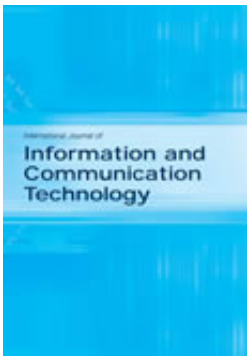
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

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

Application of 3D modelling in virtualisation of real works of art

Chujing Huang

Article History:

Received:	15 October 2024
Last revised:	02 January 2025
Accepted:	02 January 2025
Published online:	25 February 2025

Application of 3D modelling in virtualisation of real works of art

Chujing Huang

School of Fine Arts and Design,
Nanchang Institute of Technology,
Nanchang City, Jiangxi Province, 330044, China
Email: m15083500987@163.com

Abstract: In order to improve the virtualisation effect of real works of art, this paper combines the idea of 3D modelling to virtualise real works of art. This paper introduces attention mechanism on the basis of AnimeGAN network that is, adding squeeze excitation residual block to the network. And makes some optimisation in the loss function of AnimeGAN, puts forward a brand-new ExpressionGAN, and uses ExpressionGAN to transfer the image style of artistic images. From the analysis of the experimental results, the method proposed has good performance in peak signal-to-noise ratio, content loss and style loss. It shows that the network model proposed in this paper can further enrich the stylisation effect of the image without losing the quality of the stylised image, and can virtualise 3D objects through style transfer method.

Keywords: 3D modelling; real; works of art; virtualisation.

Reference to this paper should be made as follows: Huang, C. (2025) 'Application of 3D modelling in virtualisation of real works of art', *Int. J. Information and Communication Technology*, Vol. 26, No. 4, pp.39–56.

Biographical notes: Chujing Huang holds a Master's degree of Arts in Design and serves as a Lecturer at Nanchang Institute of Technology. Her research focus is on art design.

1 Introduction

The emergence of every art form will inevitably go through a development process from use to aesthetic integration, and it is based on labour, which is the only way for every art form. Early virtual reality design aimed at the simulation and immersive experience of real scenes and events. When the hardware equipment was very backward, practical VR works for aerospace and military were designed according to the application requirements. The development of technology and application demand has led to the emergence of a new design form of virtual reality design. Moreover, the aesthetic characteristics of VR works have been paid more and more attention, and excellent design works have begun to enter the market, participate in production and affect people's lives.

For the first time, the artistic design of virtual reality makes the object of people's artistic practice no longer the pure form of the external material world. It profoundly practices the design concept of integrating technology and art put forward by Gropius, and deeply combines computer technology, human-computer interaction technology and immersive art. The purpose of studying virtual reality design is from the dimensions of user experience, creation specification and content innovation. At the same time, under the vision of 6G technology, it combines the design concept of integrating ideas, technology and art to explore the road of immersive virtual reality design suitable for long-term development in the future (Bernardo and Duarte, 2022).

Traditional dynamic installation art mainly uses physical materials as the presentation medium, so that the effect of artistic ideas in the process of communication is restricted by time and space. Using art as a medium to spread culture and maximise the influence of culture is an important criterion to measure whether cultural value is realised. In this sense, dynamic installation art is influenced by physical space under the limitation of realistic conditions, so that it cannot fully exert its maximum potential of cultural value. On the other hand, the characteristics of traditional dynamic installation art are mainly manifested in the movement of its installation structure, which is a movement around the structure of the installation itself and belongs to a one-way communication mode rather than a real-time dynamic dialogue with the audience. The interaction between the audience and the device is mainly limited to the audience's appreciation of the device through audio-visual effects, but there is no deeper interaction. Therefore, the audience is in a state of passive acceptance of information. In the long run, the audience's interest and enthusiasm for such works will gradually decline, and at the same time, the artistic content will become rigid and lack of innovation, and ultimately it will be difficult to meet the needs of the times (Chiu et al., 2023).

With the vigorous development of digital technology, digitalisation has had a profound impact on human life forms and ways. People's demand for obtaining information has also undergone tremendous changes, and they have a new understanding of diversified information acceptance methods. Therefore, people are no longer used to the patterned and flat one-way information receiving mode, but like the two-way interaction and real-time communication mode of information dissemination. In the current increasingly virtual social environment, the audience is not only the viewer and receiver of information, but also the creator and publisher of information. Therefore, the virtualisation of art carrier is the inevitable requirement of the development of the times, and the research on the virtualisation transformation of traditional dynamic installation art is to meet the realistic needs of contemporary cultural communication (De Luca et al., 2023).

2 Related work

2.1 Virtualisation foundation of artistic works

The public's aesthetic needs have also changed, and the public has begun to pursue real-time, interactive and immersive virtual aesthetic experience. The research on virtual aesthetics mainly focuses on the aesthetic characteristics and elements of virtual reality and its differences from traditional aesthetics. Virtual aesthetics has been preliminarily discussed theoretically in film, television and games, but there is little research on artistic

design, especially on virtual aesthetics of virtual urban sculpture. As a means of virtual design, augmented reality technology also has the aesthetic characteristics of virtual aesthetics. At the same time, the relationship between aesthetic subject and object has also changed with the virtual design of works. In virtual design, the aesthetic process of the aesthetic subject is the experience process of its interaction with the works, and it is also the generation process of artistic works. This change represents the enrichment of the identity connotation of the aesthetic subject, and it also represents the expansion of the aesthetic initiative of the aesthetic subject. In addition, the feasibility of applying virtual aesthetics to art design has been discussed in the industry, but the research results are few (Feng and Zhang, 2022).

From traditional art design to digital art design to virtual reality art design (VRAD), the most significant difference among the three lies in the different spaces to which their drawings belong. The traditional medium of art design uses paper as the medium, which is always drawing dots, lines and surfaces on flat paper and creating on a two-dimensional plane. Until the birth of digital art design, the drawing tools were digitised, and the traditional art design effect was achieved through digital simulation, and the physical plane painting was painted into pure digital art design. Moreover, the drawing space has become a two-dimensional digital space, and the created works have thus become reproducible digital works of art. On the one hand, it simulates various traditional art design methods, and on the other hand, it carries out more innovative expansion (Gomez-Tone et al., 2022).

At the same time, many kinds of unique drawing tools for digital art design have been developed. So far, all art design is carried out in two-dimensional space until the appearance of VRAD. VRAD adopts virtual three-dimensional space art design. The three-dimensional space painting is very different from two-dimensional space painting. From the spatial point of view, there are great differences in painting methods. VRAD extends the paper of art design into three-dimensional space. At this time, the artist can design this three-dimensional space art design composed of three infinitely extending axes of XYZ, thus forming the effect of a real three-dimensional painting, and even the artist can fully feel the 'picture-in-picture' experience by controlling the camera to move inside the picture (Han and Gan, 2022).

As a new digital art, VRAD is inevitable to innovate in its expression form. Compared with traditional art design art, VRAD, which occupies the current new technology, has unique advantages. VRAD has three-dimensional vector and progressive three-dimensional space layer attributes, which enables artists to freely express their thoughts and feelings without being restricted by the plane space limitation of traditional art design. This adds a possibility to the expression of current artistic viewpoints. At the same time, it is worth mentioning the innovation and diversity of visual effects of VRAD (Hui et al., 2022).

2.2 Virtualisation application of artistic works

Kim et al. (2022) defined and classified virtual reality design art. Virtual reality design art is an immersive design art, which takes digital media as the carrier, and through the interaction and integration of science and technology and culture and art, makes users enjoy art design works in virtual space in the form of sensory interaction. Liao and She (2023) defined the concept of VRAD, and believed that virtual reality design has the

characteristics of immersion, interactivity, intuition, virtual reality and collaboration, and hold that the visual effects of virtual reality design works are generally realistic virtual, transcendental virtual and fantasy virtual.

With the improvement of VR hardware performance and the development of VR market, the research content of virtual reality design is becoming more and more detailed. Mao (2022) put forward the virtual design process from 3D modelling to virtual simulation, which focused on the immersive experience of virtual design, but ignored the practical role of information interface in virtual design. At the same time, art researchers generally pay attention to immersive scene design, virtual reality interaction design and three-dimensional interface design in virtual reality design.

Matovu et al. (2023) believed that immersive scene design is the fusion of various technologies and has remarkable aesthetic characteristics. When designing, it should aim at meeting users' needs, take psychological theory as the basis, and use digital media as the main tool to build virtual scenes to stimulate users' sensory experience, so as to achieve the state of immersion and forgetting others. Mills and Brown (2022) believed that virtual reality design should pay attention to immersion, interaction and authenticity, create a scene to tell stories to people, and even guide people to live in it.

In terms of three-dimensional interface design, Mu et al. (2024) believed that VE interface should be conceived to highlight creativity and functional application to emphasise experience, and advocated strengthening the immersive experience of virtual reality design through multi-style interface design, such as dual interaction, sensory linkage, dynamic and static combination, etc., and focused on humanistic care and emotional appeal in the design process.

In terms of virtual reality interaction, Piankarnka et al. (2023) believed that virtual reality interaction design should take digital media as the core, human perception as the main body, and combine the power of technology, artistic thinking and design methods to design interactive works with higher quality and experience. Porsani et al. (2023) made a more comprehensive summary of the interactive modes of virtual reality, and put forward the design methods and principles of VR interaction in combination with the perspective of screen technology, scene effect and experience.

Since the development of virtual reality, it is undeniable that its great success is not only a major breakthrough in human science and technology, but also expands the boundaries of the world and even breaks the limitations of time and space. However, virtual reality is only showing the tip of its iceberg at present, and it still has endless development potential. Sina and Wu (2023) believed that virtual reality is facing problems such as vertigo, screen refresh rate, clarity and rendering.

With the explosion of the Metaverse, virtual reality design has ushered in a new development direction. Tsou and Mejia (2024) detailed the completely different significance that Metaverse brings to virtual reality design. Umair et al. (2022) mentioned that Metaverse's art will not only achieve a breakthrough in the real material world in imitating perception, but also shape our perception and experience. Ververidis et al. (2022) believed that the development of digital technologies related to Metaverse will promote the all-round development of the quality, content and form of virtual reality design. In the future, it is the development direction and inevitable requirement of virtual reality design to incorporate traditional culture into the artistic design of 'Metaverse' virtual world and practice socialist core values in the virtual world.

3 Style transfer based on ExpressionGAN

When CartoonGAN transfers real-life portrait images to comic style, it is prone to serious ambiguous colour blocks, texture incongruity and other problems. AnimeGAN solves the problem of ambiguous colour blocks in CartoonGAN, but when the image is transferred, the details and textures of the generated image of AnimeGAN will be seriously lost. Therefore, this paper introduces attention mechanism on the basis of AnimeGAN network, that is, adds SE-ResidualBlock to the network, and makes certain optimisations in the loss function of AnimeGAN, proposes a new ExpressionGAN, and uses ExpressionGAN to transfer 3D image styles of images.

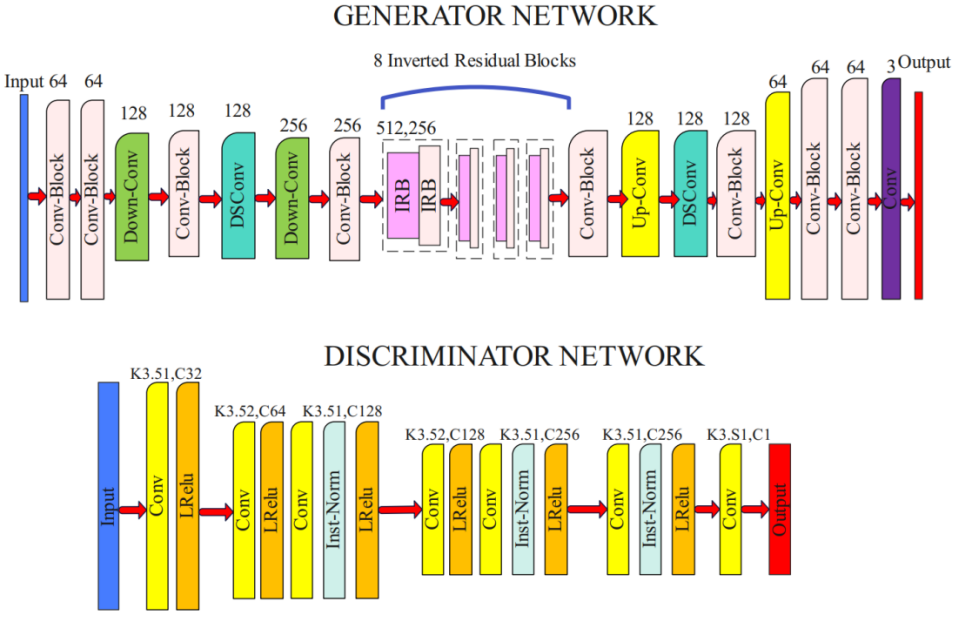
3.1 Introduction of AnimeGAN

Compared to existing research, residual block mainly improves stylisation performance by introducing squeeze and excitation (SE) modules.

The SE module adaptively recalibrates channel feature responses by explicitly modelling the interdependence between feature channels. This helps the network focus on important features and suppress unimportant features. Introducing SE module into the residual block can enhance the network's ability to process feature information, thereby improving the performance of the model.

The SE module compresses the spatial features into a channel descriptor through global average pooling operation, and then learns the nonlinear relationships between channels through two fully connected layers, generating a channel weight vector. This weight vector is used to recalibrate the channel feature response. This approach can make the network pay more attention to important feature channels, thereby improving the expression ability of features and helping the network better learn stylised information.

Figure 1 AnimeGAN network structure diagram (see online version for colours)



The SE module introduces additional attention mechanisms to enable the network to adaptively adjust the feature responses of different channels. This helps the network better adapt to different input data and enhance its generalisation ability. Introducing SE module into the residual block can make the network more robust, insensitive to small changes in input data, thereby improving the stability and performance of the model. AnimeGAN is implemented by combining neural networks and adversarial generative networks, and its purpose is to transform real-world images into comic-style images. AnimeGAN is an improvement on CartoonGAN. On the basis of CartoonGAN, it adds three novel loss functions and increases the types of input images to achieve the purpose of truly restoring the colour of real images.

AnimeGAN is designed based on an adversarial generative network. The specific network structure is shown in Figure 1 (Wang and Hu, 2022):

In Figure 1, the discriminator is mainly responsible for calculating loss and back propagation, so the discriminator does not generate an image, so the output image of the discriminator is consistent with the input image.

Firstly, AnimeGAN obtains the greyscale images of comic images and delineated comic images through greyscale matrix. The main purpose of this is to keep the original image texture and eliminate dark interference.

The total loss function of AnimeGAN is defined as formula (1):

$$L(G, D) = \omega_{adv} L_{adv}(G, D) + \omega_{con} L_{con}(G, D) + \omega_{gra} L_{gra}(G, D) + \omega_{col} L_{col}(G, D) \quad (1)$$

In formula (1), G represents the generator, D represents the discriminator, and $L_{adv}(G, D)$ and $L_{con}(G, D)$ represent the adversarial loss and content loss. $L_{gra}(G, D)$ represents the greyscale adversarial loss of the cartoon style in the texture and lines of the generated image, and $L_{col}(G, D)$ represents the colour reconstruction loss constructed by restoring the greyscale image. ω_{adv}, ω_{con}, ω_{gra}, ω_{col} represents the weight of each loss function, and the best effect is achieved when $\omega_{adv} = 300$, $\omega_{con} = 1.5$, $\omega_{gra} = 3$, $\omega_{col} = 10$ is obtained through experiments.

The real image is defined as: $S_{data}(p) = \{p_i | i = 1, 2, \ldots, N\}$, where N is the number of real images. The comic image data is defined as: $S_{data}(a) = \{a_i | i = 1, 2, \ldots, M\}$, where M is the number of comic images. The comic greyscale image data is defined as: $S_{data}(x) = \{x_i | i = 1, 2, \ldots, M\}$, the de-lined comic image is defined as: $S_{data}(e) = \{e_i | i = 1, 2, \ldots, M\}$, and the de-lined comic greyscale image data is defined as: $S_{data}(y) = \{y_i | i = 1, 2, \ldots, M\}$. The greyscale adversarial loss is specifically defined as formula (2):

$$\begin{aligned} L_{gra}(G, D) &= E_{p_i \sim S_{data}(p)}, E_{x_i \sim S_{data}(x)} \left[\left\| Gram\left(VGG\left(p_i\right)\right) - Gram\left(VGG\left(x_i\right)\right) \right\|_1 \right] \\ &= \int_{p_i}^{S_{data}(p)} \int_{x_i}^{S_{data}(x)} \left[\left\| Gram\left(VGG\left(p_i\right)\right) - Gram\left(VGG\left(x_i\right)\right) \right\|_1 \right] \end{aligned} \quad (2)$$

In formula (2), *Gram* represents the Gram matrix of the feature. p_i represents the input image, $G(p_i)$ represents the generated image, and $(VGG(x_i))$ represents the input feature map of the i^{th} layer in VGG.

Its content loss is specifically defined as formula (3):

$$L_{con}(G,D) = E_{p_i \sim S_{data}(p)}\left[\left\|VGG_l(p_i) - VGG_l(G(p_i))\right\|_1\right] = \int_{p_i}^{S_{data}(p)}\left[\left\|VGG_l(p_i) - VGG_l(G(p_i))\right\|_1\right] \tag{3}$$

In formula (3), VGG_l is the feature map of the i^{th} layer of VGG, p_i represents the input image, and $G(p_i)$ is the generated image. Both grayscale adversarial loss and content loss use L_1 sparse regularisation for loss calculation.

Figure 2 Overall frame diagram (see online version for colours)

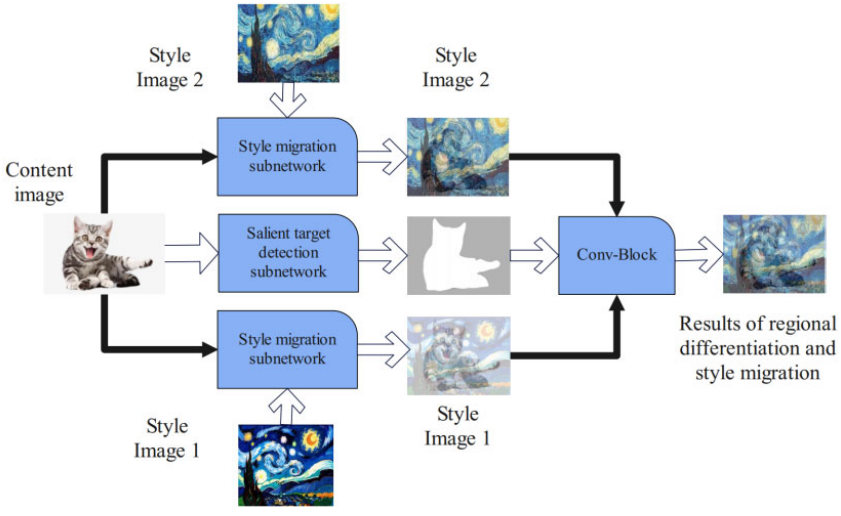
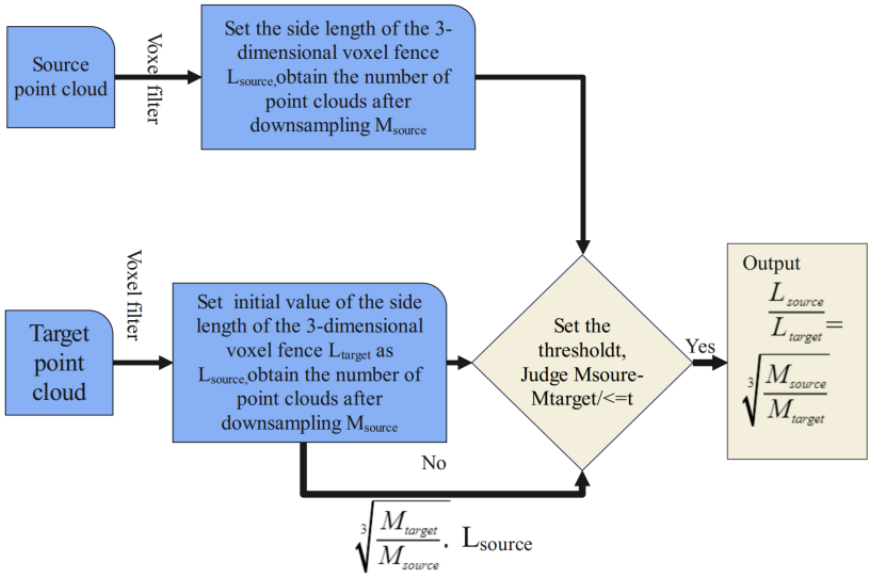


Figure 3 Construction process of improved voxel filter to solve overall density difference (see online version for colours)



3.2 Virtualisation style transfer model

In order to make the effect of style transfer richer and more natural, and further improve the quality of stylised renderings, this paper proposes a network model of regional style transfer, which mainly includes salient target detection sub-network, image style extraction sub-network and fusion sub-network. The overall framework is shown in Figure 2. Specifically, the model firstly segments the picture into different front and background regions through the salient object detection sub-network. Secondly, the two style conversion sub-networks stylise the content image by combining their respective style images. Finally, the fusion sub-network uses Poisson fusion to fuse the two images according to the front and background regions, so as to generate a new image. The network model proposed in this paper can further enrich the stylisation effect of the image without losing the quality of the stylised image

Traditional voxel filters only consider the global point cloud information, and perform poorly on local information. After passing through the voxel filter, the density of all parts will be the same regardless of the amount of information. Therefore, in the same cloud data, the side length of its three-dimensional voxel fence should be inversely proportional to the density.

This paper introduces a new concept of relative density. The concept of relative density here refers to the fact that when two point clouds of identical objects only have scale difference and point cloud density difference, when the object A changes to the size of object B through the scaling ratio between them, the relative density of the two objects is said to be equal. In order to make the relative density of the source point cloud and the target point cloud equal, the number of point clouds after the voxel filter is selected as a numerical judgement, and this value is the effective number of three-dimensional voxel fences.

The side length of the three-dimensional voxel fence is set to L_{source}, and the number M_{source} of point clouds obtained after downsampling is obtained. Then, the target point cloud P_2 is downsampled by voxel filtering, the initial value of the side length L_{target} of the three-dimensional voxel fence is set to L_{source}, and the number M_{target} of point clouds after downsampling is obtained. A threshold t is set for the absolute value of the difference between M_{source} and M_{target}. When it is greater than the threshold t, L_{target} is reassigned so that $L_{target} = \sqrt[3]{\frac{M_{target}}{M_{source}}} L_{source}$. Then, the above steps are repeated until the absolute value of the difference between M_{source} and M_{target} is less than the set threshold, so that two sets of point cloud data with similar relative density can be obtained. The specific process is shown in Figure 3.

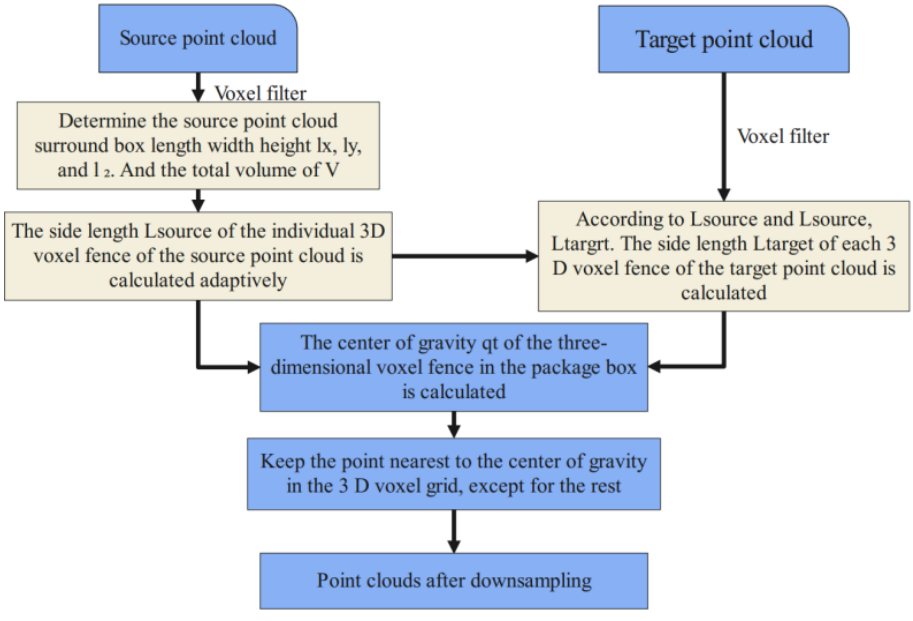
Through the above steps, the scaling approximation of the source point cloud and the target point cloud is obtained, as shown in formula (4) (Zhao and Zhao, 2022):

$$L_{source} : L_{target} = \sqrt[3]{\frac{M_{source}}{M_{target}}} \tag{4}$$

For the point cloud of real objects, the density of each part of the point cloud is also different. If this situation is not considered, a large amount of feature point information will be lost. The reason is that the denser the part, the greater the amount of information it contains. Therefore, it is necessary to adaptively adjust the side length of the

three-dimensional voxel fence for the point cloud density of each part of the point cloud data, instead of all the three-dimensional voxel fences of a point cloud data having the same length. Finally, the center of gravity of each three-dimensional voxel fence is calculated, and the point closest to the center of gravity is solved by KD-tree algorithm. This point is the sampling point that needs to be retained, and other points in the three-dimensional voxel fence other than this point are deleted. Finally, all the points retained by the three-dimensional voxel fence constitute simplified point cloud data. The construction process is shown in Figure 4:

Figure 4 Construction process of improved voxel filter to solve local density difference (see online version for colours)



The specific steps are as follows:

1 The length l_x, width l_y and height l_z of the overall bounding box of the source point cloud data are calculated, and the result is shown in formula (5):

$$\begin{cases} l_x = x_{\max} - x_{\min} \\ l_y = y_{\max} - y_{\min} \\ l_z = z_{\max} - z_{\min} \end{cases} \tag{5}$$

Among them, $x_{\max}$, $x_{\min}$, $y_{\max}$, $y_{\min}$, $z_{\max}$ and $z_{\min}$ are the maximum and minimum values in the X-axis and Y-axis and Z-axis direction of the bounding box. Therefore, the total volume of the bounding box can be obtained as shown in formula (6):

$$V = l_x \cdot l_y \cdot l_z \tag{6}$$

2 The side length L_{target} of the three-dimensional voxel fence, and the initial value of the side length is L_{source}. The basic idea of this part is: in the part with higher density

of source point cloud, the 3D voxel fence takes a larger value, and in the part with lower density of source point cloud, the 3D voxel fence takes a smaller value. If it is assumed that there are n points in any 3D voxel fence, the spatial distance of the neighbourhood is calculated based on these n points. Since the greater the density of the area, the greater the point cloud feature information. As shown in formula (7):

$$L_{source} = \lambda \sqrt[3]{\frac{g}{h}} \tag{7}$$

Among them, λ is the scaling factor of the 3D voxel fence side length, which can adaptively change the 3D voxel fence side length, and g is the scale factor of the 3D voxel fence. Among them, n can be obtained by dividing the total number of point clouds in the bounding box N by the total volume V of the bounding box, as shown in formula (8):

$$n = \frac{N_{total}}{V} = \frac{N_{total}}{l_x \cdot l_y \cdot l_z} \tag{8}$$

By substituting formula (7) into formula (8), the obtained result is shown in formula (9):

$$L_{source} = \lambda \sqrt[3]{\frac{g l_x \cdot l_y \cdot l_z}{N_{total}}} \tag{9}$$

Since there is a proportional relationship between the target point cloud and the side length of the three-dimensional voxel fence of the target point cloud, the results that can be obtained are shown in formula (10):

$$L_{target} = \sqrt[3]{\frac{M_{source}}{M_{target}}} \cdot L_{source} = \lambda \sqrt[3]{\frac{g l_x \cdot l_y \cdot l_z M_{target}}{N_{total} M_{source}}} \tag{10}$$

3 The three-dimensional voxel fence is divided. The bounding box grid is divided into A, B, and C parts according to the X-axis, Y-axis, and Z-axis, as shown in formula (11):

$$\begin{cases} A = \left| \frac{l_x}{L} \right| \\ B = \left| \frac{l_Y}{L} \right| \\ C = \left| \frac{l_Z}{L} \right| \end{cases} \tag{11}$$

Among them, L represents the side length of the three-dimensional voxel fence. Because the processing of the source point cloud and the target point cloud in the later steps is consistent, there is no difference between L_{source} and L_{target}. The symbol | | indicates a downward rounding operation, which ensures that all points in the bounding box can be contained by the three-dimensional voxel fence.

4 The centroid of each 3D voxel fence is calculated. Any point in the point cloud is (x_i), and any 3D body is the centroid $q_i(x_{qi}, y_{qi}, z_{qi})$ of the fence, as shown in formula (12):

$$\begin{cases} x_{qi} = \dfrac{1}{n}\sum_{i=1}^{k} x_i \\ y_{qi} = \dfrac{1}{n}\sum_{i=1}^{k} y_i \\ z_{qi} = \dfrac{1}{n}\sum_{i=1}^{k} z_i \end{cases} \tag{12}$$

5 The centroid of the three-dimensional voxel fence with points in all bounding boxes is calculated, and then the KD-tree algorithm is used to determine the point in the three-dimensional voxel fence that is closest to the centroid. This point is determined as the sampling point and retained, and other points are eliminated.

6 The sampling points retained by all 3D voxel fences are output, and these sampling points are finally the new point cloud obtained for subsequent point cloud registration operations. This adaptive voxel filter can not only reduce the amount of point cloud data without losing point cloud features and achieve the purpose of downsampling, but also solve the density difference under the premise of scale differences across source point clouds.

4 Experimental study

4.1 Methods

The dataset of this article is MoMA, which includes data on 98361 artworks involving 27,140 artists. Data content provides detailed information such as artist, work name, introduction, production year, size, etc. The data is provided in CSV and JSON formats, compatible with UTF-8 encoding in multiple language environments. Used in various fields such as academic research, artistic creation, application development, and big data analysis in order to ensure that the artwork in the image is centred as much as possible, all images are aligned vertically and horizontally. Divide the dataset into training set and validation set, corresponding to the train directory and val directory, both of which have cat, dog and wild folders are used to distinguish the visual domain, and 1,000 images corresponding to each category are taken from each sub folder in the validation set, while the remaining images are used as the training set.

Some of the experimental works of art studied in this paper come from the internet, and some are real-life works of art (Figure 5). The resolution of all images is 512 ∗ 512. The experimental equipment is an all-in-one computer equipped with Windows 10 system. On the one hand, it is to reduce the calculation. The reason is that too high image resolution will bring huge amount of calculation. On the other hand, it is to unify the scale of content images and style images and meet the position mapping of image reconstruction. The CPU is Intel i7 8700K 3.20 GHz, and the RAM is 16 G. The experiment is mainly based on the Ubuntu operating system under Linux, and it uses the

PyTorch framework and the script uses Python3. The specific detailed configuration is shown in Table 1.

Table 1 Configuration of experimental platform

Experimental environment	*Model or version*
Operating system	Linux Ubuntu
GPU	NIDIA Tesla P100
CPU	Intel Xeon E5
Memory	16 G
Programming language	Python

Figure 5 Example of real works of art (see online version for colours)

(a) (b) (c)

4.2 Results

The example works are virtualised through the model in this paper, and the transformation from real works to virtualised works is realised through the above-mentioned style transfer algorithm, as shown in Figure 6.

Figure 6 Virtualisation of real works of art (see online version for colours)

(a) (b) (c)

On the one hand, the effect of image style transfer method is evaluated through objective evaluation indicators such as SSIM, and on the other hand, subjective evaluation is performed through style transfer effect graph. Compared methods include ArtFlow,

Stytr2, and AdaAttN, where the ArtFlow method can be divided into ArtFlow + AdaIN and AnFlow + wCT. The constraint effect of network structure and local loss function based on detail structure enhancement proposed in this paper and the comparison of the effects of other methods on the same verification set are shown in Figure 7. The average value of SSIM results obtained by inference from more than 900 data.

Figure 7 Comparison of objective evaluation indicators of different methods (see online version for colours)

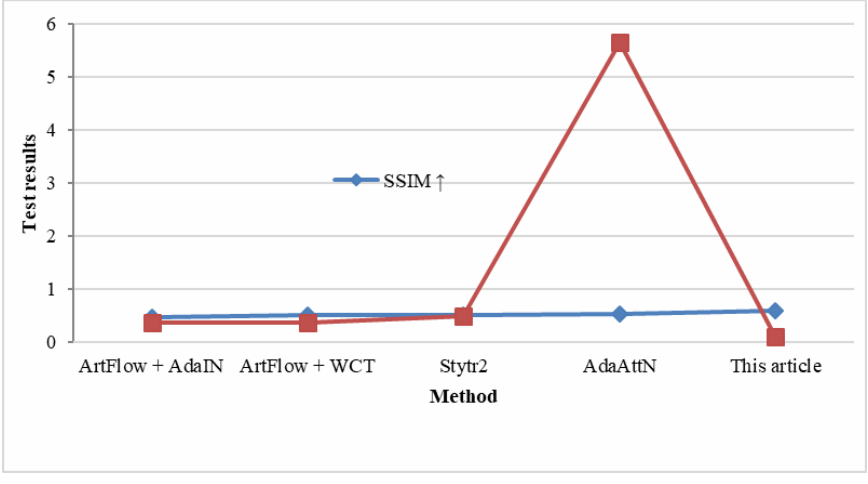
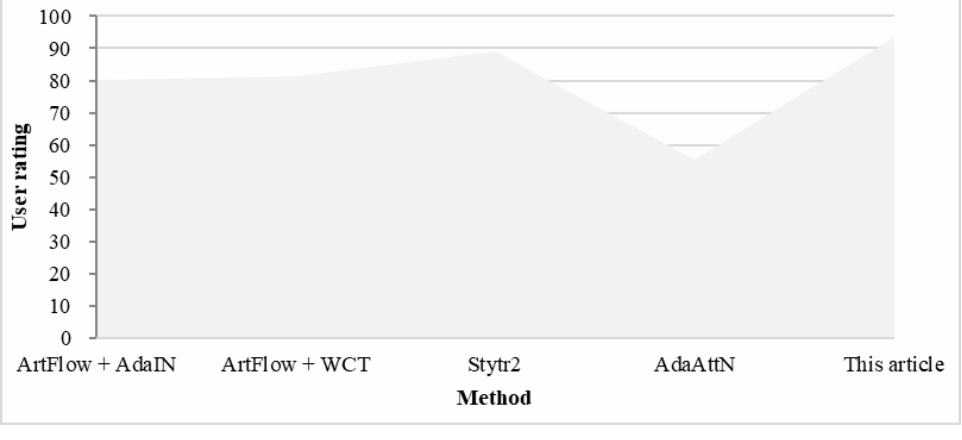


Figure 8 Subjective evaluation of users



It is a subjective problem to evaluate the quality of image virtualisation results, so it is necessary to conduct a subjective questionnaire survey. The form of the questionnaire survey is to ask 120 college graduates of the Academy of Fine Arts (who have a bachelor's degree in art history and have some theoretical research on modern art style works) to compare the results obtained by the method proposed in this paper with those obtained by the above image style transfer methods, assign a weight of 70% to the statistical results and conduct a sample survey of 100 members of society. Assign a

weight of 30% to the final results. The final result is shown in Figure 8, and the results synthesised by this method have been affirmed by most professional scholars.

In this paper, the style transfer results of different methods are compared from three aspects: peak signal-to-noise ratio (PSNR), content loss and style loss. The objective data analysis is shown in Table 2.

Table 2 Objective evaluation analysis

Method	*ArtFlow + AdaIN*	*ArtFlow + WCT*	*Stytr2*	*AdaAttN*	*This article*
PSNR	17.080	18.357	13.701	10.785	18.425
Content loss	4.277	4.428	4.202	5.778	2.235
Style loss	1.308	0.902	2.432	2.532	0.823

ArtFlow is a more comprehensive and advanced image style transfer framework, and AdaIN is one of the optional style transfer modules. The combination of the two can achieve better style transfer effects ArtFlow focuses more on lossless separation of content and style, as well as avoiding content leakage issues, while WCT achieves style transfer through covariance matching in feature space. The combination of these two methods can achieve the best style transfer effect. StyTr2 is an innovative image style transfer framework that achieves high-quality image style transfer effects by introducing Transformer architecture and content aware position encoding technologies. AdaAttN is an innovative neural style transfer method that introduces an adaptive attention normalisation mechanism to achieve local statistical alignment between content features and style features, improving the quality of image style transfer. The above modules are currently advanced image style transfer models, so comparing these models with our model can verify the effectiveness of our model.

The assumption of FID is that random variables follow a Gaussian distribution, and the mean and covariance matrix between the real image distribution and the generated image distribution are used to measure the distance between the two distributions. The calculation formula for FID is shown in equation (13):

$$FID(x, g) = \left\| u_x - u_g \right\|_2^2 + Tr\left(\sum_x + \sum_g -2\left(\sum_x \sum_g \right)^{0.5} \right) \tag{13}$$

In the formula, x represents the features of the real image; g representing the features of the generated image; u representing the mean of image features; $\sum_x$ covariance matrix representing real image features; $\sum_g$ representing the covariance matrix of generated image features.

LPIPS is a method of measuring perceptual differences in images by utilising their depth features. The main idea of LPIPS is to input the images that need to be compared into a deep network, extract the output of the images after activation function, normalise it and assign weights, calculate the L2 distance of the feature vectors between each layer, and finally take the average of the distances between all layers to obtain the similarity of the images. The calculation formula for LPIPS is shown in equation (14).

$$d\left(x, x_0\right) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(\hat{y}_{hw}^l - \hat{y}_{ohw}^l \right) \right\|_2^2 \tag{14}$$

In the formula, x and x_0 represent the comparison image of the input; l representing the l layer network; $\bar{y}$ representing the feature vector of an image; W_l assign weights to the features of the l layer.

Frechét inception distance (FID) is used to measure the quality of the generated images, and learned perceptual image patch similarity (LPIPS) is used to measure the diversity of the generated images. As shown in Table 3.

Table 3 Quantitative comparison of image-guided synthesis

Parameter	*ArtFlow + AdaIN*	*ArtFlow + WCT*	*Stytr2*	*AdaAttN*	*This article*
FID ↓	106.03	52.77	39.20	23.56	21.48
LPIPS ↑	0.17	0.31	0.31	0.38	0.40

4.3 Analysis and discussion

When using three-dimensional sensors to obtain point cloud data of objects and environments, there are noise and outliers in point cloud data due to the influence of illumination in the environment and shooting angle. In addition, in order to better display the surface feature information of objects, the point cloud data obtained by the sensor will be relatively dense and the amount of data will be large. The amount of data can reach 100,000 or even tens of millions. Such a huge amount of data requires a lot of computing resources, which increases the difficulty and efficiency of registration, and even affects the accuracy of registration. Therefore, before registering the point cloud, the original point cloud needs to be denoised and downsampled.

At the same time, in order to solve the problems of density difference and data missing among cross-source point clouds, this paper proposes a multi-filter solution. These include Gaussian filter, curvature-based conditional filter and adaptive voxel filter. Gaussian filter is mainly used to remove noise and outliers in point cloud, and at the same time, it can smooth point cloud data. The curvature-based conditional filter plays a role in eliminating areas with weak feature information in the point cloud and highlighting the feature information of the point cloud in this process. In addition, the adaptive voxel filter can solve the problem of density difference in cross-source point cloud registration and reduce the amount of point cloud data.

From the style transfer effect in Figure 6, it can be clearly seen that the factors of the style reference image are inherited, and there is good continuity in the connection between the cat body and the background and the overall style transfer expression. Moreover, the details of the edge are also well transferred from the sketch style reference image, and it has a more obvious sense of atmosphere, and the transfer effect has a strong sense of filter. At the same time, from the image style transfer effect in Figure 6, it can be seen that the local detail structure enhancement method proposed in this paper can generate high-quality stylised images more stably, and has good effects in different content images and style images.

From the SSIM index in Figure 7, it can be seen that the result of this method is 0.587, which is 0.062 higher than the suboptimal method. It shows that the method in this paper has been better improved in enhancing the detail structure. The Time indicator indicates the time required for a test data to complete the inference. As can be seen from the specific value shown in Figure 7, the inference time of a single image based on GPU Tesla graphics card proposed in this paper is 101 ms. Although it is twice that of

AdaAttN method, it is still more efficient than other methods. Compared with AdaAttN method, the efficiency loss is mainly caused by increasing the calculation of LSN network parameters. Therefore, it can be seen from this table that the method proposed in this paper has a better balance between the effect of style transfer and reasoning efficiency.

From Figure 8, the results synthesised by this method have been affirmed by most professional scholars.

Content loss indicates whether the image style migration content is well preserved and whether the content structure is distinct. It is determined by the mean square error of the feature mapping between the generated result and the original content image in a specific layer in the convolutional network. The style loss is determined by the Gram matrix of the feature mapping between the generated result and the original style image in the convolutional network, which indicates whether the style of the artistic image is well preserved to the generated result. In Table 2, the method proposed in this paper has good performance in three aspects: peak signal-to-noise ratio, content loss and style loss, which shows that the stylised image generated by this algorithm has good quality, high content structure retention, and more texture information is retained in the style transfer result.

Table 3 shows the quantitative evaluation indicators of different models to guide image style transfer through reference images. It can be seen from the table that the FID indicators of the improved model in this paper on both datasets are the best, which is lower than the lowest FID value in the baseline method. In terms of generated image diversity, other baseline methods suffer from mode collapse in the multi-domain image style transfer task and have higher FID and lower LPIPS on the dataset.

Through the above analysis, it can be seen that the 3D modelling method proposed in this paper has a good method in the process of virtualisation of physical artworks. Moreover, it can virtualise 3D physical objects through style transfer method, which promotes the application of image processing technology in art design.

In ExpressionGAN image processing, parameter adjustment and training stability are crucial, which typically involves multiple techniques and strategies.

Firstly, to address the issue of unstable training of GANs, some general strategies can be adopted to improve stability. For example, using alternative loss functions such as Earth mover distance in Wasserstein GAN to replace traditional Jensen Shannon divergence helps maintain continuity when the distributions of real and generated data do not intersect. In addition, the two timescale update rule (TTUR) is also an effective method to stabilise the training process by using different learning rates for the discriminator and generator.

Secondly, for parameter adjustment, standardising input is an important step. Normalise the input image to a range of –1 to 1 and use the tanh activation function as the output of the final layer of the generator, which helps ensure data stability and convergence. In addition, the modified loss function is also key to improving the performance of GANs. In actual code, using reverse labels to train the generator, that is, training the generated image as the label of the real image, can more conveniently optimise the generator. In addition, to avoid the common problem of mode collapse in GAN training, some specific strategies can be adopted. For example, using gradient penalty to avoid gradient explosion or vanishing caused by weight reduction. Feature matching is also an effective method that stabilises the training process by minimising the statistical differences between the features of the real image and the generated image.

In addition, micro batch identification is also a method to alleviate pattern collapse, which detects generated images by calculating the similarity between the image and other images in the same batch, thereby prompting the generator to generate images with more diversity that is closer to real images. In summary, the parameter adjustment and training stability of ExpressionGAN image processing require a comprehensive consideration of multiple techniques and strategies, including the use of alternative loss functions, normalised inputs, modified loss functions, and avoiding pattern collapse. By applying these strategies and techniques reasonably, the training stability and performance of GAN can be effectively improved. This is also the next step that needs to be done in this article.

5 Conclusions

In order to solve the problem of virtualisation of real artworks, this paper proposes ExpressionGAN. This network framework is improved on the basis of AnimeGAN. Firstly, it adds the squeeze-excited residual block (SE-ResidualBlock) to AnimeGAN, and also optimises the loss function in AnimeGAN. Secondly, in terms of dataset processing, it introduces the comic face detection mechanism to perform screening and refinement. In order to make the effect of style transfer richer and more natural and further improve the quality of stylised renderings, this paper proposes a regional style transfer network model. It mainly includes salient target detection sub-network, image style extraction sub-network and fusion sub-network. According to the analysis of the experimental results, it can be seen that the 3D modelling method proposed in this paper is a good method in the virtualisation process of physical artworks. Moreover, it can virtualise 3D objects through the style transfer method, which promotes the application of image processing technology in artwork design.

This paper uses window self-attention to complete the fusion of features of different scales, and the backbone of the network is still the CNN architecture. However, this paper has not been able to fully build a multi-domain image style transfer model with Transformer as the backbone network, which is an aspect that needs to be improved in future work. Build Transformer-based models, training models, and application models for style transfer, using transformer's self attention mechanism to capture long-range dependencies in images, thereby extracting richer feature representations. The decoder is responsible for generating new images with the target style based on the extracted content and style features.

References

Bernardo, N. and Duarte, E. (2022) 'Immersive virtual reality in an industrial design education context: what the future looks like according to its educators', *Computer-Aided Design & Applications*, Vol. 19, No. 2, pp.238–255.

Chiu, M.C., Hwang, G.J. and Hsia, L.H. (2023) 'Promoting students' artwork appreciation: an experiential learning-based virtual reality approach', *British Journal of Educational Technology*, Vol. 54, No. 2, pp.603–621.

De Luca, V., Gatto, C., Liaci, S., Corchia, L., Chiarello, S., Faggiano, F., … and De Paolis, L.T. (2023) 'Virtual reality and spatial augmented reality for social inclusion: the 'Includiamoci' project', *Information*, Vol. 14, No. 1, pp.38–46.

Feng, L. and Zhang, W. (2022) 'Design and implementation of computer-aided art teaching system based on virtual reality', *Computer-Aided Design and Applications*, Vol. 20, No. S1, pp.56–65.

Gomez-Tone, H.C., Chávez, M.A., Samalvides, L.V. and Martin-Gutierrez, J. (2022) 'Introducing immersive virtual reality in the initial phases of the design process – case study: freshmen designing ephemeral architecture', *Buildings*, Vol. 12, No. 5, pp.518–530.

Han, L. and Gan, L. (2022) 'The application of artificial neural network combined with virtual reality technology in environment art design', *Computational Intelligence and Neuroscience*, Vol. 2022, No. 1, pp.7562167–7562177.

Hui, J., Zhou, Y., Oubibi, M., Di, W., Zhang, L. and Zhang, S. (2022) 'Research on art teaching practice supported by virtual reality (VR) technology in the primary schools', *Sustainability*, Vol. 14, No. 3, pp.1246–1254.

Kim, H., So, H.J. and Park, J.Y. (2022) 'Examining the effect of socially engaged art education with virtual reality on creative problem solving', *Educational Technology & Society*, Vol. 25, No. 2, pp.117–129.

Liao, T. and She, J. (2023) 'How does virtual reality (VR) facilitate design? A review of VR usage in early-stage engineering design', *Proceedings of the Design Society*, Vol. 3, pp.2115–2124.

Mao, W. (2022) 'Video analysis of intelligent teaching based on machine learning and virtual reality technology', *Neural Computing and Applications*, Vol. 34, No. 9, pp.6603–6614.

Matovu, H., Ungu, D.A.K., Won, M., Tsai, C.C., Treagust, D.F., Mocerino, M. and Tasker, R. (2023) 'Immersive virtual reality for science learning: design, implementation, and evaluation', *Studies in Science Education*, Vol. 59, No. 2, pp.205–244.

Mills, K.A. and Brown, A. (2022) 'Immersive virtual reality (VR) for digital media making: transmediation is key', *Learning, Media and Technology*, Vol. 47, No. 2, pp.179–200.

Mu, M., Dohan, M., Goodyear, A., Hill, G., Johns, C. and Mauthe, A. (2024) 'User attention and behaviour in virtual reality art encounter', *Multimedia Tools and Applications*, Vol. 83, No. 15, pp.46595–46624.

Piankarnka, V., Lertbumroongchai, K. and Piriyasurawong, P. (2023) 'A digital painting learning model using mixed-reality technology to develop practical skills in character design for animation', *Advances in Human-Computer Interaction*, Vol. 2023, No. 1, pp.5230762–5230773.

Porsani, R.N., Trindade, A.B.C., Demaison, A., Mont'Alvão, C.R. and Paschoarelli, L.C. (2023) 'Evaluations of design and user experience in virtual reality: a systematized bibliographic review', *Revista de Ciencia y Tecnología: RECyT*, Vol. 40, No. 1, pp.38–49.

Sina, A.S. and Wu, J. (2023) 'The effects of retail environmental design elements in virtual reality (VR) fashion stores', *The International Review of Retail, Distribution and Consumer Research*, Vol. 33, No. 1, pp.1–22.

Tsou, M.H. and Mejia, C. (2024) 'Beyond mapping: extend the role of cartographers to user interface designers in the Metaverse using virtual reality, augmented reality, and mixed reality', *Cartography and Geographic Information Science*, Vol. 51, No. 5, pp.659–673.

Umair, M., Sharafat, A., Lee, D.E. and Seo, J. (2022) 'Impact of virtual reality-based design review system on user's performance and cognitive behavior for building design review tasks', *Applied Sciences*, Vol. 12, No. 14, pp.7249–7260.

Ververidis, D., Nikolopoulos, S. and Kompatsiaris, I. (2022) 'A review of collaborative virtual reality systems for the architecture, engineering, and construction industry', *Architecture*, Vol. 2, No. 3, pp.476–496.

Wang, Y. and Hu, X.B. (2022) 'Three-dimensional virtual VR technology in environmental art design', *International Journal of Communication Systems*, Vol. 35, No. 5, pp.e4736–e4744.

Zhao, J. and Zhao, X. (2022) 'Computer-aided graphic design for virtual reality-oriented 3D animation scenes', *Computer-Aided Design and Applications*, Vol. 19, No. 1, pp.65–76.