales of 64×64 and above, forcing the generator to output high-contrast edges even in low-light or high-reflectivity conditions. This achieves realistic details at the microscopic level, providing highly valuable renderings for urban environmental art design.

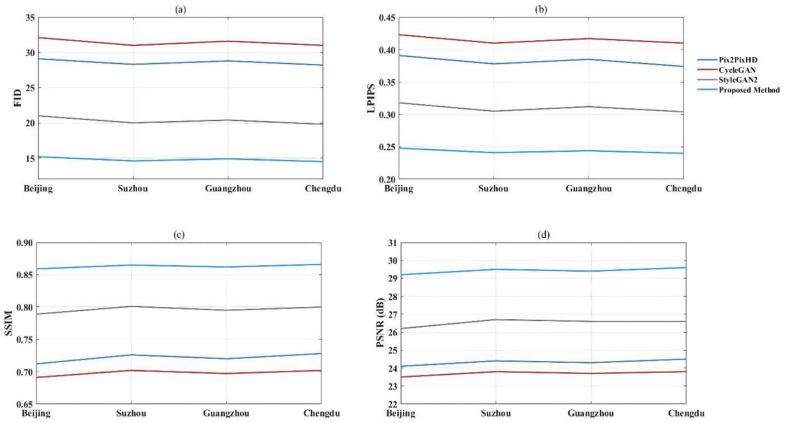

FIGURE 5. Objective Evaluation of Image Generation Quality. (a). FID; (b). LPIPS; (c). SSIM; (d). PSNR

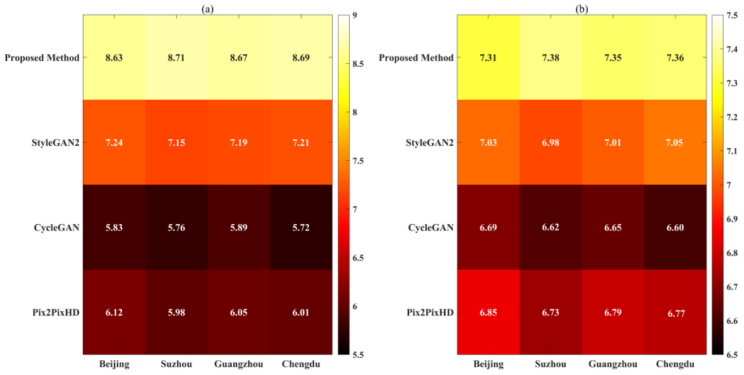

FIGURE 6. Analysis of Detail Clarity and Edge Representation. (a). Average Gradient; (b). Image Entropy

F. USER PERCEIVED REALISM AND DESIGN USABILITY TESTING

Thirty urban design experts with at least five years of experience were asked to perform a double-blind ranking of the generated and real images. They were then scored (1–5) based on three dimensions: "realism", "light and shadow rationality", and "design reference value". The mean and standard deviation for each dimension were calculated. Table 3 shows the expert evaluation results of the proposed method, Pix2PixHD, CycleGAN, and StyleGAN2 in terms of user-perceived realism and design usability. Pix2PixHD and CycleGAN received generally low scores, with realism scores of 2.83 and 2.65, respectively, and lighting and shadow plausibility scores of only 2.76 and 2.58, respectively. This indicates that their generated images are visually distorted and have confusing lighting and shadow logic, reducing their usability in urban art visualization scenarios. StyleGAN2 showed some improvement, with all three scores approaching 3.5. The proposed method achieved significant advantages across all dimensions, achieving a realism score of 4.48, a lighting plausibility score of 4.52, and a design reference value score of 4.43. Standard deviations for all scores were below 0.5, demonstrating high consensus among 30 experts on its performance, significantly outperforming other methods and demonstrating its high credibility at the human perceptual level and its potential for practical application. This outstanding performance stems from the proposed dual-driver architecture, which deeply integrates physical realism and cultural readability. Comparative methods focused solely on pixel-level reconstruction or style transfer, ignoring the physical consistency of lighting, shadow, and structure, leading to expert judgments of artificial renderings. This method embeds a physical lighting module based on sun trajectory and material reflectivity to ensure that shadow direction and light-dark transitions conform to real-world spatial logic, significantly improving lighting and shadow plausibility scores. Furthermore, a cultural semantics-guided mechanism ensures that the generated results align with regional characteristics in terms of color matching and decorative details, enhancing the communication of design intent. GLO optimization improves structural fidelity from sketches to renderings, avoids geometric misalignment, and imbues the generated images with clear design semantics. This approach has been widely recognized by experts as a high-value visualization tool, which is suitable for direct

use in visualization presentations within urban art scenarios, truly achieving the transition from generated images to design assets.

TABLE 3. User Perceived Realism and Design Usability Testing

Method	Realism Score	Lighting Plausibility Score	Design Reference Value Score
Pix2PixHD	2.83 ± 0.64	2.76 ± 0.71	2.91 ± 0.68
CycleGAN	2.65 ± 0.72	2.58 ± 0.69	2.74 ± 0.75
StyleGAN2	3.42 ± 0.58	3.36 ± 0.61	3.51 ± 0.63
Proposed Method	4.48 ± 0.41	4.52 ± 0.39	4.43 ± 0.47

G. VERIFYING STABILITY OF GENERATION ACROSS LIGHTING CONDITIONS

This paper used 10 different time-of-day and weather combinations on the test dataset and calculated the structural similarity (SSIM) and mean color difference between the generated results. The structural distortion rate was defined as the percentage of generated image edges that deviate from the true edge by more than 3 pixels. Figure 7 shows the stability trends of the generation results of the proposed method under 10 different illumination conditions. The horizontal axis represents the illumination scene number, and the vertical axes represent SSIM, average color difference, and structural distortion rate, respectively. The SSIM remained above 0.858 in the first seven conditions, dropping only to 0.837-0.851 in conditions 8-10. The average color difference remained stable below 5.5 in some conditions, though it rose to 6.82, but overall fluctuations were manageable. The structural distortion rate remained below 3.5% in the first seven scenarios, rising only to 4.6% and 5.1% in conditions 9 and 10, respectively, indicating good control of edge shifts. Overall, the proposed method maintains high structural fidelity and color consistency under most typical urban lighting conditions, demonstrating excellent cross-conditional generation stability.

The GLO strategy optimization maintains consistent geometric layout under different lighting conditions. The physical lighting module explicitly models the sun's trajectory, material reflectivity, and ambient occlusion to ensure that lighting changes adhere to real-world physical laws, avoiding structural distortion or color aberration caused by lighting estimation errors. A multi-scale discriminator continuously monitors local edges and textures during training, enabling the model to generate reasonable details even in low-contrast or complex reflective conditions. This dual guarantee mechanism which combines physics-based guidance and adversarial regularization, ensures that the generated results maintain visual consistency and design credibility in dynamic urban lighting environments, providing stable and reliable technical support for all-weather urban environment art design visualization.

H. GENERATION EFFICIENCY AND MODEL PRACTICALITY EVALUATION

The inference time from input to output for a single 512 × 512 image was measured, and the peak video memory usage was recorded. The model parameters were calculated by summing the total number of parameters in all learnable layers to verify its deployment feasibility on mainstream design workstations. The model size was estimated based on the number of parameters and storage accuracy to assess the storage and loading overhead of the model on the local workstation. Figure 8 compares the generation efficiency and model complexity of the proposed method with Pix2PixHD, CycleGAN, and StyleGAN2. Pix2PixHD, CycleGAN, and StyleGAN2 significantly were outperformed by the proposed method in terms of model size, inference time, number of parameters, and peak memory usage. Pix2PixHD's inference time reached 210 ms; its peak memory usage was 7.8 GB; its model parameters were as high as 48.6 MB, making it difficult to meet real-time design requirements. Despite improvements, StyleGAN2 still required 185 ms and 7.1 GB of peak memory. This method achieved the lowest performance across all metrics: inference time was only 118 ms; peak memory usage was as low as 5.8 GB; model size was 210 MB; parameter count was 34.2 M. This demonstrates significant computational efficiency improvements while maintaining high generation quality, making it feasible for deployment on mainstream design workstations.

This efficiency advantage stems from the dual optimization of model structure and training strategy employed in this architecture. The GLO strategy avoids redundant computation in traditional encoder-decoder architectures, significantly reducing parameter count and memory requirements. The physical lighting constraint module is embedded into the generator as a differentiable prior, reducing the discriminator's learning burden for complex lighting features and simplifying the network depth. The dual-branch structure achieves functional decoupling, allowing the generator to achieve multi-objective control without stacking too many convolutional layers. The combination of progressive upsampling and AdaIN-style modulation improves feature utilization efficiency, enabling the generation of high-fidelity details at a smaller model scale. This design paradigm, characterized by a lightweight architecture and a strong prior, enables the method to achieve improvements in both generation efficiency and deployment practicality while ensuring physical realism and cultural expression, thereby providing an integrated generation framework for urban art visualization.

I. QUALITATIVE EVALUATION OF ARTISTIC EXPRESSION AND DESIGN PLAUSIBILITY

Urban scenes exhibit significant cultural, spatial order, and material atmosphere differences within their visual expression systems. The resulting visuals often demonstrate varying degrees of visual integrity, compositional intent, and de-

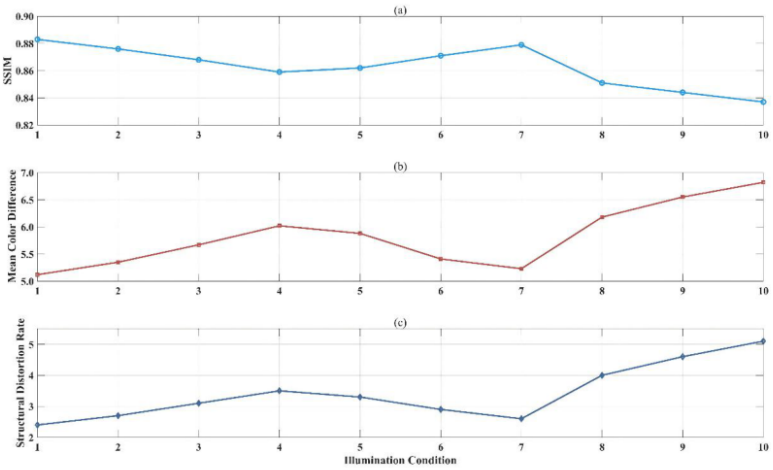

FIGURE 7. Stability Test of Generation Across Illumination Conditions. (a). SSIM; (b). Average Color Difference; (c). Structural Distortion Rate

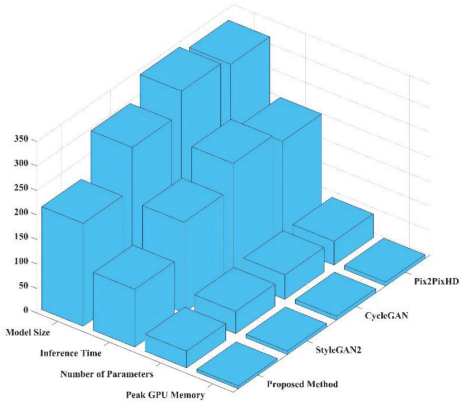

FIGURE 8. Generation Efficiency and Model Practicality Evaluation

sign potential. This section quantifies the artistic value, creative inspiration, and design feasibility of the output images within five urban contexts: Beijing, Suzhou, Guangzhou, Chengdu, and a comprehensive scene, to form an overall subjective assessment of the system's performance. These indicators correspond to the overall visual presentation, the extensibility of composition and spatial intent, and the clarity of structural expression. Their distribution is shown in Table 4 below.

TABLE 4. Subjective Evaluation Results for Urban Scene Types

Urban Scene Type	Artistic Value	Creative Inspiration	Design Feasibility
Beijing	4.37	4.18	4.29
Suzhou	4.46	4.27	4.15
Guangzhou	4.12	3.94	4.07
Chengdu	3.88	3.81	3.92
Composite Scene	4.21	4.06	4.13

The data reveals a clear regional stratification. Beijing and Suzhou led in artistic value and creative inspiration, with artistic values of 4.37 and 4.46, respectively. Both cities excelled in light and shadow organization, compositional tension, and stable output of cultural symbols, resulting in stronger visual harmony. Guangzhou's artistic value of 4.12 was slightly lower than the previous two categories, related to its higher proportion of modern architecture and complex cultural textures. Respondents' evaluation of creative inspiration tended to converge, with a value of 3.94, which was somewhat lower than the aforementioned two categories. Chengdu ranked lowest in all three indicators, with an artistic value of 3.88. This is related to its soft urban atmosphere and homogeneous scene elements, leading to a lack of significant stimulation in structural expression and emotional intent in the generated images. The Composite Scene value showed a convergent trend, with an artistic value of 4.21, reflecting an intermediate state in visual balance for multiscene mixed output. These results indicate that the system performs better in urban contexts with distinct cultural characteristics, and tends to be more stable in scenes with weaker structures or homogeneous

visual textures, forming a clear conclusion regarding urban relevance.

IV. CONCLUSION

This paper proposes a GLO-GAN method, driven by both physical constraints and cultural perception, for generating realistic light and shadow renderings in urban environmental art design. Recognizing the deep connection between material craft and cultural identity, the approach is further enriched by incorporating high-fidelity data from the industry, allowing the generator to synthesize environments where light realism is intrinsically tied to material and cultural authenticity. By constructing a decoupled generator architecture, integrating a differentiable physical lighting model with a cultural semantic guidance mechanism, and incorporating a GLO latent space optimization strategy, it achieves collaborative modeling of light and shadow physical realism and regional cultural expression. Experimental results demonstrate that this method reduces illumination parameter errors to 3.9°–4.5° azimuth RMSE and 6.8%–7.4% shadow length relative errors. The cultural style classification accuracy averages 92.7%. Furthermore, the generated images approach realistic renderings in subjective evaluations such as realism, lighting rationality, and design reference value, demonstrating their effectiveness and practicality in high-fidelity, explainable intelligent visualization generation.

AUTHOR CONTRIBUTIONS

Yue Cheng designed, collected and analyzed the data, and drafted the manuscript. Yue Cheng conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Yue Cheng participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

Research on the Integrated Development of Artistic Rural Construction and Folk Culture in the Context of Rural Revitalization Project Source: Shanxi Planning Office of Philosophy and Social Science Project No.: 2023YY303

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] B. Song and P. Puntien, "The light and shadow art in display space," Journal of Roi Kaensarn Academi, vol. 9, no. 5, pp. 88-101, 2024, [Online]. Available: https://so02.tci-thaijo.org/index.php/JRKSA/article/view/267519.

[2] O. Pidlisna, A. Simonova, N. Ivanova, V. Bondarenko, and A. Yesipov, "Harmonisation of the urban environment by means of visual art, lighting design, and architecture," Acta Scientiarum Polonorum Administratio Locorum, vol. 22, no. 1, pp. 59-72, 2023, doi: 10.31648/aspal.8214.

[3] Z. Liu, "Optimization and control of light and shadow effects in environmental art design based on particle swarm algorithm," Journal of Computational Methods in Sciences and Engineering, vol. 25, no. 4, pp. 3566-3579, 2025, doi: 10.1177/14727978251324143.

[4] H. Zhang, "Intelligent computer technology and its application in environmental art design," International Journal of Information and Communication Technology, vol. 24, no. 2, pp. 213-227, 2024, doi: 10.1504/IJICT.2024.137222.

[5] R. Chen, J. Zhao, X. Yao, S. Jiang, J. He, B. Bao, et al., "Generative design of outdoor green spaces based on generative adversarial networks," Buildings, vol. 13, no. 4, pp. 1083, 2023, doi: 10.3390/buildings13041083.

[6] H. Zhou and S. Xiang, "Applicability evaluation and reflection on artificial intelligence-based "image to image" generation of landscape architecture masterplans," Landscape Architecture Frontiers, vol. 12, no. 2, pp. 58-67, 2024, doi: 10.15302/J-LAF-1-020094.

[7] J. Yang and M. Song, "Color matching and light and shadow processing in intelligent interior environment art design analysis and application based on neural network," International Journal of Advanced Computer Science and Applications, vol. 15, no. 11, pp. 524, 2024.

[8] E. Mohammadrezaei, S. Ghasemi, P. Dongre, D. Gračanin, and H. Zhang, "Systematic review of extended reality for smart built environments lighting design simulations," IEEE Access, vol. 12, pp. 17058-17089, 2024, doi: 10.1109/ACCESS.2024.3359167.

[9] A. Jaglarz, "Perception of color in architecture and urban space," Buildings, vol. 13, no. 8, pp. 2000, 2023.

[10] Q. Lin, C. Zhang, H. Cai, X. Li, and H. Xiao, "Urban night lighting evaluation system and case study: blending popular contemporary elements, cultural traditions and advanced lighting technologies," Engineering, Construction and Architectural Management, vol. 31, no. 7, pp. 2916-2931, 2024, doi: 10.1108/ECAM-11-2021-1056.

[11] P. Cai, "Environmental landscape modeling design based on smart city public facilities," Mobile Information Systems, vol. 2022, Art. no. 6033336, 2022, doi: 10.1155/2022/6033336.

[12] C. Liu, M. Lin, H. L. Rauf, and S. Shareef, "Parameter simulation of multi-dimensional urban landscape design based on nonlinear theory," Nonlinear Engineering, vol. 10, no. 1, pp. 583-591, 2021, doi: 10.1515/nleng-2021-0049.

[13] W. Q. Ruan, G. X. Jiang, Y. Q. Li, and S. N. Zhang, "Night tourscape: structural dimensions and experiential effects," Journal of Hospitality and Tourism Management, vol. 55, pp. 108-117, 2023, doi: 10.1016/j.jhtm.2023.03.015.

[14] W. Gan, Z. Zhao, Y. Wang, Y. Zou, S. Zhou, and Z. Wu, "UDGAN: A new urban design inspiration approach driven by using generative adversarial networks," Journal of Computational Design and Engineering, vol. 11, no. 1, pp. 305-324, 2024, doi: 10.1093/jcde/qwae014.

[15] L. Y. Wang and Y. P. Huang, "Environmental landscape art design based on visual neural network model in rural construction," Ecological Chemistry and Engineering S, vol. 30, no. 2, pp. 267-274, 2023, doi: 10.2478/eces-2023-0028.

[16] K. Bai, Z. Li, L. Wang, Y. Hao, and X. Li, "The low-illumination 3D reconstruction method based on neural radiation field," Journal of Radiation Research and Applied Sciences, vol. 18, no. 2, Art. no. 101488, 2025, doi: 10.1016/j.jrras.2025.101488.

[17] X. Guo and J. Ma, "Heritage applications of landscape design in environmental art based on image style migration," Results in Engineering, vol. 20, Art. no. 101485, 2023, doi: 10.1016/j.rineng.2023.101485.

[18] D. Gao and Y. Zhang, "Automatic generation of landscape images based on deep generative modelling," International Journal of Information and Communication Technology, vol. 26, no. 30, pp. 43-59, 2025, doi: 10.1504/IJICT.2025.147762.

[19] W. Zhao, J. Zhu, J. Huang, P. Li, and B. Sheng, "GAN-based multi-decomposition photo cartoonization," Computer Animation and Virtual Worlds, vol. 35, no. 3, pp. e2248, 2024, doi: 10.1002/cav.2248.

[20] H. Zhao, W. Li, D. Huang, J. Huang, and L. Zhang, "M-GAN: multiattribute learning and multimodal feature fusionbased generative adversarial network for text-to-image synthesis," The Visual Computer, vol. 41, no. 5, pp. 3017-3035, 2025, doi: 10.1007/s00371-024-03585-y.

[21] T. Braure, D. Lazaro, D. Hateau, V. Brandon, and K. Ginsburger, "Conditioning generative latent optimization for sparse-view computed tomog-

raphy image reconstruction," Journal of Medical Imaging, vol. 12, no. 2, Art. no. 024004, 2025, doi: 10.1117/1.JMI.12.2.024004.
[22] Y. Shi, L. Han, L. Han, S. Chang, T. Hu, and D. Dancey, "A latent encoder coupled generative adversarial network (LE-GAN) for efficient hyperspectral image super-resolution," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-19, 2022, doi: 10.1109/TGRS.2022.3193441.
[23] Legito, R. Nuraini, L. Judijanto, Lubis, and AI, "The Application of Convolutional Neural Networks in Floristic Recognition," International Journal Software Engineering and Computer Science (IJSECS), vol. 3, no. 3, pp. 520-528, 2023, doi: 10.35870/ijsecs.v3i3.1827.
[24] H. Shakibania, S. Raoufi, and H. Khotanlou, "CDAN: Convolutional dense attention-guided network for low-light image enhancement," Digital Signal Processing, vol. 156, Art. no. 104802, 2025, doi: 10.1016/j.dsp.2024.104802.
[25] C. Liu, L. Wang, Z. Li, S. Quan, and Y. Xu, "Real-time lighting estimation for augmented reality via differentiable screen-space rendering," IEEE Transactions on Visualization and Computer Graphics, vol. 29, no. 4, pp. 2132-2145, 2023, doi: 10.1109/TVCG.2022.3141943.
[26] J. Hasselgren, N. Hofmann, and J. Munkberg, "Shape, light, and material decomposition from images using monte carlo rendering and denoising," Advances in Neural Information Processing Systems, vol. 35, pp. 22856-22869, 2022, [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/hash/8fcb27984bf16ca03cad643244ec470d-ABSTRACT-Conference.html.
[27] L. Wu, G. Cai, R. Ramamoorthi, and S. Zhao, "Differentiable time-gated rendering," ACM Transactions on Graphics, vol. 40, no. 6, pp. 1-16, 2021, doi: 10.1145/3478513.3480489.
[28] Y. Zhou, L. Wu, R. Ramamoorthi, and L. Yan, "Vectorization for fast, analytic, and differentiable visibility," ACM Transactions on Graphics, vol. 40, no. 3, pp. 1-21, 2021, doi: 10.1145/3452097.
[29] T. Edensor and R. Hughes, "Moving through a dappled world: the aesthetics of shade and shadow in place," Social & Cultural Geography, vol. 22, no. 9, pp. 1307-1325, 2021, doi: 10.1080/14649365.2019.1705994.
[30] M. Wada, Y. Ueda, J. Morioka, M. Adachi, and R. Miyamoto, "Dataset creation for semantic segmentation using colored point clouds considering shadows on traversable area," Journal of Robotics and Mechatronics, vol. 35, no. 6, pp. 1406-1418, 2023, doi: 10.20965/jrm.2023.p1406.
[31] J. Zhao, "Multiscale Detail Enhancement in Graphic Design Images Based on Visual Perception," International Journal of High Speed Electronics and Systems, vol. 34, no. 1, Art. no. 2540152, 2025, doi: 10.1142/S0129156425401524.
[32] X. Zeng, W. Ai, Z. Liu, and X. Wang, "Unsupervised Restoration of Underwater Structural Crack Images via Physics-Constrained Image Translation and Multi-Scale Feature Retention," Buildings, vol. 15, no. 13, pp. 2150, 2025, doi: 10.3390/buildings15132150.
[33] Q. Fang, C. Ibarra-Castanedo, Y. Duan, E. A. Jorge, G. Iván, and M. Xavier, "Defect enhancement and image noise reduction analysis using partial least square-generative adversarial networks (PLS-GANs) in thermographic nondestructive evaluation," Journal of Nondestructive Evaluation, vol. 40, no. 4, pp. 92, 2021, doi: 10.1007/s10921-021-00827-0.
[34] L. Abady, M. Barni, A. Garzelli, and B. Tondi, "Generation of synthetic generative adversarial network-based multispectral satellite images with improved sharpness," Journal of Applied Remote Sensing, vol. 18, no. 1, Art. no. 014510, 2024, doi: 10.1117/1.JRS.18.014510.
[35] X. Lu, "Interior lighting design and environmental art effect optimization based on genetic algorithm," Journal of Computational Methods in Sciences and Engineering, vol. 25, no. 1, pp. 712-727, 2025, doi: 10.1177/14727978251322024.