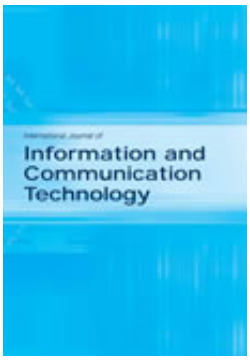
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Appearance design of art exhibits combined with computer vision rendering technology

Yujun Peng

DOI: 10.1504/IJICT.2025.10071755

Article History:

Received:	11 November 2024
Last revised:	24 March 2025
Accepted:	24 March 2025
Published online:	25 June 2025

Appearance design of art exhibits combined with computer vision rendering technology

Yujun Peng

School of Fine Arts and Design,
Nanchang Institute of Technology, China
Email: 13607082825@163.com

Abstract: In order to improve the appearance design effect of art exhibits, this paper uses the improved StyleGAN architecture to generate high-fidelity three-dimensional objects. This paper maps the image to the feature space through VAE, and then reconstructs the three-dimensional properties of the shape and surface from the well-decoupled latent vectors through the improved StyleGAN architecture. The visual-audio-to tactile (VA2T) algorithm directly generates tangential friction force and normal force data in the time domain, serving multi-dimensional data-driven tactile rendering. Combined with the experimental analysis, the VA2T algorithm based on time series force tactile data proposed has certain effects. In addition, combined with experimental analysis, it can be seen that the model proposed in this paper has a certain effect in the design of art exhibits, which can effectively improve the design effect of art exhibits and enhance the actual experience of visitors.

Keywords: computer; visual rendering; art; exhibits; appearance design.

Reference to this paper should be made as follows: Peng, Y. (2025) 'Appearance design of art exhibits combined with computer vision rendering technology', *Int. J. Information and Communication Technology*, Vol. 26, No. 22, pp.1–22.

Biographical notes: Yujun Peng is a Senior Engineer and has a Master's degree. He has long been engaged in teaching and research in art design, presiding over ten projects including MOE Industry-University Collaborative Education Project and Jiangxi Provincial Social Science Planning Project. He has published four textbooks such as *Product Design Modeling and Rendering*, and over 20 high-quality papers. His personal design works have won over 20 awards, and he has guided students to win more than 10 national-level awards. He has made notable academic influence in the fields of digital art and intangible cultural heritage innovation.

1 Introduction

Due to the continuous update of technology and concepts in display design, on the one hand, people have gradually adapted and demanded that the ways of receiving knowledge and information can be diversified and diversified. As the role of information transmission designer, display design grasps and leads the latest presentation and transmission methods, and keeps pace with the times and brings forth the new, which is the fundamental requirement and task entrusted to every practitioner by the times and the

industry. The presentation of exhibits and display methods is two important elements in display design, and it is also two key nodes in the development and application of related technologies, which has led to the attention and research on virtual exhibits and interactive displays. On the other hand, as the core carrier of exhibition content, the value of exhibits determines the quality of exhibition design to a considerable extent. Due to the development and application of virtual interaction technology, visitors' experience expectations for exhibition design are constantly improving, and endless new problems and challenges emerge from time to time. In addition, the value of exhibits plays a vital role in facing problems and challenges, and the extension of the value of exhibits benefits from the development of technology. It also has a broader extension space (Yan et al., 2022).

With the rapid development of digital technology, virtual exhibits and interactive display have become an important part of modern exhibition design. These technologies can not only provide a more immersive exhibition experience for the audience, but also increase the accessibility and malleability of the exhibition. Combining virtual exhibits and interactive displays, it presents diversified display methods, and at the same time displays multiculturalism and works of art in different historical periods, so as to help visitors better understand and appreciate cultural diversity (Mohanto et al., 2022).

In order to improve the appearance design effect of art exhibits, this paper uses the improved StyleGAN architecture to generate high-fidelity three-dimensional objects. This paper proposes an algorithm [visual-audio-to tactile (VA2T)] that combines visual images and audio to generate time-series force tactile data. The VA2T algorithm directly generates tangential friction force and normal force data in the time domain, serving multi-dimensional data-driven tactile rendering.

2 Related works

2.1 Display of art exhibits

At present, there are not only neuroscientists who provide research help in body and brain science, but also designers and artists who explore tactile experience works. Walton et al. (2021) put forward that 'touch is not only done by hands', and adopted a broader perspective to discuss touch-it, and believed that touch includes the ontological and interoceptive experience of the whole body. Kuang et al. (2022) took real plants as exhibits in the exhibition space, constructed a space close to the natural environment, allowed the audience to personally contact real plants, and carried out popular science education on plants and ecology in a large natural green space. It not only provides people with natural experience, but also meets people's learning and social needs. Therefore, the design concept is very pioneering.

Sitzmann et al. (2021) summarised the information of common exhibits in museums, and put forward some thoughts: how to really make the exhibits 'alive'? The exhibits are the 'hearts' of the exhibition planners, and the exhibits themselves are telling their own stories. Scalera et al. (2021) proposed that people can experience the information transmission brought by touch through hands and feet, skin, behavioural movement and emotional communication, and emphasised that tactile experience in space is an excellent design way.

To sum up, the research on the experience design of touchable exhibits in the exhibition space is more systematic and scientific, and there are many cases in practical application, which are quite effective. In the follow-up research and analysis, this paper focuses on the necessity and design methods of touchable experience of exhibits display and improving the visiting effect of exhibits through visual rendering.

2.2 Planar tactile reproduction rendering method

Tactile reproduction rendering method is one of the important research contents of tactile reproduction technology, which is related to whether the tactile reproduction device can present real tactile feedback. The tactile reproduction technology based on friction control can better express the tactile information of virtual objects. There are two existing tactile reproduction technologies based on friction control: one is to use ultrasonic vibration to reduce friction, and the second is to use electrostatic force to increase friction. The two technologies have something in common in tactile reproduction and rendering. Moreover, both of them control the application of driving signals to generate corresponding tactile perception and transmit tactile information of real objects (Wang, 2021).

The tactile reproduction technology applied to mobile terminal uses touch screen to transmit tactile information feedback, which belongs to planar tactile reproduction rendering. Planar tactile rendering methods mainly include the method based on image features and the method based on actual measurement. By modelling the microscopic geometric features of the image texture surface or extracting some feature information of the image, the method based on image features constructs the mapping relationship with tactile feedback, and realises the tactile reproduction of virtual graphic images. By measuring the tactile-related information data generated in the actual interaction with real objects, the method based on actual measurement extracts tactile feature information and processes it, and establishes a tactile information mapping model to generate tactile feedback (Sun et al., 2021).

By using different frequency components of the image to represent different texture feature information such as roughness, depression degree and contour, Lattas et al. (2021) used local Fourier transform method to extract image texture feature information and map it with driving signal amplitude. Fan and Li (2020) proposed a rendering method to improve 3D shape recognition. This rendering method improves the problem that the previous gradient algorithm cannot present a sharp touch to the edge, and adds an edge detection algorithm to render the edge, which improves the performance of 3D shape recognition. Zhao et al. (2022) proposed a method of jointly rendering the texture information of the image by using the amplitude and frequency of the electrostatic force driving signal. Firstly, by studying the relationship between tactile perception and driving signal, the mapping model of image texture gradient obtained by Roberts filter is established. Then, the amplitude of driving signal is used to render the height information of image texture, and the frequency of driving signal is used to render the hardness and granularity information of image texture.

Shan and Sun (2021) proposed a new texture rendering method. This method uses photometric stereo to capture the optical density of the texture and map it with the height information, which can improve the resolution of the model, and then calculate the corresponding normal force and tangential force from the height information. Meanwhile,

a considerable sense of realism can be achieved when rendering textured objects with high compliance or low friction. Based on a haptic display coupled with electrical and mechanical vibration stimuli, Guo and Wang (2021) proposed a new three-dimensional geometric bump haptic rendering algorithm that can simultaneously generate lateral, the mapping relationship between friction force and 3D bump gradient is established.

Lombardi et al. (2021) proposed to 'record' texture, and used texture acquisition tools to obtain speed and vibration data during contact, so as to construct a force tactile model of texture to reproduce texture tactile information. This algorithm is relatively simple and easy to reproduce uniform texture materials. Li et al. (2020) built a tactile acquisition system that can collect acceleration, movement position and contact force data. After the original time domain tactile signal is converted into frequency domain texture vibration signal by linear predictive coding method, the texture tactile sensation can be reproduced by voice coil motor actuator.

Fu et al. (2021) built a tactile information acquisition equipment, which includes slide rail, motion controller, accelerometer, etc., and can collect acceleration data under different interactive conditions. Neural network is used to process the acceleration data, and bilinear interpolation is used to calculate and synthesise the corresponding acceleration signals in real-time. Turchet (2023) proposed a data-driven texture rendering method for electrostatic force, which obtains and processes the acceleration data when the probe slides on the sample surface, and maps it with the voltage amplitude of the electrostatic force driving signal to reproduce a virtual texture with similar touch.

In this paper, a design system of art exhibits with strong visual infection effect and tactile feeling is proposed by combining visual rendering technology and time sequence force tactile data processing algorithm, which can improve the design effect and display effect of art exhibits.

3 Algorithm model construction

3.1 VA2T algorithm for time sequence force tactile data

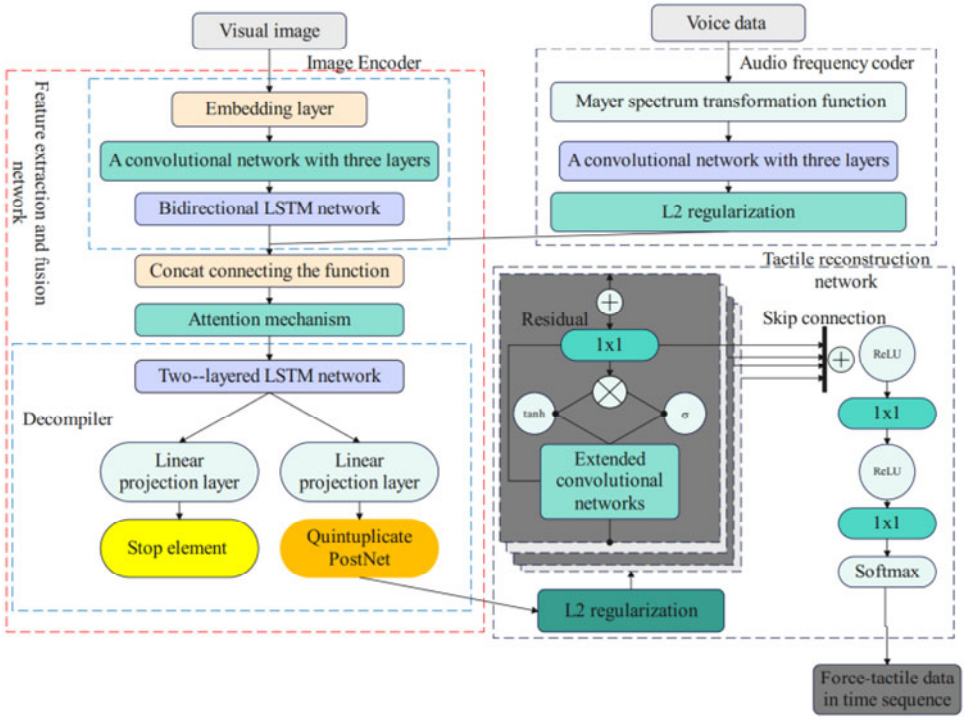
The VA2T network is shown in Figure 1, which consists of an audio encoder, a feature extraction and fusion network, and a haptic reconstruction network. The audio encoder extracts fixed-dimensional feature vectors from the audio input through the LSTM network. The image encoder in the feature extraction and fusion network captures the feature vectors in the image, connects the audio feature vectors and the image feature vectors, and forms a feature map that fuses the image and audio through the decoder module. Finally, the tactile reconstruction network transforms this feature map into friction and normal force data in the time domain.

3.1.1 Audio encoder

The function of the audio encoder is to extract the embedding vector of the input audio data, and represent the input features of the audio data through the embedding vector. The audio encoder first converts the audio input into a logarithmic Mel spectrum, denoted as (Cetinic and She, 2022):

$$S = Spectrogram\left(A_m^n\right) \tag{1}$$

Figure 1 VA2T network block diagram (see online version for colours)



Among them, A_m^n represents the n^{th} audio sample of the m^{th} material, S represents the logarithmic Mel spectrum, and *Spectrogram* represents the Mel spectrum conversion function. The resulting spectrum S is passed through a three-layer LSTM network with hundreds of units in each layer and then projected into a 256-dimensional space. The output of the final layer is L2-normalised to create the final embedding vector.

$$J_i = LSTM\left(S, J_{i-1}\right) \tag{2}$$

$$E = L2Normalise\left(Project(J3)\right) \tag{3}$$

Among them, J_i is the output vector of the i^{th} layer LSTM network, J_{i-1} is the output vector of the $i - 1^{\text{th}}$ layer LSTM network. *L2Normalise* means normalising the length (or norm) of the vector to 1 and keeping the direction of the vector unchanged. $J3$ is the output vector of the third layer LSTM network, and Project projects the output vector of LSTM into 256-dimensional space, and E is the final embedded vector.

The feature extraction and fusion network includes a visual encoder, an attention module, and a decoder. The input image is first transformed into an embedding vector through a 512-dimensional embedding layer,

$$X_e = Embedding\left(I_m^n\right) \tag{4}$$

Among them, I_m^n represents the n^{th} image sample of the m^{th} material, X_e represents the embedding vector output by the embedding layer, and Embedding represents the

embedding layer. The obtained embedding vector X_e is passed through a stacked module containing three convolutional layers to extract features layer by layer. Each convolutional layer contains 512 convolution operators with a dimension of 51, and each layer is connected to a BN layer and a ReLU, which is expressed as:

$$X_{conv} = Conv(X_e) \tag{5}$$

Among them, X_{conv} is the output of the three-layer convolutional neural network, and Conv stands for the convolutional layer. The processed vector X_{conv} passes through a bidirectional LSTM layer, and the resulting encoded feature is represented as:

$$X_{biLSTM} = BiLSTM(X_{conv}) \tag{6}$$

Among them, X_{biLSTM} is the output of the bidirectional LSTM layer, and X_{conv} represents the bidirectional LSTM network.

The encoded features X_{biLSTM} are concatenated with an embedding vector E generated by the audio encoder, and then a context vector is generated by the attention module, which encapsulates the entire encoded sequence and affects each step of the output of the decoder:

$$C_t = Attention\left(Concat\left(X_{biLSTM}, E\right), D_{t-1}\right) \tag{7}$$

Among them, C_t is the context vector at the moment t and D_{t-1} is the decoder state at the previous moment. Attention stands for Attention mechanism, which focuses on more important features by setting the weights of different regions. The context vector C_t is then fed into a decoder, which includes a two-layer LSTM network, a linear projection layer and a post-processing network (PostNet). The process is expressed as:

$$D_t = LSTM\left(C_t, D_{t-1}\right) \tag{8}$$

Among them, D_t is the LSTM network output at the current moment, and the linear projection layer is next used to convert D_t from high-dimensional space to low-dimensional space, which is denoted as:

$$L_t = LinearProjection\left(D_t\right) \tag{9}$$

Among them, L_t is the low-dimensional feature output by the linear projection layer, and *LinearProjection* represents the linear projection layer. PostNet is used to further process the low-dimensional features, which is expressed as:

$$M_t = PostNet\left(L_t\right) \tag{10}$$

Among them, M_t is the feature refined by PostNet. PostNet stands for post-processing network. Each layer of PostNet contains 512 convolution operators with a dimension of 5 × 1. Except for the last layer, all layers are batch normalised and activated by tanh function. The calculation of tanh activation function is:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{11}$$

The haptic reconstruction network is responsible for generating haptic data in the time domain. The network has autoregressive nature, and the generation of each haptic data

sample depends on all previous samples, forming an autoregressive process. The input features M_t are processed by an extended causal convolutional layer and are expressed as:

$$Y_t = CausalConv\left(M_t, Y_{t-1}\right) \tag{12}$$

$$Z_t = DilatedConv\left(Y_t, d\right) \tag{13}$$

Among them, CausalConv represents causal convolution, DilatedConv represents dilated convolution, and the superposition of the two is called dilated causal convolution. Dilated causal convolution, as the core component of the tactile reconstruction network, will be described in detail in the next section. Y_t is the output of the causal convolution layer at time t, Y_{t-1} is the output of the previous time, Z_t is the output of the dilated convolution layer at time t, and d is the dilation rate. In the network, the dilated convolution layers are arranged according to the exponentially increasing dilation rate.

The last extended convolutional layer uses two extended convolutions to generate frictional force and normal force features respectively, the output of this layer contains two channels, and a gated activation unit is set to process the output of the extended convolutional layer to enhance the representation capability of the network (Ettalibi et al., 2024).

$$G_t = \tanh\left(W_f * Y_f\right) \odot \sigma\left(W_g * Y_t\right) \tag{14}$$

Among them, W_f and W_g are the weights of the convolution layer, * represents the convolution operation, $\odot$ represents the element-wise multiplication, $\tanh(\cdot)$ represents the tanh activation function, and $\sigma(\cdot)$ represents the Sigmoid function. The output G_t then passes through a Softmax layer to obtain the probability distribution of the next tactile sample value:

$$O_t = Softmax\left(G_t\right) \tag{15}$$

According to the probability distribution O_t of the output, tangential friction force and normal force data are generated in two channels by random sampling:

$$x_{t+1} = Sample\left(O_t\right) \tag{16}$$

Among them, x_{t+1} is the force tactile data sample generated at the moment $t + 1$. The generated samples are then used as inputs at the next moment, thereby iteratively constructing complete time-domain force haptic data. This autoregressive method can predict the next sample based on the previously generated tactile data samples, and step by step construct the whole time domain tactile data. Each prediction relies on the history of all past samples, ensuring the temporality of the generated force tactile data.

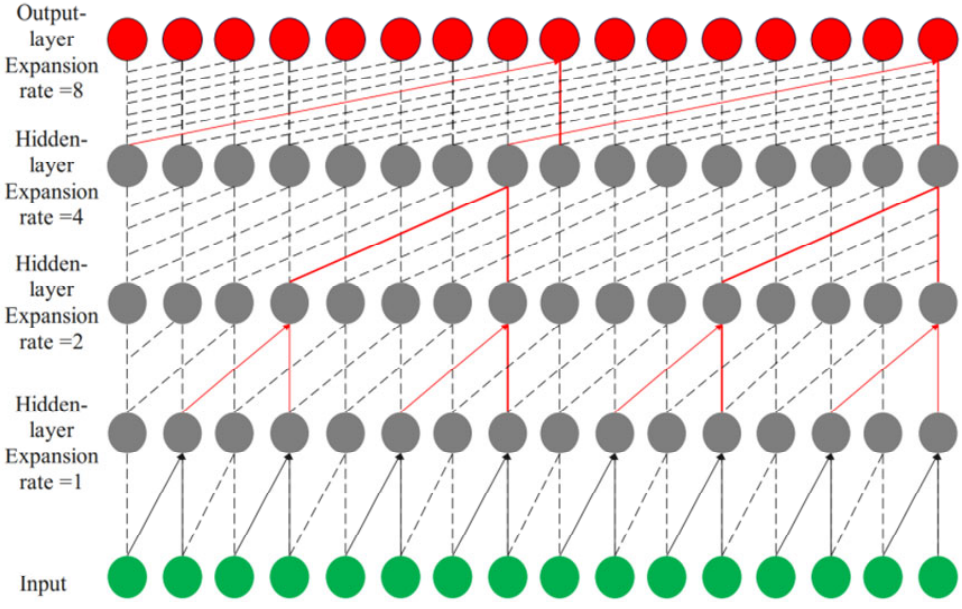
For force tactile data $x_1, x_2, \ldots, x_T$, each data sample is conditioned based on all previous samples, and the VA2T network decomposes the joint probability of the tactile signal into a series of conditional probabilities, which is expressed as:

$$p(x) = \prod_{t=1}^{T} p\left(x_t \middle| x_1, \ldots, x_{t-1}\right) \tag{17}$$

Extended causal convolution is an advanced convolution technology, which adds expansibility to the traditional causal convolution, and its structure is shown in Figure 2. This convolution method has a broad acceptance domain, can effectively maintain and

identify remote dependencies in the data, and guarantee the timing of haptic data in the whole modelling process.

Figure 2 Schematic diagram of extended causal convolution structure (see online version for colours)



This paper assumes that the input sequence is $x_1, x_2, \ldots, x_T$, where T is the length of the sequence, and the output sequence is $y_1, y_2, \ldots, y_T$. The output y_t of the causal convolution at time t can be expressed as:

$$y_t = \sum_{k=0}^{K-1} f_k \cdot x_{t-k}, t = 1, 2, \ldots, T \tag{18}$$

Among them, y_t is the output at time t, f_k is the value of the convolution kernel at position k, causal convolution is ideal for dealing with long time series and is very fast to train due to the absence of cyclic connections. However, it requires increasing the receiving domain by setting many layers or using a large number of convolution operators, which will increase the computational cost of the network.

The dilated convolution is implemented by adding a margin to the stride of the standard convolution operator, which effectively allows the network to operate at a coarser scale than normal convolution. The output $y_t^{'}$ of the dilated causal convolution at the t^{th} time step is expressed as:

$$y_t^{'} = \sum_{k=0}^{K-1} f_k \cdot x_{t-d.k}, t = 1, 2, \ldots, T \tag{19}$$

Among them, $y_t^{'}$ is the output at time t, f_k is the value of the convolution kernel at position k, d is the dilation rate, and $x_{t-d,k}$ is the value of the input sequence at time $t - d$, k. The dilation rate u controls the spacing of the elements in the convolution kernel. To

maintain causality, $x_{t-d,k}$ is set to 0 when $t - d$, $k - 1$. In this way, the dilated causal convolution allows each output y_t' to consider more distant previous information in the input sequence, rather than just the information of the next few time steps.

An L2 loss function as shown below is used during training:

$$L\left(f, \hat{f}\right) = \sum_{i=1}^{N} \left(f_i, \hat{f}_i\right)^2 \tag{20}$$

Among them, f is the measured force tactile data, $\hat{f}$ is the generated force tactile data, and N is the number of samples. Due to the existence of the square term, it is more sensitive to outliers and can therefore more effectively penalise larger errors. Then, the optimal parameter set Θ is obtained by minimising the loss function.

3.2 Visual rendering algorithm

The differential shader R links the 3D attribute space to the 2D image space and defines pixel values by interpolating the abstract vertex attributes u_0, u_1, and u_2. Since the renderer expects the input to be a mesh, vertex position is one of these attributes, but a large number of other vertex attributes are supported at the same time, drawing images directly using vertex colour or textures. To define the base colour of a mesh, supported vertex attributes are vertex colour or coordinate u and v from learned or predefined texture maps. The pixel values are determined by bilinear interpolation of vertex colours or projected texture coordinates, respectively. Then, differential rendering is used to render the resulting three-dimensional attributes into two-dimensional images. The image encoder and the three-dimensional attribute generator are optimised by two-dimensional image supervision, and the formula is as follows (Moragane et al., 2024):

$$\theta = \arg\min \frac{1}{N} \sum_{i=1}^{N} Dist\left(R\left(G_\theta\left(E_\theta\left[X_i, M_j\right]\right)\right), X_i\right) \tag{21}$$

Among them, *Dist*() represents the distance between the reconstructed data $X_i^r = [I_i^r, M_i^r] = R(G_\theta(E_\theta[X_i, M_i]))$ and the input data X_i. DIB-R is used as the differential renderer in this chapter.

In order to constrain the above process, a loss function is designed, including reconstruction loss and GAN loss. These losses will be described below.

3.2.1 Reconstruction loss

The reconstruction loss is mainly achieved by calculating pixel level and feature level L1 losses:

$$L_{rgb} = \frac{1}{N} \sum_{i=1}^{N} \left\| X_i \odot M_i - X_i^r \odot M_i^r \right\|_1 \tag{22}$$

$$L_{frat} = \frac{1}{N} \sum_{i=1}^{N} \left\| VGG\left(X_i \odot M_i\right) - VGG\left(X_i^r \odot M_i^r\right) \right\|_1 \tag{23}$$

Among them, ⊙ represents element-by-element multiplication.

Finally, the minimised distance loss for overall two-dimensional spatial supervision is weighted as follows:

$$L_{recon} = \lambda_{rgb} L_{rgb} + \lambda_{feat} L_{feat} \tag{24}$$

λ_{rgb} and λ_{feat} represent weights, and L_{recon} can back-propagate the loss gradient to E_θ and G_θ through the renderer R.

3.2.2 GAN loss

The adversarial loss can be expressed as:

$$L_D = \frac{1}{N}\sum_{n=1}^{n}\left[D(X_i) - D(G)\right] + \lambda \frac{1}{m}\sum_{j=1}^{m}\left(\left\|\nabla x_j^r D\left(x_j^r\right)\right\|_2 - 1\right)^2 \tag{25}$$

The weight t of the GP is the path length regularisation set in consideration of the stability during the training process.

Path length regularisation can be expressed as:

$$L_{pl} = \left(\left\|J_W^T y\right\|_2 - a\right)^2 \tag{26}$$

Among them, w is a latent space point, y is a unit normally distributed random variable in the generated image space (the dimension of RGB images is 3 * w * h), J is the Jacobian matrix, and a is a global value representing the expected gradient scale. When the Jacobian matrix J is orthogonal, the internal expectation is (approximately) minimised.

4 System design and experimental analysis

The related theories such as multi-projection geometry and colour rendering studied in this paper are verified in practical application scenarios. In the third part, the VA2T algorithm and visual rendering algorithm of time-series force tactile data have been combined to construct an intelligent system that can be used for the design of art exhibits. system.

4.1 Immersive multi-projection system

This paper puts forward a multi-projection system solution for art exhibit design, which consists of four layers: application layer, support layer, data layer and hardware layer. As shown in Figure 3.

The application layer mainly realises the functions of system management, interactive function, configuration management, visual simulation, etc. The support layer provides technical support for the operation of the whole system, mainly providing communication services, projector and camera array control support, clock synchronisation, correction and rendering, etc. The data layer provides data support for each layer, mainly including 3D model files, animation and special effects files, configuration files, fusion correction data, etc. The hardware layer provides hardware platform support for system operation, including network switches, system control computers, scene generation computer arrays,

projector arrays, large-scale special-shaped projection screens, industrial camera arrays, interactive devices, etc.

Figure 3 Framework design of art exhibit design system (see online version for colours)

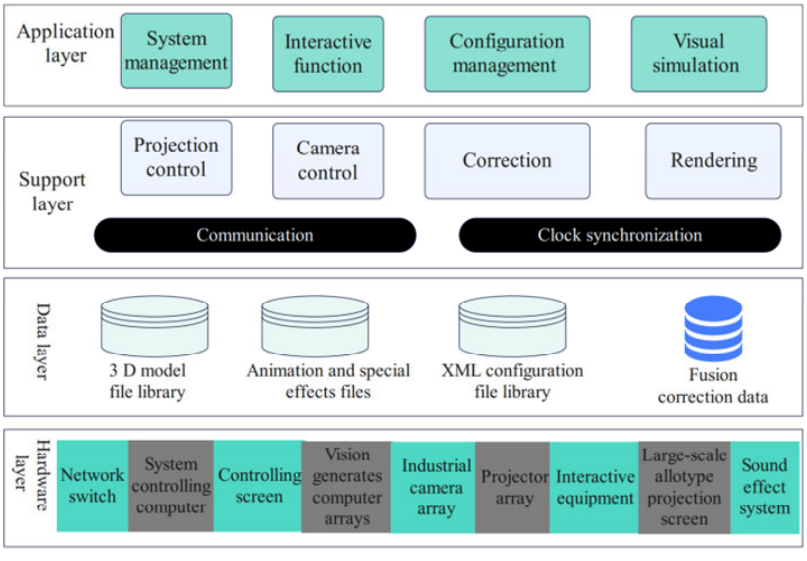
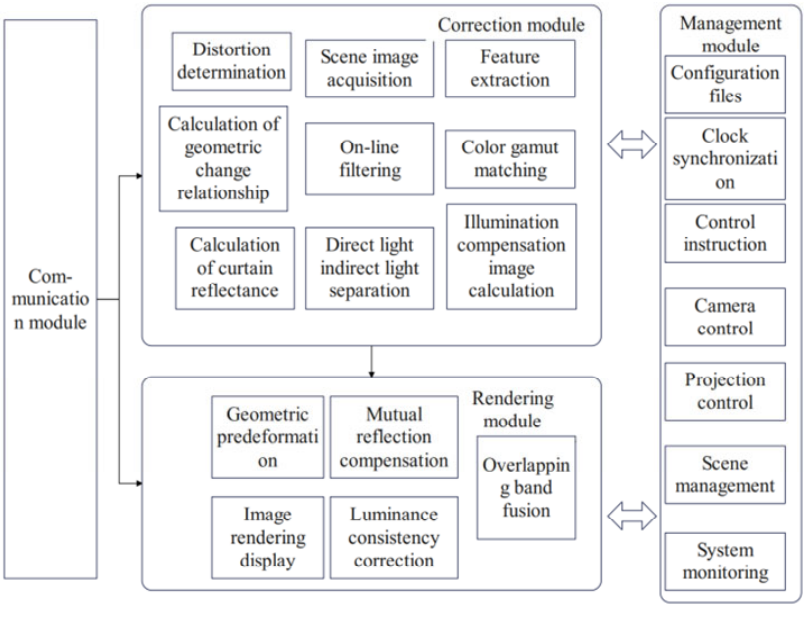


Figure 4 Functional module diagram of art exhibit design system



This immersive multi-projection system mainly designs and develops several main functional modules as shown in Figure 4, including correction module, rendering module, communication module and management module.

All the image correction and fusion work of the system is completed by the correction module, and its main functions include geometric correction, colour adaptive correction and ambient light elimination. The operation flow of the whole calibration module is shown in Figure 5.

Figure 5 Art exhibit appearance design system image correction module

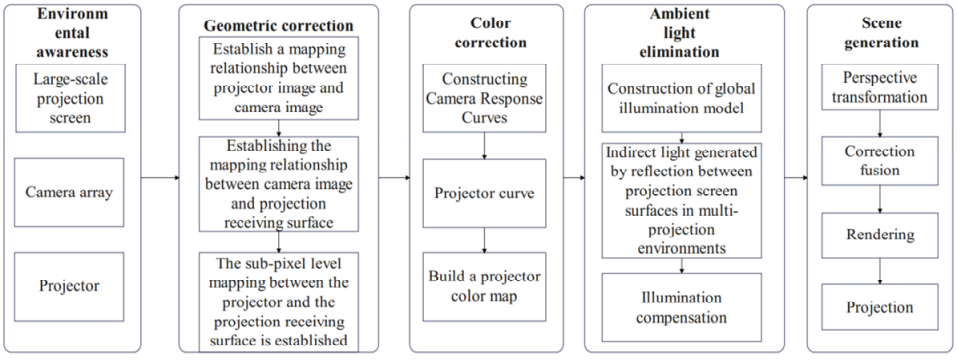
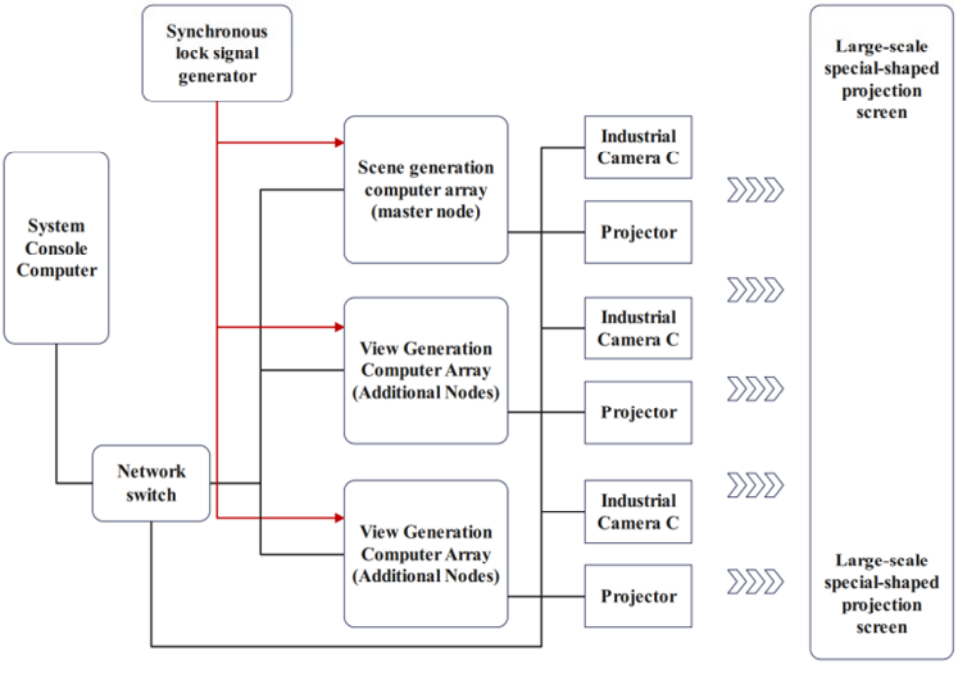


Figure 6 Networking structure of art exhibit display system (see online version for colours)



The correction process of the system consists of five steps: environmental perception, geometric correction, colour correction, ambient light elimination and scene generation. The rendering module is implemented based on the UE5 rendering engine, integrates geometric correction data from the correction module to complete picture splicing, and receives colour correction data and ambient light elimination data to rebalance colour and brightness.

The management module is mainly responsible for the unified control of the whole system, including the on-off of the whole system, the camera array photography control, the projector array projection control, and the synchronisation of image output frames among all projectors. The module designs an easy-to-use human-computer interaction interface, displays the system status, and provides the functions of scene data management and system configuration parameter setting. Its networking structure is shown in Figure 6.

The master node computer in the array of scene generation computers is the scheduler of nDisplay, which manages the timing information across the computer cluster and manages and distributes the copy content of possible rendering roles and data to other computers.

The dataset used in this article is Oxford Art Online, ARTstor Digital Library,JSTOR, According to the random pairing method for model matching, 6500 sets of data were obtained, with 80% of the data used as the training set and 20% as the testing set.

4.2 Experimental methods

The experimental model proposed in this paper can simulate the touchable exhibit simulation structure with certain rendering effect in the experimental terminal. Moreover, the VA2T algorithm combined with time-series force tactile data can realise audio and vision fusion to generate a tactile terminal projection image. In addition, this paper combines the algorithm in Part 3.2 to realise the appearance and colour rendering of art exhibits, and combines the touch to enhance the simulation design effect of art exhibits, so as to further improve the appearance design effect and display effect of art exhibits.

The experiment is carried out on a computer system equipped with an IntelXeonE5-2620 processor and four Nvidia Tesla V100 GPUs and is implemented with the help of the PyTorch framework deployment in a Python environment.

The experimental environmental parameters of this paper are shown in Table 1.

Table 1 Experimental environment parameters

Category	*Configuration*
CPU	Intel Core i9-9900KS
GPU	GeForce RTX4080 SUPER
Memory	16 GB
Render window resolution	1,920 * 1,080
Operating system	Windows11
Development tools	Unity 6
API	DirectX 12

4.3 Results

Mean absolute error (MAE) and mean square error (MSE) are two commonly used loss functions in regression tasks, which affect the optimisation direction and performance of the model through different error calculation methods.

Figure 7 MAE loss curves of tangential friction force and normal force (see online version for colours)

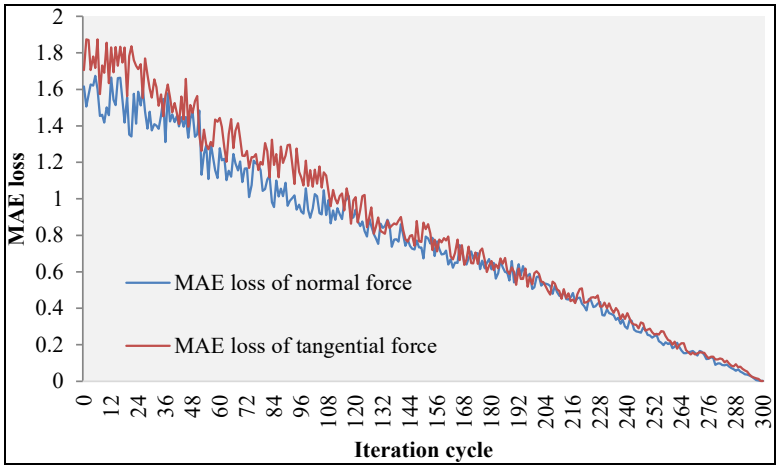
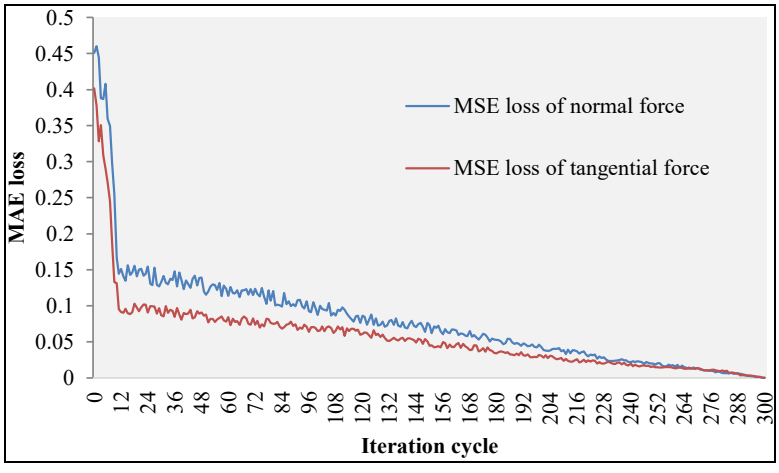


Figure 8 MSE loss curves of tangential friction force and normal force (see online version for colours)



MAE has low sensitivity to outliers and is suitable for scenarios with high data noise or outliers, such as sensor data.

MSE, due to its square operation, significantly amplifies the impact of large errors and focuses more on reducing extreme biases, making it suitable for scenarios that require high accuracy and clean data.

The MAE and MSE loss curves of tangential friction force and normal force are shown in Figures 7 and 8, respectively, and the MAE and MSE of tangential force and normal force converge to 0.001 after 300 iteration periods. The Pearson correlation coefficient curves of tangential friction force and normal forces are shown in Figure 9.

Figure 9 Pearson correlation coefficient curves of tangential friction force and normal force (see online version for colours)

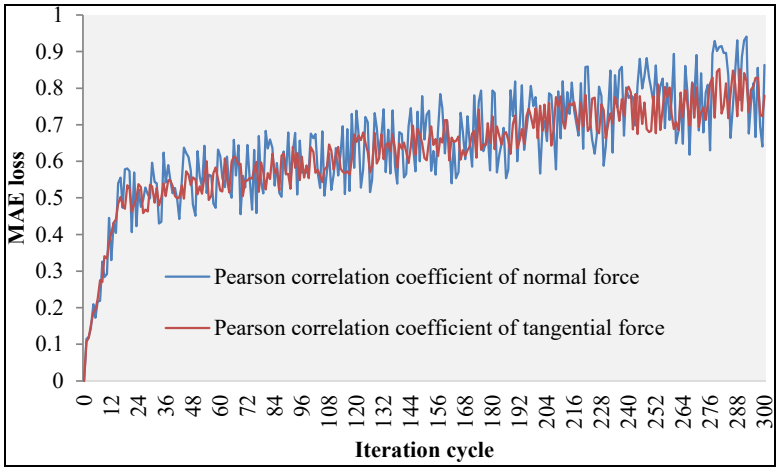
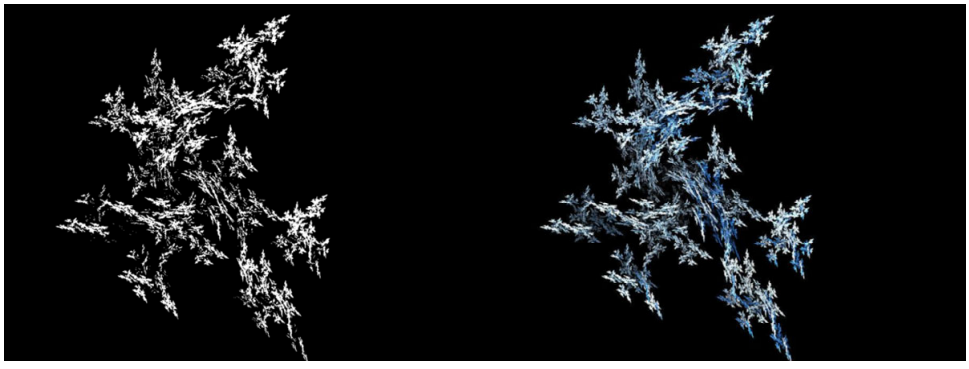


Figure 10 Example of rendered image effect using StyleGAN-VA2T algorithm, (a) before rendering (b) after rendering (see online version for colours)



(a) (b)

Figure 10 shows an example image rendered by the stylegan-va2t algorithm. It can be seen from the figure that the original image is fuzzy and monotonous in colour. After the stylegan-va2t rendering, not only the image definition is improved, but also the colour richness of the image is improved, which effectively improves the artistic effect of the image.

The algorithm proposed in this paper is named StyleGAN-VA2T, and the rendering scene is Sponza. In order to compare this algorithm with mainstream algorithms in many aspects, this paper divides it into static comparison and dynamic comparison in Sponza rendering comparison experiment. Static comparison can more intuitively compare the rendering results of different algorithms, and dynamic comparison can obtain more experimental data, which fully verifies the effectiveness of the algorithm. Then, this paper compares the algorithm proposed in this paper with the algorithms in Wang (2021), Lattas et al. (2021), Guo and Wang (2021) and Li et al. (2020). This paper compares the

appearance of two art exhibits with the static scene reflection rendering frame rates (FTs) under different algorithms, as shown in Table 2.

Table 2 Comparison of FT of static scene reflection rendering under different algorithms

Algorithm	*Rendering FT of art exhibit 1*	*Rendering FT of art exhibit 2*
Wang (2021)	172.43	112.27
Lattas et al. (2021)	151.32	80.66
Guo and Wang (2021)	130.32	61.79
Li et al. (2020)	177.73	116.04
StyleGAN-VA2T	208.02	167.64

Structural Similarity Index Measure (SSIM) is a comprehensive reference evaluation index that measures the similarity between two images. It quantifies the perceived characteristics of brightness, contrast, and structure by simulating the human visual system.

The comparison of SSIM for static scene reflection rendering under different algorithms is shown in Table 3.

Table 3 Comparison of SSIM for static scene reflection rendering under different algorithms

Algorithm	*Art exhibits 1 SSIM*	*Art exhibits 2 SSIM*
Wang (2021)	0.94	0.93
Lattas et al. (2021)	0.95	0.94
Guo and Wang (2021)	0.92	0.91
Li et al. (2020)	0.80	0.76
StyleGAN-VA2T	0.96	0.95

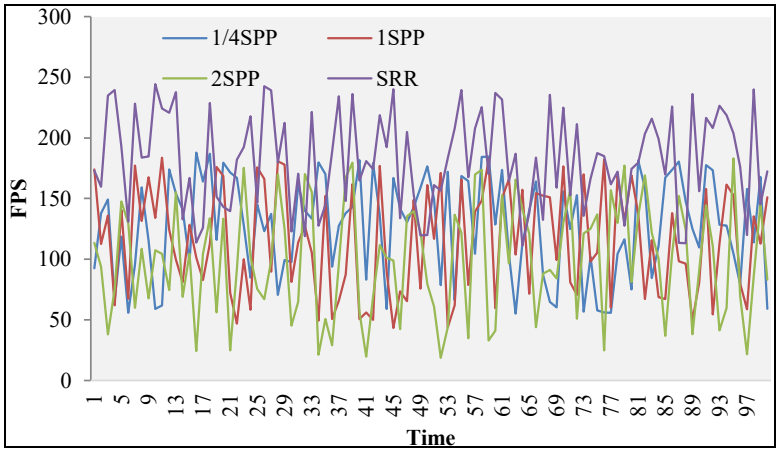
The comparison of MSE for static scene reflection rendering under different algorithms is shown in Table 4.

Table 4 Comparison of MSE for static scene reflection rendering under different algorithms

Algorithm	*Art exhibit 1 MSE*	*Art exhibit 2 MSE*
Wang (2021)	5.01	5.21
Lattas et al. (2021)	4.95	5.16
Guo and Wang (2021)	6.46	6.94
Li et al. (2020)	9.34	12.98
StyleGAN-VA2T	4.90	5.10

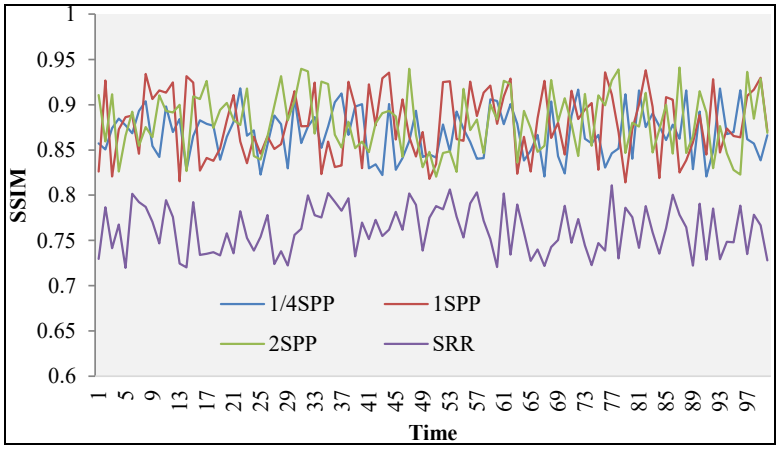
In the comparison experiment of dynamic scene reflection rendering FT, this paper will set a fixed speed to move in the same path in the scene. Then, the recorded FT is drawn into a line chart to compare the FT of dynamic scene reflection rendering. The FT of different algorithms in dynamic scenes are shown in Figure 11. In the figure, the ordinate is the FT, the unit is Hz/S, and the abscissa is the time, the unit is S.

Figure 11 Comparison of FT of dynamic scene reflection rendering under different algorithms (see online version for colours)



Taking the high-precision picture obtained by the ray tracing algorithm under 4096SPP as a reference, the intermediate frame per second in 100 s of dye FT comparison in the dynamic scene is used for comparison, and the SSIM values of each algorithm and the reference picture are compared as shown in Figure 12, in which the ordinate is SSIM, the abscissa is time, and the unit is s.

Figure 12 Comparison of SSIM for dynamic scene reflection rendering under different algorithms (see online version for colours)



4.4 Analysis and discussion

It can be seen from Figures 7 and 8 that the MAE and MSE of tangential force and normal force converge to 0.001 after 300 iteration periods. It is also verified that the VA2T algorithm based on time series force tactile data proposed in this paper has certain effects, and it can realise reliable system functions after many iterative learning.

It can be seen from Figure 9 that the Pearson correlation coefficients of tangential force and normal force fluctuate around 0.8 after 300 periodic iterations, which means that there is a strong correlation between the generated tangential force, normal force and measured force respectively.

- In the comparison of static scene FT in Table 2, art exhibit 1: The model in Lattas et al. (2021) is reduced by 12.24% compared to the model in Wang (2021), the model in Guo and Wang (2021) is reduced by 24.42% compared to the model in Wang (2021), the model in Lattas et al. (2021) is reduced by 2.98% compared to the model in Li et al. (2020), and it is reduced by 17.11% compared to StyleGAN-VA2T. Art exhibit 2: the model in Lattas et al. (2021) is reduced by 28.1% compared with the model in Wang (2021), the model in Guo and Wang (2021) is reduced by 44.96% compared with the model in Wang (2021), the model in Lattas et al. (2021) is reduced by 3.25% compared with the model in Wang (2021), and it is reduced by 33.03% compared with StyleGAN-VA2T.
- In the SSIM comparison of static scenes in Table 3, art exhibit 1: The model in Lattas et al. (2021) is improved by 1.30% compared with the model in Wang (2021), STYLEGAN-VA2T is improved by 2.06% compared with the model in Wang (2021), the model in Lattas et al. (2021) is improved by 1.99% compared with the model in Guo and Wang (2021), and is improved by 17.13% compared with the model in Li et al. (2020). Art exhibit 2: the model in Lattas et al. (2021) is improved by 1.20% compared with the model in Wang (2021), STYLEGAN-VA2T is improved by 2.30% compared with the model in Wang (2021), and the model in Wang (2021) is improved by 2.00% compared with the model in Guo and Wang (2021) and is improved by 23.32% compared with the model in Li et al. (2020).
- In the static scene MSE comparison in Table 4, art exhibit 1: The model in Lattas et al. (2021) is improved by 1.34% compared with the model in Wang (2021), STYLEGAN-VA2T is improved by 2.26% compared with the model in Wang (2021), the model in Wang (2021) is reduced by 22.32% compared with the model in Guo and Wang (2021), and is reduced by 46.28% compared with the model in Li et al. (2020). Art exhibit 2: the model in Wang (2021) is improved by 1.02% compared with the model in Lattas et al. (2021), STYLEGAN-VA2T is improved by 2.20% compared with the model in Wang (2021), and the model in Lattas et al. (2021) is reduced by 24.85% compared with the model in Guo and Wang (2021), and is reduced by 59.83% compared with the model in Li et al. (2020).

In the dynamic scene FT comparison in Figure 11, as the number of samples gradually increases, the STYLEGAN-VA2T dynamic scene FT decreases by several frames. The average FT in 100s under 1/4 SPP (SamplesperPixe) is 19.35 fps, the highest FT is 188.45 fps, and the lowest FT is 189.36 fps. The FT is 50.23 fps. The average FT of 100s under 1SPP is 103.12 fps, the highest FT is 185.25 fps, and the lowest FT is 37.45fs. The average FT at 2SPP is 85.13fps, the highest FT is 183.14fps, and the lowest FT is 16.21 fps. The average 100s FT under SSR (Screen space reflection) is 162.99 fps, the highest FT is 245.49 fps, and the lowest FT is 109.11 fps.

In the dynamic scene SSIM comparison in Figure 12, as the number of samples increases, the STYLEGAN-VA2T dynamic scene SSIM improves very little. The average SSIM in 100s under 1/4 SPP is 0.908, the max value is 0.919, and the min value is 0.817. Under 1SPP, the average SSIM of 100s is 0.9113, the max value is 0.939, and

the min value is 0.812. The average SSIM of 100 s under 2SPP is 0.919, the max is 0.941, and the min is 0.819. The average SSIM of 100s under SSR is 0.769, the max value is 0.811, and the min value is 0.719.

As an improved attention mechanism, VA2T's computational complexity and potential limitations can be inferred based on similar architectures such as group vector attention and separable self attention, and analysed in conjunction with algorithm design principles

- Time complexity: If a divide and conquer strategy similar to group vector attention (GVA) is adopted, VA2T can reduce the complexity of traditional multi head attention from O (k^2) to O (k), specifically by grouping input features and processing them in parallel. For example, MobileViTv2 optimises complexity to linear level through separable self attention. If a dynamic weight allocation mechanism is introduced (such as adjusting the number of attention groups based on environmental feedback), it may increase the cost of dynamic decision-making and lead to fluctuations in complexity.
- Space complexity: Grouping strategy can reduce parameter storage requirements (such as Point Transformer V2 reducing the number of parameters caused by channel growth through grouping), but if multiple sets of historical states or pre trained parameters need to be stored (such as Meta CT's iterative optimisation process), it may increase memory usage.

Based on the mechanism and practical requirements of the algorithm, the limitations are analysed as follows:

1 Insufficient adaptability to dynamic scenes:

In high dynamic environments such as real-time obstacle avoidance and streaming data sorting, the grouping strategy of VA2T may cause response delays due to fixed grouping rules. For example, tree search algorithms can cause overthinking or underestimation issues due to node redundancy and uneven computation allocation.

Some variants rely on static positional encoding (such as the positional encoding multiplier of Point Transformer V2), which limits their generalisation ability to non-uniform distribution data (such as point clouds).

2 Parameter sensitivity and training cost:

Hyperparameters such as the number of groups and attention heads need to be finely tuned, otherwise they may lead to underfitting or overfitting. For example, the accuracy improvement of MobileViTv2 relies on the parameter balance of separable self attention, while the performance of tree search algorithm shows a marginal decrease with the increase of the number of child nodes. If combined with online learning (such as the dynamic reward mechanism of reinforcement learning), frequent updates of the policy network are required, which exacerbates the consumption of training resources.

Combined with the above analysis, it can be seen that the model proposed in this paper has certain effects in the design of art exhibits, which can effectively improve the design effect of art exhibits and enhance the actual experience of visitors.

Sensory design guides audience emotions through environmental sound effects, tactile feedback, and other means, which may lead to controversy over 'perceptual manipulation'. For example, using high-frequency sound waves or specific light frequencies to induce audience behaviour requires defining the boundary between 'reasonable guidance' and 'involuntary intervention'. The promotion of multi sensory exhibition design in public spaces needs to face the dual proposition of technological empowerment and ethical constraints. Through dynamic regulation, inclusive design, and cultural sensitivity prediction, public value can be maintained while stimulating perceptual potential. In the future, it is necessary to continue exploring the symbiotic model between humanism and technological innovation, in order to avoid the 'sensory utopia' from becoming an ethical testing ground.

The VA2T algorithm demonstrates full scene coverage potential in the field of art and design, from individual creation to industrial applications, through its ability to generate generalisation and multimodal fusion. Its scalability is not only reflected in the compatibility of the technical architecture with heterogeneous data (such as vision, language, sensor signals), but also in adapting to rapidly changing industry demands through dynamic learning mechanisms. In the future, combining embodied intelligence and high-precision rendering technology, this system is expected to become the core infrastructure of the digital creative economy The VA2T algorithm is reshaping the technological paradigm in the field of art and design through its dual innovation of generating logic reconstruction (multimodal alignment) and executing link reinforcement (dynamic optimisation). Its value lies not only in improving the efficiency of single point creation, but also in building an open platform to promote the evolution of the digital creative industry towards standardisation and modularisation.

5 Conclusions

This paper combines virtual exhibits with interactive displays as a design strategy, and takes multi-sensory integration and diversified presentation as the research direction of the article, thereby extending the cultural and commercial value of the exhibits. Meanwhile, this paper combines computer vision rendering technology to study the design system of art exhibits. The experimental results show that the VA2T algorithm based on time series force tactile data proposed in this paper has certain effects, and can realise reliable system functions after multiple iterations of learning. In addition, combined with experimental analysis, it can be seen that the model proposed in this paper has a certain effect in the design of art exhibits, which can effectively improve the design effect of art exhibits and enhance the actual experience of visitors.

In this paper, a mesh model with texture is proposed based on StyleGAN model. By utilising feature extraction and vector decoupling, StyleGAN's powerful two-dimensional image generation capabilities are extended to three-dimensional space, thus generating meshes with textures, whose performance outperforms the reconstruction performance of existing methods.

In response to the insufficient adaptability of algorithms to dynamic scenarios, parameter sensitivity, and training costs, the main research directions in the future are as follows: firstly, the algorithm model is improved through dynamic grouping strategy,

combined with meta learning (such as iterative optimisation of Meta CT), and the number of groups is adaptively adjusted according to the complexity of input features. Secondly, a hybrid computing framework is constructed to integrate the parallel advantages of quantum reinforcement learning and allocate high load computing tasks (such as attention weight generation) to dedicated hardware acceleration. The next step will be to build the above platform and continue to conduct model performance analysis through experiments to further improve the model's effectiveness.

Acknowledgements

This paper is supported by Jiangxi Provincial University Humanities and Social Sciences Research Project (Red Culture Education Special Project), Project No. HSWH24109.

Declarations

All authors declare that they have no conflicts of interest.

References

Cetinic, E. and She, J. (2022) 'Understanding and creating art with AI: review and outlook', *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, Vol. 18, No. 2, pp.1–22.

Ettalibi, A., Elouadi, A. and Mansour, A. (2024) 'AI and computer vision-based real-time quality control: a review of industrial applications', *Procedia Computer Science*, Vol. 231, No. 2, pp.212–220.

Fan, M. and Li, Y. (2020) 'The application of computer graphics processing in visual communication design', *Journal of Intelligent & Fuzzy Systems*, Vol. 39, No. 4, pp.5183–5191.

Fu, H., Jia, R., Gao, L., Gong, M., Zhao, B., Maybank, S. and Tao, D. (2021) '3D-future: 3D furniture shape with texture', *International Journal of Computer Vision*, Vol. 129, No. 3, pp.3313–3337.

Guo, S. and Wang, B. (2021) 'Application of computer aided modeling design in the expression techniques of sculpture art space', *Computer-Aided Design and Applications*, Vol. 19, No. S3, pp.1–12.

Kuang, Z., Olszewski, K., Chai, M., Huang, Z., Achlioptas, P. and Tulyakov, S. (2022) 'Neroic: neural rendering of objects from online image collections', *ACM Transactions on Graphics (TOG)*, Vol. 41, No. 4, pp.1–12.

Lattas, A., Moschoglou, S., Ploumpis, S., Gecer, B., Ghosh, A. and Zafeiriou, S. (2021) 'Avatarme++: facial shape and BRDF inference with photorealistic rendering-aware GANs', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 12, pp.9269–9284.

Li, J., Yang, J., Zhang, J., Liu, C., Wang, C. and Xu, T. (2020) 'Attribute-conditioned layout GAN for automatic graphic design', *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, No. 10, pp.4039–4048.

Lombardi, S., Simon, T., Schwartz, G., Zollhoefer, M., Sheikh, Y. and Saragih, J. (2021) 'Mixture of volumetric primitives for efficient neural rendering', *ACM Transactions on Graphics (ToG)*, Vol. 40, No. 4, pp.1–13.

Mohanto, B., Islam, A.T., Gobbetti, E. and Staadt, O. (2022) 'An integrative view of foveated rendering', *Computers & Graphics*, Vol. 102, No. 1, pp.474–501.

Moragane, H.P.M.N.L.B., Perera, B.A.K.S., Palihakkara, A.D. and Ekanayake, B. (2024) 'Application of computer vision for construction progress monitoring: a qualitative investigation', *Construction Innovation*, Vol. 24, No. 2, pp.446–469.

Scalera, L., Seriani, S., Gallina, P., Lentini, M. and Gasparetto, A. (2021) 'Human-robot interaction through eye tracking for artistic drawing', *Robotics*, Vol. 10, No. 2, pp.54–62.

Shan, P. and Sun, W. (2021) 'Research on landscape design system based on 3D virtual reality and image processing technology', *Ecological Informatics*, Vol. 63, No. 2, pp.101287–101298.

Sitzmann, V., Rezchikov, S., Freeman, B., Tenenbaum, J. and Durand, F. (2021) 'Light field networks: neural scene representations with single-evaluation rendering', *Advances in Neural Information Processing Systems*, Vol. 34, No. 3, pp.19313–19325.

Sun, Q., Wang, C., Qiang, F., **ong, D. and Wolfgang, H. (2021) 'End-to-end complex lens design with differentiable ray tracing', *ACM Trans. Graph*, Vol. 40, No. 4, pp.1–13.

Turchet, L. (2023) 'Musical Metaverse: vision, opportunities, and challenges', *Personal and Ubiquitous Computing*, Vol. 27, No. 5, pp.1811–1827.

Walton, D.R., Dos Anjos, R.K., Friston, S., Swapp, D., Akşit, K., Steed, A. and Ritschel, T. (2021) 'Beyond blur: real-time ventral metamers for foveated rendering', *ACM Transactions on Graphics*, Vol. 40, No. 4, pp.1–14.

Wang, R. (2021) 'Computer-aided interaction of visual communication technology and art in new media scenes', *Computer-Aided Design and Applications*, Vol. 19, No. S3, pp.75–84.

Yan, H., Zhang, H., Liu, L., Zhou, D., Xu, X., Zhang, Z. and Yan, S. (2022) 'Toward intelligent design: an AI-based fashion designer using generative adversarial networks aided by sketch and rendering generators', *IEEE Transactions on Multimedia*, Vol. 25, No. 3, pp.2323–2338.

Zhao, F., Jiang, Y., Yao, K., Zhang, J., Wang, L., Dai, H. and Yu, J. (2022) 'Human performance modeling and rendering via neural animated mesh', *ACM Transactions on Graphics (TOG)*, Vol. 41, No. 6, pp.1–17.