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

Generating an adversarial network-driven optimization method for 3D sculpture modeling

Wanglong Yu[1,*]

[1] Academy of Fine Arts, Shanghai University, Shanghai, 200436, China

SUMMARY: *With the rise of digital technology, the automated manufacturing capability for complex and fine models can be improved. For the 3D sculpture such as the process of fine design complex products, can be presented through the digital 3D model. In this paper, additional conditions are added as constraints on the basis of the original GAN generator, so that the generator and discriminator generate images or discriminate images according to the given conditions. At the same time, the diffusion model is utilized to remove the generated noise in the generated graphics. The 3D-GAN model is constructed by combining the body convolutional network with the generative adversarial network to generate 3D objects from the probability space. On this basis, the Transformer point cloud coding generation is introduced to refine the global features of the point cloud, and the Sigma model is used to reconstruct the 3D point cloud continuously from the sparse image of the sculpture points, so as to realize the reconstruction and optimization of the 3D modeling of the 3D sculpture point cloud. Using structural similarity and other indicators to evaluate the effect of 3D sculpture design, when* μ *is 0.5, the LPIPS value is the best, and the values are 0.1052, 0.1204, 0.1348, 0.1248 on four datasets, respectively, and comparing the error of the actual coordinate points and the coordinate points in the point cloud data, the error range is between pairs of 0.0001~0.0083, which indicates that sculpture point sparse image for 3D point cloud continuous reconstruction with better accuracy.*

KEYWORDS: *Diffusion model; 3D-GAN; Transformer point cloud coding; Sigma model; 3D sculpture; shape reconstruction optimization*

1 Introduction

The last several years saw digital art not so much integrate with traditional art and design processes as replace them in certain instances. The forms of digital expression are widely used in music, cinema, television, photography, animation, gaming, industrial design, visual communication, advertising, fashion design and numerous other related areas. Specifically, application of three-dimensional modelling in 3D film and television effects, digital game and animation has introduced a novel form of artistic representation, allowing new sculptures to find their way into virtual spaces [1] . More and more sculptors have turned towards computers to produce virtual sculptures. In the course of this, computer programs are used not only to create the sculptural forms themselves, but also to recreate the spatial environment around the work, which makes it look more attractive at the end result [2, 3]. Sculptural activity has increased out of physical materials into digital three-dimensional domains, and the creative process along with the works themselves are slowly transcending the limitations

*ywlsh_ea@163.com
https://doi.org/10.65102/is2026448
Published: 30 April, 2026

of material objects due to the use of 3D virtual modeling [4].

At present, the use of three-dimensional digital modeling in sculpture art still lacks a complete theoretical system. As the participation of three-dimensional digital design in sculpture design is a relatively new design method, coupled with three-dimensional digital molding technology is in its infancy. Most of the research results on this topic are three-dimensional digital design, three-dimensional digital design involved in sculpture design and other relatively one-sided, fragmented discussion and research, about the combination of the two systematic and comprehensive research information is still relatively small. For example, literature [5] started from the traditional hand-carved modeling method, taking into account the characteristics of conceptual design and artistic modeling design, using computer three-dimensional modeling software, put forward a polygonal modeling based on sculpture modeling modeling method of thought. Literature [6] proposed an efficient 3D-AWE (3D Weighted Architecture Estimation) method to analyze urban sculptures. 3D-AWE uses a min-max estimation model to compute the features in the image, which reveals a quantifiable representation of the sculpture's attributes through feature extraction. As outlined in reference [7] , the use of 3D modeling concepts to style sculptures can help solve issues that exist with conventional sculpture making, including a high level of manual labor and an increase in the cost of physical pieces as well as contribute to the further integration of modern art and computer graphics. The virtual restoration of sculptures using 3D modeling software is used by reference [8] to provide an opportunity to present the three-dimensional nature of urban sculptures more clearly. In such a manner, the design scheme is not only represented in the spatial form but also has more artistic characteristics, which results in the formation of a special artistic mood in the urban space. As suggested in reference [9] , the three-dimensional modeling technology can help in the creative process of sculptors by converting the ideas produced in a virtual world into physical objects of art and eventually coming up with products that can elicit human senses and communicate a material body. According to reference [10] , this technology presents a novel stage of sculpture creation, allowing the display of sculptural works in digital format, the exchange and distribution of information, and simultaneously suggesting a fuzzy affiliation-based surface genetic algorithm of image segmentation to overcome issues connected with the development of traditional sculpture works.

Generative artificial intelligence is developing rapidly and is profoundly changing the creative landscape of digital art, and in the field of digital sculpture, its technological innovations have brought unprecedented possibilities for artistic expression and creative methods [11, 12]. Around the application research of generative artificial intelligence in digital art, literature [13] analyzes the key features of Generative Adversarial Network (GAN) art design, which provides a new rational and perceptual, aesthetic and stochastic integration of ideas for the development of this field. Literature [14] investigates various applications of visual art, music, and literary texts generated with the help of GAN, and also compares and describes the performance of different generative AI architectures, and finally points out the key challenges of using generative AI for art design, and gives suggestions for future work. Literature [15] proposed a Deep Convolutional Generative Adversarial Network (DCGAN) model improved based on the dual attention mechanism, which is able to overcome the limitations of traditional sculpture color design with low texture accuracy. Literature [16] incorporates generative artificial intelligence and machine learning techniques in the process of digital modeling of sculpture shapes, which improves the accuracy and flexibility of creating digital sculptures, and the study achieves 92% model fidelity and a rendering speed of 45 frames per second, which provides a new perspective for sculpture design in terms of high efficiency and fast art creation. Literature [17] proposes a sculpture modeling generation

method based on a hybrid architecture of GAN and Convolutional Neural Networks (CNN), a framework that reveals the complex relationships between composition, texture and shape of digitally generated sculptures, aiming to break through the limitations of traditional art creation. At the intersection of digital sculpture and generative artificial intelligence, foreign researchers are creating a new era of digital art that integrates technology, art, and society to bring more rich and appealing possibilities to sculpture art.

In this paper, based on the generative principle of three-dimensional adversarial network, conditional generative adversarial network is proposed, and additional conditions are introduced as constraints to make the generator and discriminator generate images or discriminate images according to the given conditions. Based on the generative adversarial network model, using Transformer, geometric affine transform local features and multi-head attention mechanism local features, the generative encoder passes the features through the multilayer perceptual machine and the maximum pooling layer to get the encoded features, and the corresponding loss function is designed to realize the reconstruction and optimization of the sculpture point automated cloud three-dimensional modeling. The results of 3D sculpture modeling optimization design are evaluated by SSIM, PSNR, LPIPS and other evaluation indexes. The beauty degree evaluation method is proposed, and the evaluation index is applied to analyze the characteristics of higher-scoring 3D sculptures, and suggestions are made for the next application of 3D sculptures.

2 3D Sculpture Modeling Generation Based on Generative Adversarial Networks

2.1 3D Sculpture Modeling

2.1.1 Principle of Three-Dimensional Confrontation Generation

The original GAN generator takes random noise as input, which means that the generated image data is random and uncontrollable, in order to make image generation can be carried out in the specified direction, some scholars proposed conditional generative adversarial network (CGAN), which introduces additional conditions as constraints, so that the generator and discriminator can generate an image or discriminate an image in accordance with a given condition, where the constraints introduced can be the data types such as semantic images, category labels and text and other data types.CGAN is mostly applied in the fields of image translation, image restoration, and text-guided image generation, and has achieved better results.Figure 1 shows the schematic diagram of conditional generative adversarial network.

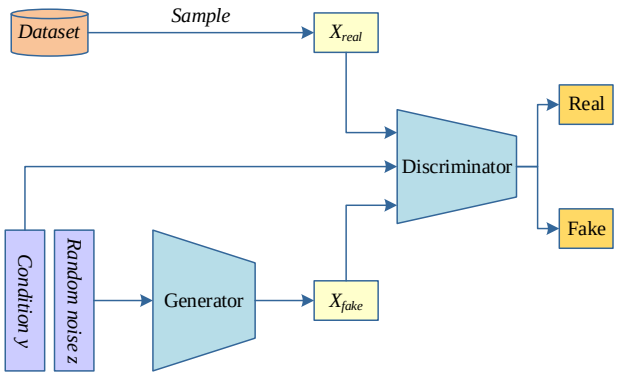

Figure 1: Schematic diagram of conditional Generative adversarial networks

The random noise z and the introduced constraints y are connected as hidden vectors input to the generator G to generate the image X_{fake}, and the discriminator takes the real sample data X_{real}, the fake data X_{fake}, and the constraints y as the inputs, and determines whether the generated image is real or not and whether it is in line with the conditional information or not.

The training objective function of the conditional generative adversarial network is basically the same as that of the original GAN, the difference is that the generator G and the discriminator D of the conditional generative adversarial network both increase the constraints y, and its objective function is as follows:

$$\begin{aligned}\min_G \max_D v(D,G) = & E_{x\sim P_r}[\log D(x,y)] \\ & + E_{z\sim P_z}\left[\log\left(1-D(G(z,y))\right)\right]\end{aligned} \tag{1}$$

2.1.2 Diffusion models

The denoising diffusion model takes random Gaussian noise as input in the reverse noise removal process, and after a certain number of steps of noise removal operations, the final generated results are obtained. However, only using random Gaussian noise as input generates random and uncontrollable results, and it is not possible to control the generated content according to the user's wishes.

Given a pixel-aligned paired dataset $\mathcal{D}=\{x_i, y_i\}_{i=1}^N$, where y denotes the conditional information and x denotes the target image. The conditional diffusion model starts from pure Gaussian noise $x_T \sim \mathcal{N}(0, I)$, and generates sequential images $x_T, x_{T-1}\cdots x_0$ based on the learned conditional transformed distribution parameter representations $p_\theta(x_{t-1}\mid x_t, y)$ through successive iterative refinement denoising so that the final sample data x_0 has the same distribution $x_0 \sim p(x\mid y)$ as the original data. The forward Gaussian diffusion process and the inverse noise removal process are viewed as two Markov chains with opposite directions to each other, with the difference that the conditional diffusion model introduces conditional information in the inverse noise removal process to control the direction of image generation. The intermediate image distribution in the image sampling process is defined by the forward Gaussian diffusion process, which gradually adds Gaussian noise to the original image data through a Markov chain, which can be expressed as $q(x_t\mid x_{t-1})$, and the trained parametric model recovers the target image from the noise through the Markov chain conditioned on y, thus realizing the reversal of the Gaussian diffusion process, reversing the Markov chain of Gaussian diffusion process can be described by Eq. (2):

$$p_\theta(x_0, x_1\cdots x_{T-1}\mid x_T) = \sum_{t=1}^{T} p_\theta(x_{t-1}\mid x_t, y) \tag{2}$$

where $p_\theta\left(x_{t-1}\mid x_t, y\right) = \mathcal{N}\left(x_{t-1}; \mu_\theta\left(x_t, \left(1-\bar{\alpha}_{\bar{t}}\right), \sigma^2 I\right)\right.$. To optimize the reversal process of Gaussian diffusion, the image y is used as a constraint input to the neural network parametric model f_θ, and the target image x_0 is recovered by taking the target image $\tilde{x}$, which contains noise, also as an input.

In addition to the conditional image y and the target image $\tilde{x}$ containing noise, the

conditional denoising model $f_\theta(y, \tilde{x}, \beta_t)$, which takes as input the statistical distribution of the variance of the noise, is iteratively trained to predict the noise vector z. By adjusting the scale size of β_t so that the parametric model f_θ can predict the noise vector z with different scale sizes, the training objective function of the whole parametric model can be described by Equation (3):

$$\mathbb{E}_{y,x,z,\beta} \left\| f_\theta(y, \sqrt{\beta} x_0 + \sqrt{1-\beta} z, \beta) - z \right\|_p^p \tag{3}$$

where (x, y) is the data sample of the training set, $p \in \{1, 2\}$ and $\beta \sim \mathcal{N}(0, I)$. Let $\beta_t = 1 - \bar{\alpha}_{\bar{t}}$, we can get $p_\theta(x_{t-1} \mid x_t, y) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, \beta_t)), \sigma^2 I)$, in addition to deriving the sampled image at any t moment after some algebraic operations:

$$x_{t-1} = \mu_\theta(x_t, y, t) + \sigma_\theta(x_t, y, t) \cdot z \tag{4}$$

2.2 Assessment of indicators

2.2.1 Peak signal-to-noise ratio

The Peak Signal-to-Noise Ratio (PSNR) is used as a measure of image or video quality and has been commonly employed in image generation literature. Performance of a generation model can be measured through computation of the PSNR value using the mean square error between the produced image and the reference image. Lower PSNR usually means more distortion on the synthesized output, and higher PSNR means that generated image is more close to the ground truth and thus more of a good quality. PSNR formula can be represented as:

$$PSNR = 10 \cdot \log_{10} \left(P^2 / L_2 \right) \tag{5}$$

where L_2 is the mean squared error between the produced result and the reference image, and P stands for the maximum pixel intensity.

2.2.2 Structural similarity

Structural Similarity (SSIM) is a measure that can be used to compare the similarity of two images as well as often used in image generation problems. SSIM takes into account three main attributes of visual data, i.e., luminance, contrast and structural consistency. It uses the mean, variance and covariance of an image to approximate these three factors before adding them together multipliatively to get the final SSIM score. The higher the SSIM value, the clearer the quality of the image or video and the lesser the level of perceived distortion. Its computation is shown below:

$$SSIM_l(X,Y)=\frac{2\mu_X\mu_Y+K_1}{\mu_X^2+\mu_Y^2+K_1}$$
$$SSIM_c(X,Y)=\frac{2\sigma_x\sigma_y+K_2}{\sigma_x^2+\sigma_y^2+K_2} \quad (6)$$
$$SSIM_s(X,Y)=\frac{\sigma_{xy}+K_3}{\sigma_x+\sigma_y+K_3}$$

$$SSIM(X,Y)=SSIM_l\times SSIM_c\times SSIM_s \quad (7)$$

where X,Y are the real image and the generated image respectively, K_1,K_2,K_3 are constants denoting the smoothing factors, μ_x,μ_y denote the mean of the real image and the generated image respectively, σ_x,σ_y denote the image's standard deviation, and σ_{xy} denotes the covariance of the two images.SSIM takes values in the range of $[0,1]$, with 1 denoting that the real image and the generated image are exactly the same, and 0 denoting that there is no similarity whatsoever.

2.2.3 Perceived similarity

Learned Perceptual Image Patch Similarity (LPIPS) is a measure of comparing images based on visual difference measurement that is closer to human perception. In contrast to traditional pixel-based or structural similarity metrics, LPIPS aims at modeling how humans really perceive images content, rendering it more useful when evaluating perceptual consistency. The approach uses a trained convolutional neural network to extract deep feature representations of images, and then computes the similarity of two images by estimating the distance between the corresponding features in the feature space. One common way to compute this distance is using Euclidean distance or cosine distance to calculate the similarity score based on the generated result compared to the ground-truth image. As an example, let us take Euclidean distance to express LPIPS in terms of the formula:

$$LPIPS(X,Y)=\sum_{j=0}^{H}\sum_{i=0}^{5}\left(f_i(x_j)-f_i(y_j)\right)^2 \quad (8)$$

where X,Y denotes the input predicted and real images, $f_i(x_j)$ and $f_i(y_j)$ denote the features extracted from the jth element at the ith layer, and $f(\cdot)$ denotes the pre-trained neural network model. After calculation, the final similarity score is in the range between 0 and 1, where a larger value indicates a lower similarity and vice versa.

3 GAN-based 3D sculpture ensemble reconstruction

3.1 GAN-based point cloud shape reconstruction

The 3D-GAN algorithm was presented in 2016 as the first architecture that combines a volumetric convolutional network with a generative adversarial network to generate 3D objects based on a probabilistic space, which enhances the completion of point-cloud data. In this approach, 3D structures are derived using latent representations using a volumetric convolution-based GAN. A random 200-dimensional latent vector is converted to a 64x64x64

voxel cube by the generator in 3D-GAN. The discriminator will then assess the validity of the generated 3D shape by giving a confidence score following the representation of the shape in voxel space, utilizing this cube.

3.1.1 Transformer-based point cloud encoding methods

The network model based on Transformer has better network performance in 2D image and text tasks, Transformer is an encoder-decoder architecture with a multi-head self-attention mechanism at its core that generates attention weights based on the global context of the input.

On this basis, a point agent based Transformer point cloud coding method is proposed in the field of point cloud processing, where the point cloud and points are regarded as sentences and words in a textual information task, and the global features of the point cloud are refined by the Transformer by learning based on the positional relationships of the points. The point agent obtains the geometric relations in the point cloud by using the GAN structure through an encoder with N multiple self-attentive layers and a feedforward network layer, which queries the nearest features by coordinates based on the given query point coordinates. Local geometric structures are learned through feature aggregation and maximum pooling operations in the linear layer. The structure of transformer based point cloud encoder is shown in Fig. 2.

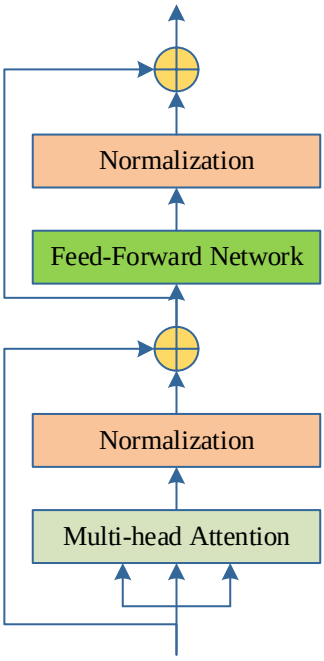

Figure 2: Point cloud encoder architecture based on transformer

The multi-attention mechanism allows the network to jointly attend to information from different representational subspaces at different locations, given the input value V, key K and query Q, the multi-attention is computed by the following formula:

$$MulHead\left(F2,K,V\right)=Concat\left(head_1,head_2,\cdots,head_n\right)W^O \tag{9}$$

where: W^O outputs the linear layer and the weights of each head feature can be obtained by:

$$head_i = soft\max\left(\frac{QW_i^Q\left(KW_i^K\right)^T}{\sqrt{d_k}}\right)VW_i^V \tag{10}$$

3.1.2 Geometric affine transform-based local feature attention coding

In the point cloud segmentation task, due to the presence of sparse and irregular geometrical structures in the local regions of the point cloud model, which causes the accuracy and stability degradation in the MLP feature extraction process, and makes the model less robust, different geometrical structures between different local regions may require different extractors, but it is difficult to realize this with common shared MLPs. Inspired by this, this paper introduces a lightweight geometric affine module in the point cloud complementation task to address the sparsity of local structures in point cloud models.

In this paper, local points are extracted before and after the aggregation operation of the point cloud data respectively, let $\left\{f_{i,j}\right\}_{j=1,2,3,\cdots,k} \in \mathbb{R}^{k\times d}$ be the grouped local neighborhood of $f_i \in \mathbb{R}^d$ contains k points, and each adjacent point $f_{i,j}$ is a d-dimensional vector. In this paper, the local neighbor points are transformed by the following formula:

$$\left\{f_{i,j}\right\} = \alpha \odot \frac{\left\{f_{i,j}\right\} - f_i}{\sigma + \varepsilon} + \beta \tag{11}$$

where $\alpha \in \mathbb{R}^d$ and $\beta \in \mathbb{R}^d$ are learnable parameters, $\odot$ denotes the Hadamard product and $\varepsilon = 1e^{-5}$, and σ is a parameter that describes the characterization of all the local group and channel deviations of the scalar. By transforming the point cloud into a geometric affine transformation, this paper transforms the local points into a normal distribution while maintaining the original geometric properties of the point cloud model.

3.1.3 Localized Feature Encoder Based on Multiple Attention Mechanisms

The retrieval of local geometry facts is important but challenging when completing point clouds and existing GAN-based approaches to local shape modeling are extremely sensitive to changes in point cloud density. In order to solve this problem, the paper proposes to create a feature extraction module to learn shape representation based on the Transformer architecture. Namely, it involves the use of multi-head cross-attention as well as self-attention mechanisms that allow capturing geometric properties of the point cloud in a more effective manner and learning its local structure features indirectly.

The feature extractor receives an input (X, d), where X is a matrix of size $n \times c$, each row of X can be regarded as a feature vector corresponding to a point, and d is the downsampling rate.By applying the farthest point sampling (FPS) algorithm, this paper obtains a downsampled point cloud feature matrix Y of size. Then, the feature matrix F is learned by utilizing multiple cross connects in the form of residuals. Secondly, receiving the acquired features F_1 and setting the downsampling ratio d for the farthest point sampling algorithm, the downsampled point cloud feature matrix F_2 of size $(n/d) \times a$ is obtained, and then, the residual form of the multi-head cross-attention mechanism is utilized to learn the features F_2 and acquire the corresponding features F_3:

$$F_3 = f\left(F_2 + MultHead\left(F2, K, V\right)\right) \tag{12}$$

$$F_4 = cat\left(F_3 + FFN\left(F_2\right), F_2\right) \tag{13}$$

where: f denotes the summation and normalization operation by Layer Norm, F_2 denotes the input features, $MultHead\left(F2, K, V\right) = Concat\left(head_1, head_2, \cdots, head_n\right)W^O$ ($n = 4$ is taken in this paper), the size of feature F_3 is $(n/d)\times 2a$, FFN denotes the feed-forward network and connects the existing feature F_2 to further update the feature F_3, and $cat(\ )$ denotes the feature F_3 is updated via the torch.nn.cat function to perform the feature vector connection operation to improve the model fitting ability by feed-forward neural network FFN. The feature F_4 is passed through a multilayer perceptron and a maximum pooling layer to obtain the coded feature F_5.

3.1.4 Loss function

In this network, CD loss is chosen as the measure.CD loss calculates the average nearest point distance between the predicted point cloud S and the real point cloud S':

$$d_{CD}(S, S') = \frac{1}{|S|}\sum_{x\in S}\min_{y\in S'}\|x-y\|_2^2 + \frac{1}{|S'|}\sum_{y\in S'}\min_{x\in S}\|y-x\|_2^2 \tag{14}$$

Assuming the seed point cloud, the output point clouds of the two cascade point generators are denoted as $P0$, $P1$, and $P2$, respectively. Meanwhile, the ground truth point cloud is downsampled by FPS to obtain three sub-clouds $S0$, $S1$, and $S2$, which have the same densities $P0$, $P1$, and $P2$, respectively. Then the loss of the model can be defined as follows:

$$L_1 = \sum_{i=0}^{2}\lambda_i d_{CD}(P_i, S_i) \tag{15}$$

where: $\lambda_i = 1$.

3.2 Optimization of 3D reconstruction of the automatic cloud of sculpture points

3.2.1 Sculpture point sparse image feature extraction

The process described above uses 3D corner-point detection and edge-contour feature extraction to optimize the existing automatic point-cloud-based 3D reconstruction algorithm in the field of sculpture design, and the process is based on identifying local features of sparse sculpture point images. The use of a gradient-based operator is used to decompose features and restore contour-edge information using the sparse image information of sculpture points. The transversality principle is then applied and a group of sparse linear equations is formulated to achieve progressive segmentation of the sparse sculpture point image, and finally, the matching value of the three-dimensional shape template is achieved:.

$$\frac{\left\langle \tau_d u', \tilde{u} \right\rangle_{\varphi_{x_0}}}{\left\| \tau_d u \right\|_{\varphi_{x_0}} \left\| \tilde{u} \right\|_{\varphi_{x_0}}} - \frac{\left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}} \left\langle \tau_d u', \tau_d u \right\rangle_{\varphi_{x_0}}}{\left\| \tau_d u \right\|_{\varphi_{x_0}}^3 \left\| \tilde{u} \right\|_{\varphi_{x_0}}} \tag{16}$$

The resulting sparsity decomposition process for a sparse image of a sculptured point is obtained as:

$$\begin{aligned} \frac{\partial}{\partial d}\left(\frac{\left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}}}{\left\| \tau_d u \right\|_{\varphi_{x_0}} \left\| \tilde{u} \right\|_{\varphi_{x_0}}} \right) &= \frac{\left(\left\| \tau_d u \right\|_{\varphi_{x_0}} \frac{\partial}{\partial d}\left(\left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}} \right) - \left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}} \frac{\partial}{\partial d}\left(\left\| \tau_d u \right\|_{\varphi_{x_0}} \right) \right)}{\left(\left\| \tau_d u \right\|_{\varphi_{x_0}} \right)^2 \left\| \tilde{u} \right\|_{\varphi_{x_0}}} \\ &= \frac{\frac{\partial}{\partial d}\left(\left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}} \right)}{\left\| \tau_d u \right\|_{\varphi_{x_0}} \left\| \tilde{u} \right\|_{\varphi_{x_0}}} - \frac{\left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}} \frac{\partial}{\partial d}\left(\left\| \tau_d u \right\|_{\varphi_{x_0}} \right)}{\left(\left\| \tau_d u \right\|_{\varphi_{x_0}} \right)^2 \left\| \tilde{u} \right\|_{\varphi_{x_0}}} \\ &= \frac{\left\langle \tau_d u', \tilde{u} \right\rangle_{\varphi_{x_0}}}{\left\| \tau_d u \right\|_{\varphi_{x_0}} \left\| \tilde{u} \right\|_{\varphi_{x_0}}} - \frac{\left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}} \left\langle \tau_d u', \tau_d u \right\rangle_{\varphi_{x_0}}}{\left(\| \ \tau_d u \|_{\varphi_{x_0}} \right)^2 \left\| \tilde{u} \right\|_{\varphi_{x_0}}} \\ &= \frac{\left\langle \tau_d u', \tilde{u} \right\rangle_{\varphi_{x_0}}}{\left\| \tau_d u \right\|_{\varphi_{x_0}} \left\| \tilde{u} \right\|_{\varphi_{x_0}}} - \frac{\left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}} \left\langle \tau_d u', \tau_d u \right\rangle_{\varphi_{x_0}}}{\left(\left\| \tau_d u \right\|_{\varphi_{x_0}} \right)^3 \left\| \tilde{u} \right\|_{\varphi_{x_0}}} \end{aligned} \tag{17}$$

where $\left\| \tau_d u \right\|_{\varphi_{x_0}}$ denotes the merged weakly convex component eigenvolume, $\left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}}$ denotes the residual of the signal decomposition over the best atom, and $\left\langle \tau_d u, \tilde{u} \right\rangle_{\varphi_{x_0}}$ denotes the signal decomposition on the best atom, $\left\langle \tau_d u', \tau_d u \right\rangle_{\varphi_{x_0}}$ denotes the residual component after the best matching, $\left\| \tilde{u} \right\|_{\varphi_{x_0}}$ denotes the atoms used for coefficient decomposition. The gradient operation method is used for feature decomposition, and the information fusion process is performed on the detected point cloud data of the sparse image of sculpture points to obtain the edge information feature components of the sparse image of sculpture points:

$$is\_visible(M_d(C_i), TC) = \begin{cases} 1, if \begin{cases} j \neq i, \\ C_j \ may \ call \ M_{mi} \end{cases} \\ 0, otherwise \end{cases} \tag{18}$$

where TC is the sculpture point sparse image reconstruction smoothing operator, $M_d(C_i)$ denotes the key feature point in the point cloud C_i, and the successive reconstructed feature component at the point (a, b_m) of the reconstructed sculpture point sparse image surface is:

$$L(a, b_m) = \log\left(M_d(C_i) \frac{|V||V_m \cap V_n|}{|V_m||V_n|} \right) \tag{19}$$

In the above equation, $|V|$ denotes the set of sparse images, $|V_m|$ denotes the set of sparse images in the neighborhood, $|V_n|$ denotes the set of sparse images corresponding to the sampling points, and $|V_m \cap V_n|$ denotes the set of sparse images corresponding to the sampled points in the domain. In the point cloud reconstruction, the local surface at the point is fitted with the moving least squares method, and the image 3D reconstruction is performed based on the sparse scattered point 3D reconstruction and sharpened template feature matching method.

3.2.2 Automatic cloud 3D reconstruction output of sculpture points

The Sigma model is used to reconstruct the sculpture point sparse image with 3D point cloud successive reconstruction, and the block segmentation and template feature matching process is performed on the sculpture point sparse image in the neighborhood with a search radius in the range of r, i.e.:

$$L(a,b_m)=\sum_{V_m\in P^{res}}\sum_{V_n\in P^{ture}}\frac{|V_m\cap V_n|}{|V|}\log\left(\frac{|V||V_m\cap V_n|}{|V_m||V_n|}\right) \tag{20}$$

In the above equation, p^{res} denotes a hypothetical measure of the degree of image distortion, and p^{ture} denotes a real measure of the degree of image distortion. On the matching points on the grid model, the information enhancement process of the sparse image of the sculpture point is carried out, thus obtaining the matching points of the grid model of the sparse image of the sculpture point:

$$\overline{x_T}=\frac{1}{T}\sum_{i=1}^{T}x_i \tag{21}$$

where: $x_1,x_2,x_3\cdots x_T$ is the feature subsample of the template shape, and T is the feature component of the sculpture point auto-cloud 3D reconstruction. The statistical shape model of the sparse image of the sculpture point is established, and Taubin smoothing is implemented on the whole grid model to obtain two neighboring pixel sets as:

$$F=\tilde{p}(x,y)=p(x,y)\left(\frac{v(x)}{v(y)}\right)^{1/2} \tag{22}$$

In the above equation, $\tilde{p}(x,y)$ denotes the gray value of the sparse component reconstructed image, $p(x,y)$ denotes the pixel approximated gray value after noise cancellation, $v(x)$ denotes the deformation vector in the horizontal direction and $v(y)$ denotes the deformation vector in the vertical direction. Where:

$$p(x,y)=\frac{k(x,y)}{v(x)},v(x)=\sum\nolimits_y k(x,y) \tag{23}$$

Neighborhood search method is used for fast feature point localization and information enhancement processing of sparse images with sculptured points, and the diameter of the computational grid model is obtained:

$$E\,\mathrm{int}(vi)=\frac{1}{2}F\left(\partial i\left|\mu-\left|vi-vi-1\right|\right|^{2}+\beta i\left|vi-1+2vi+vi+1\right|^{2}\right) \tag{24}$$

In the above equation, μ denotes the scale reference value, v_i denotes the deformation vector of the ith curve, ∂i denotes the derivative of the gray level of the sparse image of the sculpture point, and βi denotes the control point relative to the ith curve. Where:

$$\mu=\frac{1}{n}\sum_{i=0}^{n-1}\left|vi-vi-1\right| \tag{25}$$

According to the above processing, the sculptured point auto-cloud feature decomposition is performed for each layer l in the longitude direction, and P_n and P_{n+1} denote the sampled point cloud after the nth and the $n+1$th point acquisition, to obtain the output gradient vector $T(g_i)$ of the point cloud reconstruction formulated as follows:

$$G_{new}=(1+\mu T)(1+\lambda T)G_{old} \tag{26}$$

$$T(g_i)=\frac{1}{\sum_k T_n}\cdot\sum\nolimits_k T_n\left(P_{n+1}-P_n\right) \tag{27}$$

where G_{new} and G_{old} sample the point cloud matching coefficients and deformation parameters, respectively. $T_n=\left(f,g,h\right)_n$, which denotes the deformation parameter from P_n to P_{n+1}, and in summary, the sculpture point auto cloud 3D reconstruction expression is obtained as:

$$\begin{aligned} d_j^{k+1}&=(1-\mu)d_j^k \\ &\sum_{j\in N(i)}\left(E\,\mathrm{int}(vi)\frac{g_i+g_j}{2}+(1-\mu)\right)d_j^{k+1} \\ &+\sum_{j\in N(i)}\left(E\,\mathrm{int}(vi)\frac{g_i+g_j}{2}+(1-\mu)\right)d_j^{k}-\frac{T(g_i)}{G_{new}} \\ &+\frac{}{\sum_{j\in N(i)}\left(E\,\mathrm{int}(vi)\frac{g_i+g_j}{2}+(1-\mu)\right)} \end{aligned} \tag{28}$$

The above equation represents the grayscale pixel characteristics of the sparse image of the sculpture point, in summary, the implementation of the algorithm to improve the design of the reconstructed image maintains and enhances the low-frequency component, which improves the reconstruction effect.

3.3 3D sculpture modeling results

3.3.1 Structural similarity

Structural similarity measures are used to measure the similarity of the produced images with the reference images test1.jpg in the 3D sculpture design object dataset and cat1.jpg in the 3D sculpture-source inspired dataset, hence, assessing the performance of the 3D-GAN model in

generating the images. In particular, test1.jpg is the evaluation sample following 3D-GAN training of 500, 700, 1000, 1300, 1500, 1700, 2000, 5000, 10000, 20000, 30000, and 50000 steps along with 12 generated images. The structural similarity values of the evaluation samples and test1.jpg are calculated as SSIM1, and those between the evaluation samples and cat1.jpg are represented as SSIM2. The SSIM values averaged at identical training stages and the results of each stage are provided in Table 1. As per the SSIM analysis, besides the image produced after 500 training steps with an SSIM2 value of 0.7556, the similarity scores of all other models, as well as their mean values, do not fall below 0.8. In addition, the variations between SSIM values at different training stages are insignificant and are only recorded in the second decimals. Thus, based on the data in Table 1, we can conclude that all training stages in the 3D sculpture morphology generation experiment attained desirable generation quality, and the number of training iterations did not have significant effects on the end results of the generation.

Table 1: Calculation value of SSIM

Rating criteria	Training steps											
	500	700	1000	1300	1500	1700	2000	5000	10000	20000	30000	50000
SSIM1	0.8869	0.8269	0.8345	0.8345	0.8758	0.8269	0.8045	0.8636	0.8563	0.8636	0.8466	0.8348
SSIM2	0.7556	0.8245	0.8166	0.8324	0.8469	0.8369	0.8036	0.8545	0.8648	0.8618	0.8433	0.8316
SSIM	0.82125	0.8257	0.82555	0.83345	0.86135	0.8319	0.80405	0.85905	0.86055	0.8627	0.84495	0.8332

3.3.2 Signal-to-Noise Ratio and Perceived Similarity

The training set is derived from 800 high-definition images from the publicly available dataset DIV2K, and 32515 non-overlapping sub-images of size 480×480 obtained by cropping the images using a sliding window are expanded, and these images are interpolated bicubically to obtain the corresponding downsampled images, which are used in the generator. In order to fit 3D-GAN, the experiments only consider 4x zoom factor. In this paper, four widely used benchmark datasets, Set5, Set15, BSD100 and Urban120, are used as the test set, and the number of images included are 5, 15, 100 and 120, respectively.PNSR, SSIM and LPIPS are used as the evaluation metrics for the experiments in this paper, where the lower the value of LPIPS is, the higher the perceived similarity is, i.e., the reconstructed image is visually quality-wise closer to GT.For a fair comparison, bars of 4 pixel width size were removed from each boundary of all evaluated images and the Y channel of the image was used for PSNR and SSIM measurements to go to the computation. Table 2 shows the evaluation metrics for different μ values.

Firstly, the β is fixed and set to $\beta_{1.3}$, and then the experimental effects under different μ values are compared. Table 2 shows the evaluation indexes under different μ values, giving more convincing proofs from the metrics that all the indexes are greatly improved after adding the double perceptual loss under the appropriate μ. At μ of 1, the best PNSR values are obtained for each dataset, which are 27.4425, 26.5221, 25.9645, and 25.1965 on Set5, Set15, BSD100, and Urban120, respectively, while the LPIPS values are next to the best, which are obtained at μ of 0.5 on Set5, Set15, BSD100, and Urban120 with LPIPS values of 0.1052, 0.1204, 0.1348, and 0.1248, respectively, while most of the PNSR values are second best. Here in this paper, we focus on de-considering the LPIPS value, i.e., perceived similarity, because LPIPS is more in line with human perceptual habits in terms of evaluation compared to PNSR, and therefore the value of μ is determined to be 0.5.

Table 2: The evaluation metrics under different μ values

Test set evaluation index	Set5		Set15		BSD100		Urban120	
	PNSR	LPIPS	PNSR	LPIPS	PNSR	LPIPS	PNSR	LPIPS
μ=∞	26.8423	0.1088	26.0485	0.1248	25.3455	0.1486	24.6321	0.1406
μ=0.2	20.3345	0.1536	20.1365	0.1638	19.8263	0.2048	19.6425	0.1836
μ=0.5	27.3693	**0.1052**	26.3485	**0.1204**	25.8454	**0.1348**	25.0753	**0.1248**
μ=1	**27.4425**	0.1082	**26.5221**	0.1229	**25.9645**	0.1458	**25.1965**	0.1369
μ=5	27.3645	0.1093	26.4152	0.1234	25.8265	0.1486	25.0425	0.1385
μ=10	27.1399	0.1082	26.2845	0.1235	25.7185	0.1473	24.8694	0.1395
μ=20	27.1554	0.1093	26.3644	0.1254	25.6578	0.1493	24.8642	0.1423

3.3.3 Optimization of 3D sculpture modeling reconstruction based on point cloud data

Table 3 shows the coordinate transformation parameters. The Sigma model is used to reconstruct the 3D point cloud continuously from the sparse images of the sculpture points, and the volume measurements of five 3D sculptures are taken as an example, and the automatic measurement of the 3D shape of the sculptures is implemented using the automatic measurement method of the simulation sculpture of the proposed method. When the point cloud data are aligned, the coordinate conversion parameters are shown in Table 2. The point cloud data alignment of the 3D sculpture is realized by the coordinate conversion in Table 2. The point cloud data after coordinate conversion is consistent with the spatial relationship of the actual 3D sculpture. Due to the complex structural characteristics of the experimental target, the number of scans was set to 1-4 times when laser scanning was carried out on the 3D sculpture. A point cloud interval setting of 10 to 20 mm was used to ensure that the smallest detailed features of the 3D sculpture could be completely reflected. The final measurement results are compared with the actual 3D sculpture 3D shape values, and compared with the traditional 3D sculpture 3D shape automatic measurement methods. Comparing the error between the actual coordinate points and the coordinate points in the point cloud data, the error range is in [0.0001,0.0083], the error value of the two sets of data is small, and the accuracy of the 3D point cloud continuous reconstruction of the sculpture point sparse image is better.

Table 3: Coordinate transformation parameters

Project	Point number	X	Y	Z
Actual coordinate points	Point 1	201.2242	878.7254	578.9654
	Point 2	252.7584	871.2235	554.0632
	Point 3	267.1867	875.1364	555.1238
In point cloud data The coordinate point	Point 1	201.2241	878.7255	578.9655
	Point 2	252.7583	871.2318	554.0628
	Point 3	267.1868	875.1348	555.1243

The proposed Sigma model for 3D point cloud continuous reconstruction of sparse images of sculpture points is used for automatic measurement of 3D shape of simulation sculpture, traditional 3D shape measurement method 1 for simulation sculpture, traditional 3D shape measurement method 2 for simulation sculpture, respectively, and the comparison results of the measurement accuracy obtained are shown in Fig. 3.

For the five 3D sculptures, using the traditional measurement method 1, the obtained measurement error range is 1.4863~3.9865 m³, and the sculpture with the largest measurement error value is numbered 1. Using the traditional measurement method 2, the obtained measurement error value range is 1.2345~3.1235 m³, and the sculpture with the largest measurement error value is numbered 1. Using the proposed Sigma model for the sculpture point sparse Using the proposed Sigma model for the sculpture point sparse image 3D point cloud continuous reconstruction simulation sculpture 3D shape automatic measurement method, the resulting measurement error value range of 0 ~ 0.0001m³, the largest measurement error sculpture number 2, the rest of the sculpture does not exist measurement error. Through comparison, it is found that the proposed automatic measurement method of 3D shape of simulation sculpture with Sigma model adopts multiple measurement stations for splicing processing, and further processes the 3D shape of simulation sculpture through denoising and streamlining, so that the accuracy of the measurement results is higher.

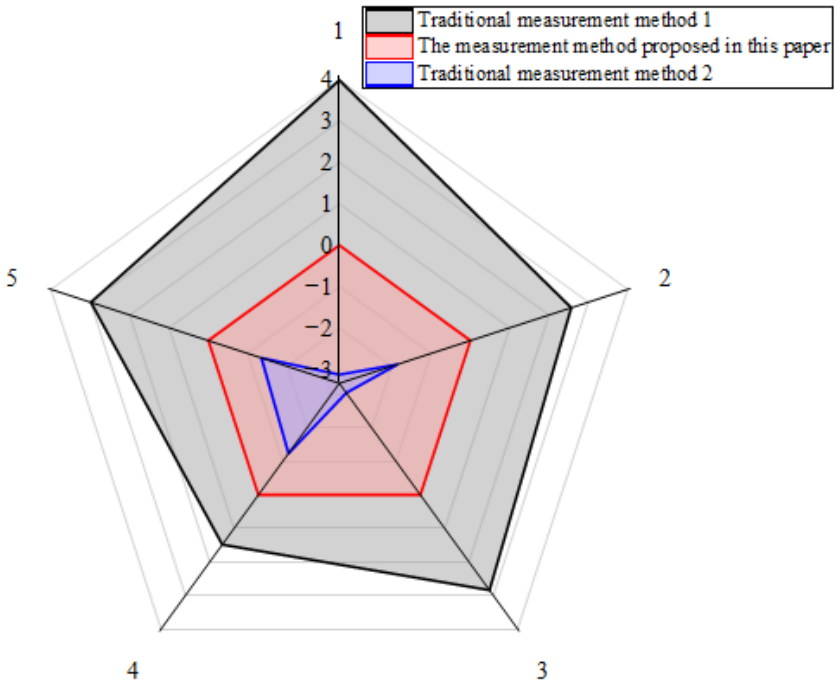

Figure 3: Comparison results of measurement accuracy

4 3D sculpture modeling optimization results analysis

4.1 SBE-based 3D sculpture analysis

In this chapter, based on the beauty evaluation method (SBE), the optimization effect of 3D sculpture styling in the research city is further analyzed and evaluated. It compares the aesthetic preferences and needs of different groups, and analyzes the characteristics of 3D sculptures with higher scores. The SBE model is derived, and the factors affecting the evaluation of 3D sculpture's SBE are ranked, so as to make suggestions for the next application of 3D sculpture.

According to the traditional standardized processing method of scoring values, the SBE standard values of 20 3D sculptures optimized by 3D-GAN-driven modeling are obtained, as shown in Table 4. The SBE value of the 3D sculpture with the first score is 0.88585, and the SBE value of the 3D sculpture with the lowest score is the smallest -0.79545. Among the 20

3D sculptures with SBE values, there are 10 3D sculptures with positive values, accounting for 50% of the total number. Overall the evaluation group was more positive and satisfied with the 3D sculptures. The highest SBE value in the professional group was 0.9154 and the lowest value was -0.8124. The highest SBE value in the non-professional group was 0.8563 and the lowest value was -0.7785.

Table 4: SBE values of 3D sculpture

SBE value sorting	Total value	Professional	Non-professional	Three-dimensional sculpture serial number
1	0.88585	0.9154	0.8563	5
2	0.60745	0.6024	0.6125	6
3	0.3435	0.4825	0.2045	3
4	0.3316	0.1136	0.5496	8
5	0.28055	0.1348	0.4263	2
6	0.24585	0.2369	0.2548	4
7	0.17455	0.2855	0.0636	18
8	0.1741	0.2354	0.1128	4
9	0.1582	0.2039	0.1125	10
10	0.08135	0.0142	0.1485	11
11	-0.05145	-0.0493	-0.0536	13
12	0.05245	-0.0966	0.2015	12
13	-0.12435	0.0458	-0.2945	15
14	-0.20335	-0.4525	0.0458	14
15	-0.31555	-0.2785	-0.3526	1
16	-0.33805	-0.2636	-0.4125	19
17	-0.39565	-0.4255	-0.3658	20
18	-0.41895	-0.3954	-0.4425	7
19	-0.49605	-0.5136	-0.4785	9
20	-0.79545	-0.8124	-0.7785	16

4.2 SBE evaluation and group correlation and variability

To determine whether the survey respondents of different genders have consistency in judging the beauty of 3D sculpture, SPSS software was used to verify. Figure 4 shows the SBE standardized value folds for different genders and groups, plotting the different genders, different groups and total SBE value scores into three folds, and the changes of the folds in the figure tend to be consistent.

For gender groups, the highest scoring 3D sculpture #8 corresponds to the highest scores for both genders, with a score of 0.65187 for men and 0.97851 for women.The lowest scoring 3D sculpture #14 corresponds to the lowest scores for women and lower scores for men, with a score of -0.62234 for men and -0.82144 for women.The ratings of 3D sculptures between male and female genders also There are differences.

For the different groups, the number of ratings for 3D sculptures #1, #2, #5, #8, #11, #14, #16, #17, #18, and #20 were all relatively close (score differences less than 0.1). For the 3D sculptures with the highest and lowest total scores, both correspond to the highest and lowest scores for the professional and non-professional groups. The total scores for the highest and lowest scores are 0.89576 and -0.77166, respectively.The scoring for the professional group is more rigorous in terms of considering the 3D sculptures as a whole.

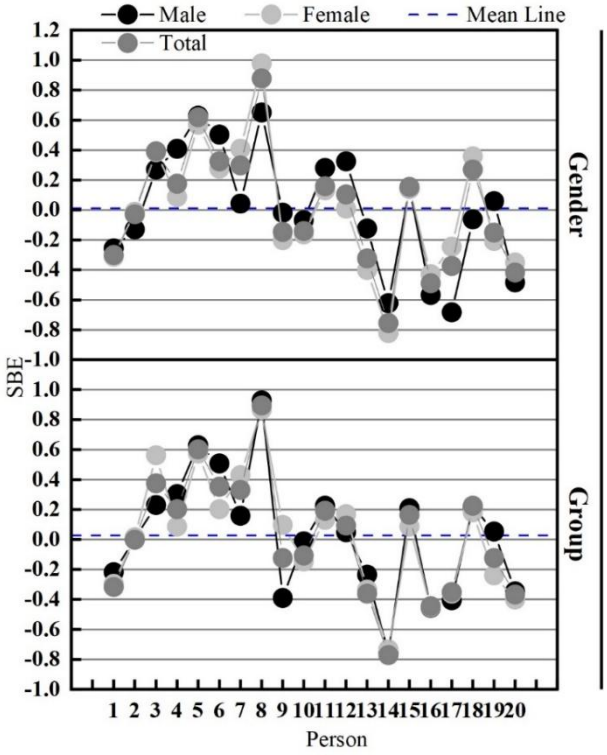

Figure 4: Different gender and group SBE standard value lines

5 Conclusion

In this paper, the conditional generative adversarial network is used as the basis for adding a denoised diffusion model to realize the reversal of the Gaussian diffusion process and recover the target image. The 3D shapes are represented in the stereo space using a body convolution based GAN to generate 3D shapes from the latent space. The Sigma model is used to reconstruct the 3D point cloud continuously from the sparse image of the sculpture points to optimize the shape design of the 3D sculpture.

The generation results of 3D sculpture modeling are judged by the evaluation indexes such as peak signal-to-noise ratio (PSNR), and according to the results of the SSIM calculations, the SSIM values and their average values of the training models in all stages reach more than 0.8, except for the SSIM2 value of 0.7556 for iteration 500steps. In the analysis of PSNR values, when μ is 1, the PNSR has the best values on each training set, which are 27.4425, 26.5221, 25.9645, and 25.1965, respectively.

Based on the beauty evaluation method, the optimization effect of 3D sculpture modeling is further analyzed and evaluated, among the SBE values of 20 3D sculpture models, there are 10 3D sculptures with positive values, accounting for 50% of the total. Overall the group is more satisfied with the evaluation with the 3D sculpture, the highest SBE value in the professional group is 0.9154, and the lowest value is -0.8124. The professional group has more considerations on the 3D sculpture as a whole, and the rating is more strict.

About the Author

Wanglong Yu was born in Shanghai, China, in 2004. I am currently studying at the Academy of Fine Arts, Shanghai University. My main research direction is Modern Art and

Experimental Art.

References

[1] Gumulcine, A., & Coskun, N. (2019). Model design study with three dimensional modeling in fine arts education. In EDULEARN19 Proceedings (pp. 7847-7850). IATED.

[2] Wan, W. (2025). Intelligent pattern design using 3D modelling technology for urban sculpture designing. Systems and Soft Computing, 7, 200176.

[3] Sun, X., Liu, X., Yang, X., & Song, B. (2021). Computer-aided three-dimensional ceramic product design. Computer-Aided Design and Applications, 19(S3), 97-107.

[4] Liu, Y. (2024). Digital interactive design of art sculpture decoration based on augmented reality technology. International Journal of Art Innovation and Development, 5(1).

[5] Guo, S., & Wang, B. (2021). Application of computer aided modeling design in the expression techniques of sculpture art space. Computer-Aided Design and Applications, 19(S3), 1-12.

[6] Liu, L., & Li, X. (2024). Application of Relying on 3D Modeling Technology in the Design of Urban Sculpture. Journal of electrical systems, 20(1).

[7] Zhang, K. (2023, May). Application of 3D Modeling Technology in Sculpture Design. In The World Conference on Intelligent and 3D Technologies (pp. 221-230). Singapore: Springer Nature Singapore.

[8] Xu, X. (2024). Based on 3D Virtual Reconstruction of Modern City Landscape Sculpture Planning Design. EAI Endorsed Transactions on the Energy Web, 11(1).

[9] Ghoneim, H., El-karanfeily, A., & Abdoh, S. A. (2025). Digital Modeling and Its Role in Designing and Applying 3D Functional sculpture models. International Design Journal, 15(2), 55-63.

[10] Yang, Z. (2022). Application and development of digital enhancement of traditional sculpture art. Scientific Programming, 2022(1), 9095577.

[11] Steinfeld, K., Tebbecke, T., Grigoriadis, G., & Zhou, D. (2022, September). Artificiale Rilievo GAN-generated architectural sculptural relief. In Design Modelling Symposium Berlin (pp. 133-148). Cham: Springer International Publishing.

[12] Nelson, P., Mai, J., & Au, R. (2025). The 3D tree dataset: an artistic experiment using a voxel-based GAN. Multimedia Tools and Applications, 84(24), 28519-28533.

[13] Hertzmann, A. (2020). Visual indeterminacy in GAN art. In ACM SIGGRAPH 2020 Art Gallery (pp. 424-428).

[14] Shahriar, S. (2022). GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network. Displays, 73, 102237.

[15] Fang, Y., Ismail, I., & Hadi, H. A. (2025). Digital Restoration of Sculpture Color and Texture Using an Improved DCGAN with Dual Attention Mechanism. Applied Sciences, 15(17), 9346.

[16] Fang, C. (2025). AI-driven digital sculpture design: optimising fusion algorithms with deep learning and virtual reality. International Journal of Information and Communication Technology, 26(22), 55-71.

[17] Kalita, S. S., Mahajan, P., & Sharanya, S. (2024, April). Generative Adversarial Network Art Generator for Sculpture Analysis. In 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT) (pp. 1-6). IEEE.