Augmented Reality Animation Image Information Extraction and Modeling Based on Generative Adversarial Network

Xiaoxuan Tian[1] and Chengkun Li[2]

[1]Xi'an International Studies University, Xi'an 710128, China, tianxiaoxuan@xisu.edu.cn
[2]Information Center, POWERCHINA Northwest Engineering Corporation Limited, Xi'an 710065, China, m17791520860@163.com

Corresponding author: Xiaoxuan Tian, tianxiaoxuan@xisu.edu.cn

Abstract. Modeling is the first step of 3D animation design and production, and it is the core and foundation of the 3D world. Without a good model, other good effects are hard to show. However, the traditional geometric modeling method has complex modeling process and poor realism. In this article, Gan (Generative Adversarial Networks) model in AI (Artificial Intelligence) is applied to the information extraction of AR (Augmented Reality) animation images, and the modeling process of 3D animation CAD is optimized. In the information extraction model of AR 3D animation image based on GAN, a MFFM (Multi-Feature Fusion Module) is introduced into the traditional generator, which combines the characteristics of different expansion convolution rates. Moreover, the loss function of feature reconstruction based on perception is introduced into the generator, which is convenient for the network to extract more abundant features. In order to verify the reliability of this method, the GAN constructed in this article is compared with several other neural network models. Simulation results show that compared with the classic CNN model and the classic BPNN model, the error of the proposed GAN model is the lowest, and finally it is stable at about 0.54. And the accuracy of the network model is excellent, which can basically reach more than 90% and the highest can reach 95.24%. This shows that the research on information extraction and modeling of AR 3D animation images in this article has achieved good results.

Keywords: Artificial Intelligence; Generative Adversarial Network; Computer Aided Design; Augmented Reality; Three-Dimensional Animation; Image Information Extraction
DOI: https://doi.org/10.14733/cadaps.2024.S3.77-91

1 INTRODUCTION

AR is a synthesis technology that seamlessly integrates virtual objects into images or videos. AR technology, a technology of superimposing virtual images in reality, can enhance the sci-fi scenes and animation scenes seen in movies into real life. Aberman et al. [1] analyzed and constructed a motion style-based animation data framework. It reconstructed and tracked the motion style transformation of animated characters in 3D. Through centralized learning and observation driven by different data, it analyzed the transfer of training style in 3D sports stylization. It is necessary to preprocess the input video in order to extract keyframes and key actions from the video. Computer vision algorithms and motion estimation techniques can be used to automatically extract keyframes and key actions from videos for subsequent unpaired motion style transformations. Use CAD software or animation modeling tools to create animation models, and adjust and edit them based on the keyframes and key actions in the video. After the animation model is created, it is necessary to perform a non-paired motion style conversion. Deep learning techniques or style conversion algorithms can be used to convert the motion style of the animation model into the target motion style. This may require adjustments and modifications to the joints, bones, muscles, etc. of the animation model to achieve the conversion of unpaired motion styles. After non paired sports style conversion, it is necessary to render and optimize the animation. CAD software or animation rendering tools can be used to render and optimize animations to achieve more realistic and smooth animation effects. Looking at AR from another angle, the purpose of embedding virtual objects into the real scene is to enhance the understanding of the real world. Digital animation plays an important role in visual presentation in the application of AR technology, and they are inseparable. Bao [2] constructed and analyzed a breakthrough design method for rendering game animations using virtual engines. It combines many professional virtual animation engines to obtain high-quality images and improves the frame rate based on real-world 3D simulation inference. The bone and skin animation technology it adopts plays a key role in the playback of character animation. Intelligent algorithms can be applied to the planning and adjustment of object motion. Through optimization algorithms and machine learning techniques, the trajectory and speed of object motion can be automatically planned and adjusted. These algorithms can automatically adjust the motion mode and rhythm of objects according to the designer's requirements, making the movement of objects more natural and smoother. Intelligent algorithms can be applied to scene transformation and adjustment. Through machine learning and deep learning techniques, elements in the scene can be automatically recognized and adjusted, such as buildings, trees, roads, etc. These algorithms can automatically adjust the position, proportion, and color of elements in the scene according to the designer's requirements, making the scene more realistic and vivid. The application of digital animation enriches the visual form of AR technology and improves the friendliness and interest of interaction. There are many 3D modeling methods in AR 3D animation. The establishment of a 3D model can be obtained by several methods. However, due to the limitation of real-time computer and screen size, the traditional 3D roaming demonstration system can only display part of a certain area at the same time, which makes the observer lack of understanding of the overall situation. Ben et al. [3] conducted quality control analysis on automated image detection in the aviation industry. Firstly, it is necessary to prepare a 3D CAD model of the aviation machinery components. These models can be created or imported into existing model files through CAD software. Ensure model accuracy and accuracy for subsequent automatic detection. Before preparing for automatic detection, it is necessary to prepare real 2D animated images. These images can be generated through animation software, or captured or scanned from actual aviation machinery components. Ensure image quality and accuracy for subsequent automatic detection. Match 3D CAD models with 2D animated images for comparison and analysis during automatic detection. You can use the matching function in Computer-aided design software or special image processing software to complete this step. Analyze and process aviation machinery components based on the results of automatic detection. If abnormal or non-compliant situations are found, corresponding corrective measures can be taken, such as modifying CAD models, adjusting image parameters, etc. After completing automatic detection, the detection results can be organized and reported, including detailed explanations and suggestions for abnormal situations. At the same time, subsequent processing and optimization can also be carried out to improve the quality and reliability of aviation machinery components. Moreover, traditional 3D modeling and model rendering need to configure model materials, lighting and other elements, and then get the results through complex operations, which often takes a long time and requires high computer hardware. With the extensive use of 3D models in animation, machinery, medical care and other fields, Internet technology and the substantial improvement of graphic display hardware, many 3DCAD model products have been accumulated. Because the production of 3D animation pursues accuracy, authenticity and infinite operability, it is necessary to comprehensively consider and choose a more suitable modeling method when making 3D animation. How to quickly and accurately establish the AR 3D animation scene and two-dimensional electronic map is also the core problem of establishing the model.

Bodini [4] carried out face video analysis and recognition based on Computer-aided design 3D images. Through data analysis and investigation of deep learning, it has constructed performance learning indicators for 3D facial recognition reconstruction. CNN is a commonly used technique for image processing in deep learning. It can extract image features by performing convolution operations on the image. In facial feature extraction, CNN can be used to extract facial features in images, such as eyes, nose, mouth, etc. They can extract the target area from the image, classify and locate it. In facial feature extraction, these algorithms can be used to detect the position and size of facial features. Extract the position and size of the target from the entire image without the need for target candidates first. In facial feature extraction, these algorithms can be used to detect the position and size of facial features. Cao et al. [5] conducted a dynamic reference analysis on the animation industry. It conducted CAD virtual assisted simulation of cartoon animation scene aided design and proposed a material image stylization feature based on deep learning. A large number of "moe" style cartoon images need to be collected and preprocessed. The preprocessing steps may include image scaling, cropping, rotation, brightness adjustment, etc., in order to ensure that all images have a unified size and color. For each "moe" style cartoon image, its features need to be extracted for subsequent emotion classification. Deep learning models such as Convolutional neural network (CNN) or other models can be considered for feature extraction. GAN in AI is a popular generation model based on deep learning. It is a kind of neural network that uses unsupervised learning method. It does not need to process the input data in a complicated way, and is very suitable for processing the data with large amount of data and incomplete information. The original GAN model generally consists of two neural networks: one is the generating network used to reproduce the image style of the training data set, and the other is the discriminating network used to evaluate a sample from the training data set. GAN uses the confrontation training of generator and discriminator to make the generated false graph distribution close to the real data distribution. The generating network takes the noise that obeys Gaussian distribution as input and learns the true distribution of sample data. The ultimate goal is to make the judging network unable to judge whether the image comes from real data or generated data. GAN can not only generate a complex model, but also avoid the problem that the parameters of mixed Gaussian model are not enough to accurately describe the model, and it is not necessary to resample the samples by Markov chain, thus avoiding complex probability approximation. Moreover, it can better capture the implicit association of data distribution and generate samples with better quality. Based on GAN model and CAD theory, this article studies the information extraction and modeling of AR 3D animation images in detail. Its innovations are as follows:

⊖ In this article, the GAN model in AI is applied to the information extraction of AR animation images, and the modeling process of 3D animation CAD is optimized. It has certain theoretical and practical significance.

⊜ In this article, an MFFM is introduced into the traditional generator, which combines the features of different extended convolution rates, and at the same time, a feature reconstruction loss function based on perception is introduced into the generator, which is convenient for the network to extract more abundant features.

⊛ The images generated by ⊛GAN are more realistic, but the way of confrontation training is very unstable and prone to pattern collapse. Therefore, in the training of local discriminators, Euler distance based on feature variance is introduced as the loss function of perceptual feature matching, which facilitates the extraction of image edge information and enhances the structural consistency of 3D animation images. Moreover, the risk penalty term is introduced into the anti-loss function to satisfy the Lipschitz continuity condition, so that the network can converge quickly and stably during training.

The specific parts discussed in this article include: the first section, the introduction part explains the research background of information extraction and modeling of 3D animation images. The second section is the literature review. The third section is the method. This section discusses the related concepts and algorithms of GAN model and CAD technology in detail. An AR animation image information extraction model is designed and implemented, and the 3D animation CAD modeling process is optimized. The fourth section is the experimental analysis part, which simulates, trains and evaluates the performance of the AR 3D animation image information extraction model based on GAN. The fifth section is the conclusion and research prospect.

2 RELATED WORK

Ding and Li [6] analyzed the speed and accuracy of 3D animation pose recognition based on the improved depth Convolutional neural network. By recognizing gestures and postures in a 3D virtual simulation environment, it constructed the database required for character animation. The depth Convolutional neural network (CNN) is used for feature extraction. When training CNN, pre trained models or randomly initialized models can be used to determine the number of layers and parameters of the network based on the size of the dataset. In the feature extraction stage, multiple convolution operations, pooling operations, and batch normalization operations can be used to improve the performance of the network. In order to improve the generalization ability of the network, data augmentation techniques can be used to increase the diversity of training data. For example, random rotation, scaling, flipping, and other operations can be performed on input data to increase the robustness of the network to different postures. Generally speaking, the geometric construction of models for non-destructive animation evaluation is helpful for digital model imaging of physical environments. Feng et al. [7] analyzed a refined geometric reconstruction method for 3D facial shapes. It controls the expression details under 3D image supervision by evaluating the training model of 3D facial shape. It has prepared a database containing CAD facial models. These models should contain detailed 3D facial geometric structure and texture information, and require preprocessing and annotation for subsequent model learning and animation generation. For the restored 3D face model, some existing animation generation algorithms can be used, such as animation generation based on the Skeleton, animation generation based on the Muscular system, etc., to generate an animatable detailed 3D face model. Learning detailed 3D face models that can be animated from CAD images requires certain computer graphics and in-depth learning knowledge, and requires the use of some professional algorithm libraries and software. At the same time, it is also necessary to continuously iterate and optimize to adapt to different scenarios and requirements. Holland et al. [8] simulated and analyzed the data challenge of digital CAD animation based on natural light to Geometric modeling. It has developed an animated open-source library for pulse thermal imaging. After extracting features, it is necessary to conduct in-depth analysis of these data to determine the performance and quality of the animation model. This may include Data and information visualization, data clustering, data dimensionality reduction and other analysis methods, as well as machine learning and deep learning technologies. Based on the results of data analysis, a model can be established to predict the performance and quality of animation models. This may include establishing regression models, classification models, or clustering models to predict the various indicators and performance of animation models. The process of analyzing and modeling animation non-destructive testing data can be achieved in a 3D CAD environment. This helps to improve the quality and performance of animation models, as well as reduce risks and costs in the animation production process.

Jin and Yang [9] used the method of Computer-aided design to conduct case art design research on environmental art teaching. It analyzes the color connotations and requirements of computer art environment design. By comparing the artistic information models in the 3D model, a design concept of artistic color was constructed. Computer-aided design software provides a variety of visualization tools, which can enable students and teachers to better understand and present the design. For example, students can use software to create 3D models and use virtual reality technology for interactive design. Jing and Song [10] conducted continuous image analysis of the virtual animation world through computer 3D animation. By analyzing the design of animated characters in 3D images, a virtual character motion trajectory with specific materials as the model was constructed, which improved the education system of 3D animation. Compared with traditional methods, its differentiation in shaping three-dimensional animated characters has been demonstrated. By utilizing 3D reality technology, realistic 3D models can be created in virtual space, including various forms of objects, scenes, and characters. These models can be accurately edited and modified through CAD software to meet the needs of animation design. 3D reality technology can simulate materials and textures in the real world, making virtual objects appear more realistic. Through CAD software, the materials and textures of objects can be finely adjusted and edited to achieve the visual effect of animation design. 3D reality technology can simulate real-world lighting and environmental effects, making virtual scenes more realistic. Through CAD software, lighting and environmental effects can be controlled and adjusted to achieve the atmosphere and visual effects of animation design. The approximate point algorithm for curves and surfaces has very high requirements for studying the design of animation differentiation. Li [11] analyzed the approximation domain interpolation analysis under the development of free form surfaces. It constructs animation instance analysis methods under different conditions. In the motion path design of animated characters, cubic B-spline curves can be used to create smooth and continuous motion paths. By adjusting the control points of the cubic B-spline curve, you can control the shape and bending degree of the motion path, so as to realize the natural transition and change of the character in the motion process. In the modeling of animation scenes, you can use cubic B-spline curves to create camera tracks. By adjusting the control points of the cubic B-spline curve, you can control the trajectory and motion speed of the camera, so as to achieve smooth transition and change of the camera in the scene. By adjusting control points and nodes, curve optimization and adjustment can be achieved, making the curve more in line with design requirements and improving the quality and effect of animation scenes. Liu and Yang [12] analyzed and created a computer-assisted teaching model for animation art. By debugging and deploying an open network model database, advanced technologies and tools such as artificial intelligence, virtual reality, augmented reality, etc. are introduced to provide students with broader innovation space and richer learning resources. These technologies and tools can help students achieve more unique artistic effects and animation performance. Through the combination of practice and theory, students can learn the principles and theories of art animation while learning Computer-aided design. This can make students more purposeful and targeted in practice, and better utilize their creativity and innovation abilities.

Lu et al. [13] constructed an animation driver based on audio signal output to simulate a realistic speech space for network target projection. Through the conditional generation of the regression Statistical model, the head pose image under the CAD simulation prediction is rendered realistically. Texture mapping is applied to the head model to provide more realistic details such as skin and hair. Texture mapping can be done using some image processing software or CAD software. By setting different light sources and shadows, the head model has a more realistic appearance and shadow effects. The current computer graphics cannot be very realistic face animation model creation. Paier et al. [14] analyzed the granularity deformation and realistic rendering visual construction of small parameter animations. It conducts a hybrid framework flexibility analysis on the detailed actions of dynamic textures, expressing a potential dynamic texture sequence recognition for three-dimensional faces. A large amount of facial animation data needs to be collected, including facial expressions, speech patterns, etc. Then, it is necessary to preprocess the data, such as facial marker point detection and tracking, and speech signal preprocessing. For each facial animation data segment, it is necessary to extract its features for subsequent deep neural network modeling. Convolutional neural network (CNN) or other models can be considered for feature extraction. Based on the results of feature extraction, a deep neural network model can be trained to achieve interactive facial animation generation. The model can be optimized, such as adjusting its parameters, structure, etc., to improve its accuracy and robustness. For new facial expressions or speech patterns, a trained deep neural network model can be used to generate corresponding animations. In CAD image statistics, color concept association is an important research direction. By estimating the correlation between color concepts, it is possible to better understand the relationship between different colors in images and concepts such as objects, scenes, and emotions. Rathore et al. [15] analyzed the visual color matching semantic analysis effect of image models. After image segmentation, color feature extraction is performed on each segmented area. Some color spaces, such as RGB, HSV, etc., can be used to describe the color features of each region. At the same time, some color filters, such as average filter, Median filter, can also be used to extract the color features of each region. After extracting the color features of each region, color concept correlation analysis can be performed. Some machine learning algorithms can be used to establish correlation models between colors and concepts. Some Feature selection methods, such as variance-based analysis and correlation-based analysis, can be used to select the features most related to the color concept. Wang et al. [16] used the computer Assistive technology of 3D model to analyze and study the limb animation tomography. Through the exponential analysis of tomography, it determined the precise printing results of computer-aided modeling. Based on surgical planning and design, 3D printing technology can be used to create a solid model of the toe model, in order to more accurately simulate the surgical process and effects. This model can include tissue structures such as bones, tendons, blood vessels, and nerves to provide more accurate surgical simulations. Zhang et al. [17] proposed a visual information image network analysis architecture. By analyzing the robustness training data of visualized two-dimensional images, it compensates for the visual missing points generated by two-dimensional layout. It designs an encoder network that converts visual images into hidden vector representations. The encoder can be a Convolutional neural network (CNN) or other types of neural network, which is used to extract the features of visual images. These features can be a set of continuous numerical vectors used to represent the content and structure of an image. For visualized images embedded with information, decoder networks are used for decoding and reconstruction. The decoder network can decode the fused hidden vectors into reconstructed visual images and display the embedded information. The encoder and decoder are trained and optimized using standard optimization algorithms, such as random gradient descent (SGD) or Adaptive learning rate. The Loss function of the reconstructed visual image can be calculated, such as Mean squared error (MSE) or Cross entropy, and these functions can be used to optimize the parameters of the network.

Based on the in-depth discussion of previous literatures, this article proposes and constructs a new AR animation image information extraction model by using the generative confrontation network and CAD method in AI, and optimizes the 3D animation CAD modeling process. Simulation shows that this method has achieved good results.

3 INFORMATION EXTRACTION AND MODELING OF AR 3D ANIMATION IMAGES

3.1 GAN-Based AR Animation Image Information Extraction Modeling

GAN model in AI has a good performance in semi-supervised learning. Moreover, compared with the traditional machine learning algorithm, GAN model has more powerful feature representation and learning ability, and can mine deeper information of images, and has achieved great success in various fields of image processing in recent years. As a generation model, GAN hopes to gain the ability to generate samples that are fake and genuine through learning. However, the training of GAN is very unstable, and the optimization of discriminator is much easier than that of generator, that is to say, discriminator can distinguish true graph and false graph very easily, but generator is difficult to generate false graph close to true graph. In the GAN model, the generating network transforms the noise image into the target image through deconvolution, and the ultimate goal is to generate an image that can "fool" the discriminant network. The discriminant network uses a CNN to identify image features and judge whether it is true or false. The ultimate goal is to continuously improve the discriminant ability and correctly identify whether the image is from the training data set or "forged". When the input data comes from real training data, it is judged that the expected output of the network is relatively high. When the generated data is input to the discriminant network, it is expected that the discriminant model will output a low probability, while the discriminant network will output a high probability, thus forming competition and confrontation. In order to simulate the activation frequency function of brain neurons proposed in the field of neuroscience, some researchers have proposed Softplus function and modified linear element, as shown in Figure 1.

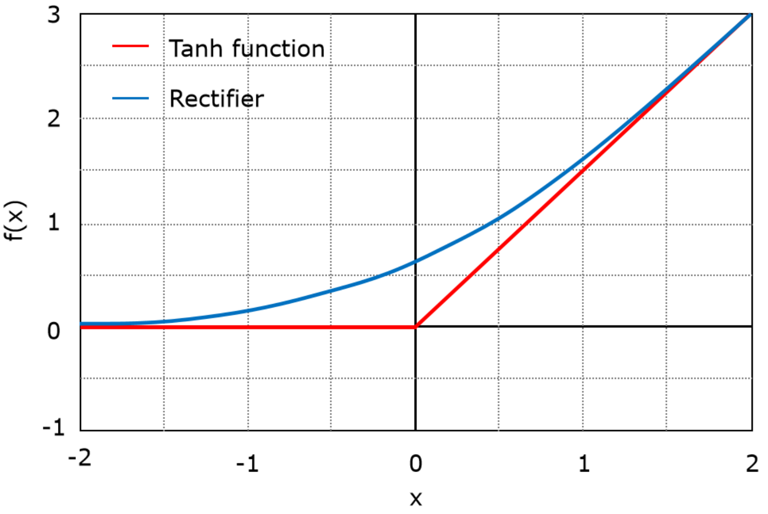

Figure 1: Softplus function and modified linear element.

Softplus function has the first two characteristics of biological brain neuron activation function, but it does not have sparse activation characteristics. The modified linear element is used as the approximation of Softplus function. In this network model, the fully connected layer is removed from the convolution feature, and the global average pooling is helpful to the stability of the model, but it slows down the convergence speed. Both discriminant network and generative network use batch standardization to solve the training problem caused by poor parameter initialization, which makes the gradient spread further. In the generated network, except for the output layer, Tanh activation function is used, and all other layers use ReLU activation function, which makes the model converge faster. LeakyRelu is used as the activation function of discriminant network. The generator is generally composed of multi-layer neural network, which reconstructs the noise obtained by random sampling in some way into a false picture consistent with the sample dimension. The generator in the task of image conversion generally adopts the structure of encoder and decoder, that is, the features are obtained by continuous down-sampling through the encoder, and then the images are obtained by up-sampling through the decoder, which easily leads to the low quality of the generated images. The usual solution is to add cross-layer connection between encoder and decoder, which can enhance the image quality. The original GAN unsupervised learning method is just like copying the target image on white paper, and the result error is large. However, the GAN model with generation constraints becomes supervised learning, which makes up for this defect. The GAN model with generation constraints is transformed from a specific base map to a target image, which is equivalent to filling colors according to the existing wire draft, and the generation result is more controllable. Therefore, using this feature, it is possible to realize 3D animation image conversion through GAN.

Semi-supervised learning of GAN is mainly realized by discriminator, which not only considers the probability that input samples belong to real samples, but also considers the probability of labeled input samples to label categories. The traditional generator uses a convolutional encoder as the generator model, and its ability to extract features is limited, mainly by expanding convolution kernel or expanding convolution. This will lead to some problems, such as: the kernel of extended convolution is sparse, so many pixels used for calculation are skipped; However, the use of large convolution kernel introduces a large quantity of model parameters, which will lead to a waste of a lot of computing resources. In this article, an MFFM is introduced into the traditional generator, which combines the features of different extended convolution rates, and at the same time, a feature reconstruction loss function based on perception is introduced into the generator, which is convenient for the network to extract more abundant features. The discriminator formula in this article is as follows:

$$f = F(x;\varnothing_r) \tag{1}$$

$$D_{\varnothing}(s) = sigmoid(\varnothing_1, F(s;\varnothing_r)) = sigmoid(\varnothing_1, f) \tag{2}$$

The generator formula is as follows:

$$g, h_t^M = M\left(f_t, h_{t-1}^M; \theta_m\right) \tag{3}$$

$$O_t, h_t^W = W\left(x_t, h_{t-1}^w; \theta_W\right) \tag{4}$$

Set m and n equal when training with mini-batch. Due to the constraint of optimal transmission quality conservation, the transmission matrix T needs to meet the following conditions:

$$\sum_{j=1}^{n} T_{ij} = \frac{1}{m} \tag{5}$$

$$\sum_{i=1}^{m} T_{ij} = \frac{1}{n} \tag{6}$$

The elements in the transmission matrix T are a probability distribution of transmission, in which the sum of all elements is 1. The loss function of feature reconstruction introduced into the generator is shown in the following formula:

$$L_{Gmf} = \sum_{p=1}^{5} \omega^p \frac{\left\| \theta_{I_{output}}^p - \theta_{real}^p \right\|_2^2}{N_{\theta_{I_{real}}^p}} \tag{7}$$

Where: $\theta_{I_*}^p$ is the activation diagram of relup_1 layer given the original input I_*; $N_{\theta_{I_{real}}^p}$ represents the quantity of elements of $\theta_{I_{real}}^p$; p represents the quantity of layers where the active feature map is located; I expression input. In order to introduce the unsupervised mean square error loss of the model in different States, the parameters of different outputs predicted by the same input are punished. Moreover, the SENet module is added to the discriminator to automatically learn the importance of each feature channel and perform feature recalibration, thus improving the performance of semi-supervised learning.

The original generating network maximizes the probability that the generated 3D animation image is a real image. Feature matching requires that the generated data match the statistical features of the real data as much as possible, and the discriminant network is used to specify the statistical features that need to be matched. In the implementation, the expected value of network matching and distinguishing the characteristics of network middle layer is generated by training. The framework of 3D animation image information extraction model mainly includes a multi-feature fusion block generator with perceptual loss, a local discriminator with perceptual loss, a down-sampled 2 times multi-scale discriminator, a down-sampled 4 times multi-scale discriminator and a global discriminator. The information extraction model framework of AR 3D animation image based on GAN is shown in Figure 2.

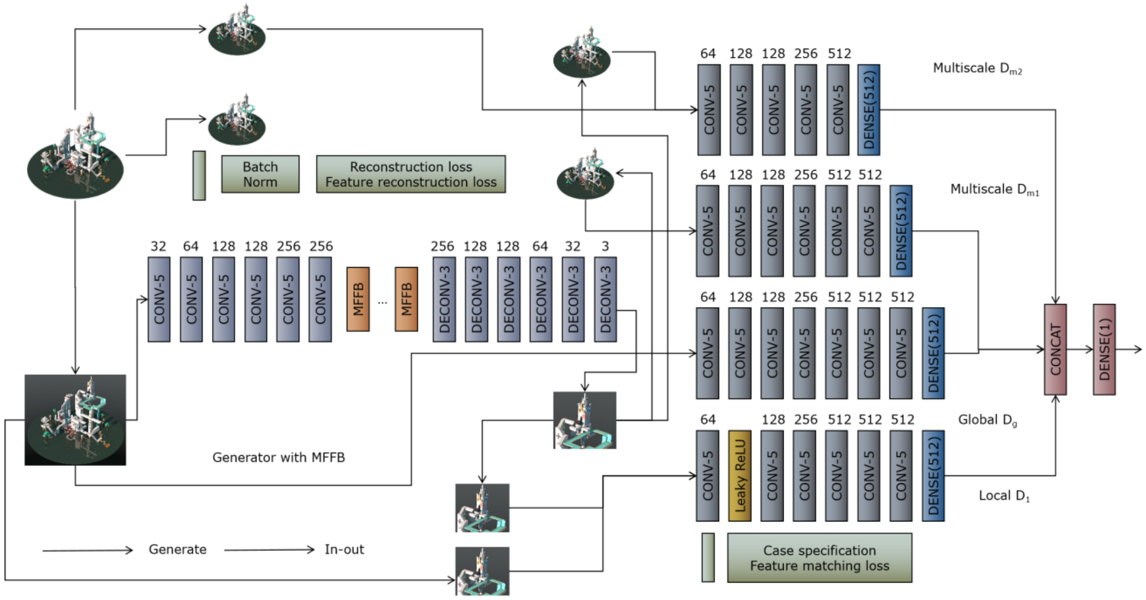

Figure 2: GAN-based information extraction model framework for AR 3D animation images.

Because the whole model includes two stages, the model consists of two GAN's, which are connected in series by finite state machine. Finally, the model includes two pairs of generators and two pairs of discriminators, and the model is trained by paired images. The input of the discriminator includes generated samples and real samples, and the real samples are mainly unlabeled samples and limited labeled samples. Labeled samples only participate in the supervised loss of the discriminator, while unlabeled samples and generated samples participate in the unsupervised confrontation loss and unsupervised mean square error loss. A source domain classifier is added to the discriminator to judge which domain the image is converted from. Moreover, the generator generates a conversion map from which the source domain cannot be judged, so as to remove the source domain information in the generated map, which enables the generator to completely transfer the image to the target domain. GAN uses deconvolution layer for up-sampling, convolution layer instead of pooling layer to keep the image scale unchanged, cancels the full connection layer to reduce the calculation amount, and adds batch normalization to improve the convergence speed of the network. In the GAN-based AR 3D animation image information extraction model, the generator is a repair network, which consists of six layers of encoders and six layers of decoders.

3.2 Optimization of 3D Animation CAD Modeling Process

CAD model has strict requirements on accuracy, and it is generally represented by Brep model. Brep model is also called boundary model. A Brep model can be regarded as the basic element of boundary representation composed of three kinds of geometric information-face, edge, vertex and topological information related to them. In order to realize mutual drive, coordinate transformation between two-dimensional GIS subsystem and three-dimensional visual subsystem is necessary. The two-dimensional GIS subsystem uses the two-dimensional plane coordinate system, while the three-dimensional scene subsystem uses the three-dimensional right-hand coordinate system adopted by Open GVS. Two-dimensional GIS and 3D coordinate system are shown in Figure 3.

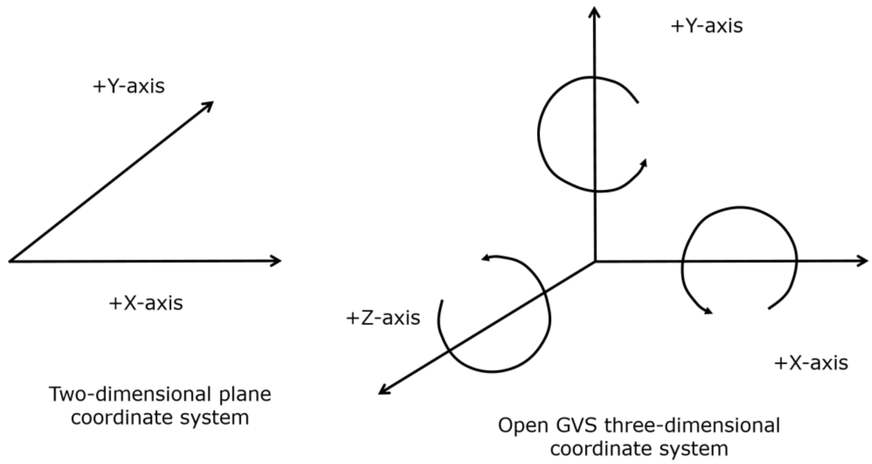

Figure 3: 2D GIS and 3D coordinate system.

In order to describe the internal features and specific relationships of 3D models, the features of 3D models can be expressed as a vertex attribute of the attributed feature adjacency graph, and the internal feature relationships of 3D models can be expressed as the relationship between vertices in the graph. In the 3D scene system, if a central point in the scene is selected as the origin of the 3D scene coordinate system, the coordinate transformation relationship between the viewpoint in the 3D scene and the coordinate in the 2D GIS is shown in the following formula:

$$\begin{aligned} X' &= (X - OffsetX)/1000 \\ Y' &= 1.7 \\ Z' &= -(Y - OffsetY)/1000 \end{aligned} \tag{8}$$

Where X' is X coordinate in the coordinate system of 3D visual system; Y' is the Y coordinate in the coordinate system of 3D visual system; Z' is the Z coordinate in the coordinate system of 3D visual system; X coordinates of $OffsetX$ custom center point; $OffsetY$ defines the Y coordinates of the center point. Solid Works software has rich software interfaces and powerful application functions, and can provide various converters for input and output functions, and can output various software file formats. Modeling by Solid Works is based on solid modeling, which can accurately describe the contour line, surface and volume of the workpiece, so its 3D solid model contains information such as volume and surface. 3DS Max is based on surface modeling. Solid Works software model can't be directly imported into VA software for interactive design, but it can be realized by third-party software, while 3DS Max can. Therefore, this article combines the two software, uses Solid Works for modeling and assembly, and then uses 3DS Max for rendering and AR animation simulation. When it is necessary to record and play 3D animation, first render the animation as a video audio file, so that general video software can recognize and watch it. For screen recording function, plug-ins need to be installed, and plug-ins are needed to start the recording function, and attention should be paid to the file format.

The model uses two stages of learning to generate clear and high-quality AR 3D animation images. The main idea is to generate low-resolution images, then generate high-resolution images from low-resolution images, and finally generate clear and detailed AR 3D animation images. In the training process, due to the existence of the regular term of the network model dropout and random noise, the network output will change. Based on this, this article considers that the same data will be randomly processed in the same batch iteration, and its output will also change. In order to guide the direction of model parameter optimization, it is hoped that the output from the same input will be as same as possible, that is, the probability of belonging to a certain category is as close as possible for prediction classification. In addition, in the training of local discriminators, Euler distance based on feature variance is introduced as the loss function of perceptual feature matching, which facilitates the extraction of image edge information and enhances the structural consistency of 3D animation images. Moreover, the risk penalty term is introduced into the anti-loss function to satisfy the Lipschitz continuity condition, so that the network can converge quickly and stably during training.

4 GAN MODEL TRAINING AND RESULT ANALYSIS

PyTorch is a Python software package, which provides two functions: using GPU to accelerate the calculation of tensor, and constructing GAN based on tape to automatically solve gradient system. PyTorch provides a tensor that supports both CPU and GPU computing, which can speed up the computing process. The experiment in this section is implemented by using the open source framework pytorch. The generator of the experiment is based on GAN generator, and the discriminator introduces SENet module to increase the nonlinearity of the model WN, BN and Dropout strategies are also used in the model. The parameter configuration of the network model in the experiment is shown in Table 1.

Discriminator	*Builder*
Input 32×32 Three-channel label y∈R10	Input Noise∈R100
Dropout =0.2 3×3 Conv, 64, lReLU, WN 3×3 Conv, 64, lReLU, WN 3×3 Conv, 64, lReLU, WN	MLP 8192 units ReLU, BN Reshape 512×4×4 5×5 DeConv, 256, stride 4
Dropout =0.2 3×3 Conv, 128, lReLU, WN 3×3 Conv, 128, lReLU, WN 3×3 Conv, 128, lReLU, WN 3×3 Conv, 128, lReLU, WN SENet block	5×5 DeConv, 128, stride 4 ReLU, BN
Dropout =0.2 3×3 Conv, 256, lReLU, WN NIN, 128, lReLU, WN NIN, 128, lReLU, WN NIN, 128, lReLU, WN Global pool layer Denselayer 10 units with WN	5×5 DeConv, 3, stride 4， Tanh, WN

Table 1: Parameter configuration of network model.

The experiment is mainly divided into two processes. In the training process, the network trained 500 groups of learning samples of "contour map-line sketch map" and "line sketch map-effect map" for 200 rounds in stage 1 and stage 2 respectively. The former of each group of images is the constraint image and the latter is the target image. Because the 3D animation image generated by the multi-feature fusion generator with MFFM is fuzzy, in order to make the generated 3D animation image more realistic, it is necessary to get the 3D animation image that meets the target through confrontation training. Therefore, it is necessary to introduce a network-discriminator to identify the true and false images in the experiment to supervise and guide the generation of images. Figure 4 is the root mean square error of the training sample.

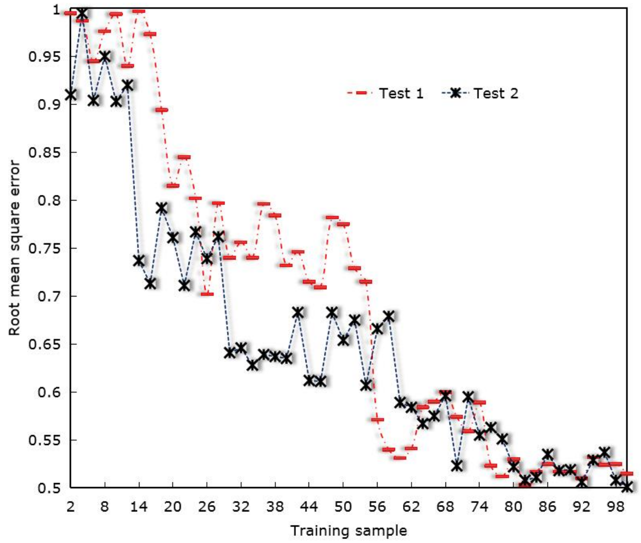

Figure 4: Root-mean square error of training samples.

From the training results of Figure 4, it can be seen that the root mean square error of the curve decreases very quickly at the beginning of training. After training, the sub-networks are integrated by finite state machine. The finite state machine hides the conversion process of the line drawing and simplifies the operation of the test process. In the test process, the target image can be output only by inputting the test sample.

In all experiments, the learning rate of GAN model is initialized to 0.001. In the process of learning, the learning rate becomes 10% every 10000 iterations. The weight of convolution layer is initialized to orthogonal matrix, and the weight of linear layer is also initialized using orthogonal matrix. In this article, Adm optimization algorithm is used to update the weights, and the size of mini-batch is 128. The training of the model is divided into three steps: the first step is to train the generator. Step 2: Train the discriminator. Step 3: The generator and the discriminator conduct confrontation training. In the process of training, the generator and the discriminator are trained alternately, and the generator trains the discriminator every five times. The generator has modified different data sets on the basis of GAN. The classic CNN model, the classic BPNN model and the proposed GAN model are compared, and the training effects of different algorithms are shown in Figure 5.

With the increase of training times, the error rates of the three network models are getting lower and lower. Compared with the classic CNN model and the classic BPNN model, the training error rate of GAN model is the lowest, and finally it is stable at around 0.05.

In this section, when training local discriminators, Euler distance based on feature variance is introduced as the loss function of perceptual feature matching, which facilitates the extraction of image edge information and enhances the structural consistency of 3D animation images. Moreover, the risk penalty term is introduced into the anti-loss function to satisfy the Lipschitz continuity condition, so that the network can converge quickly and stably during training.

The training process inputs the images in the learning sample set. Because GAN is a generative model, the output image of each training is not unique, and the Loss value of GAN model cannot converge gradually like that of classic CNN, so the training effect can only be reflected by directly observing the difference between the generated image and the target image.

The animation image renderings constructed by "outline drawing-line drawing-renderings" show a good training effect, and the renderings output by the network are very close to the learned target images. The classic CNN model, the classic BPNN model and the proposed GAN model are tested in the test samples respectively. Figure 6 shows the error situation of the model in the test sample, and Figure 7 shows the accuracy of the model in the test sample.

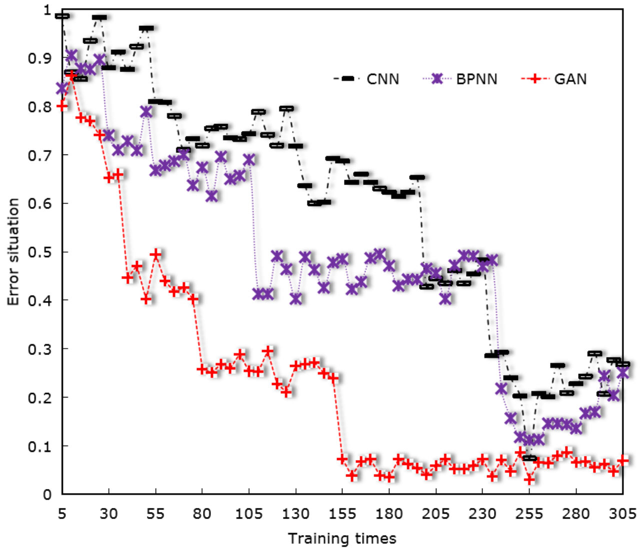

Figure 5: Training effect diagram of different models.

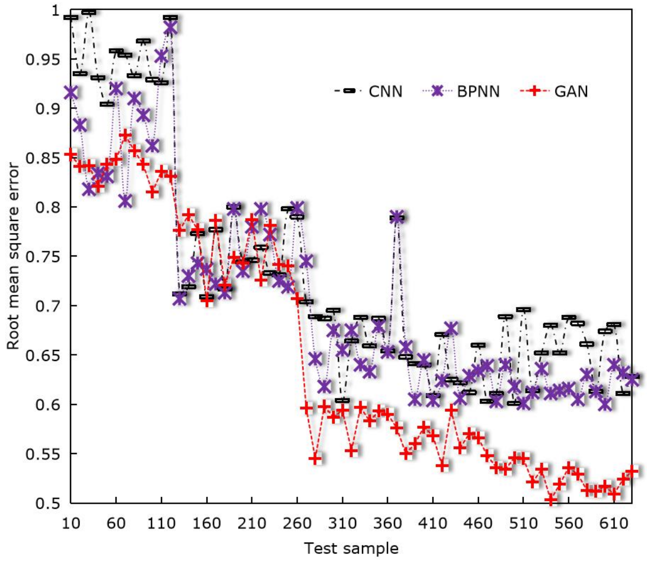

Figure 6: Error of the model in the test sample.

From the test results of Figure 6, it can be seen that the error of the classic CNN model is high in the test samples, and it finally stabilizes at around 0.68; The error of the classical BPNN model is in the middle, and it finally stabilizes at around 0.63; The error of GAN model is the lowest, and it finally stabilizes at about 0.54.

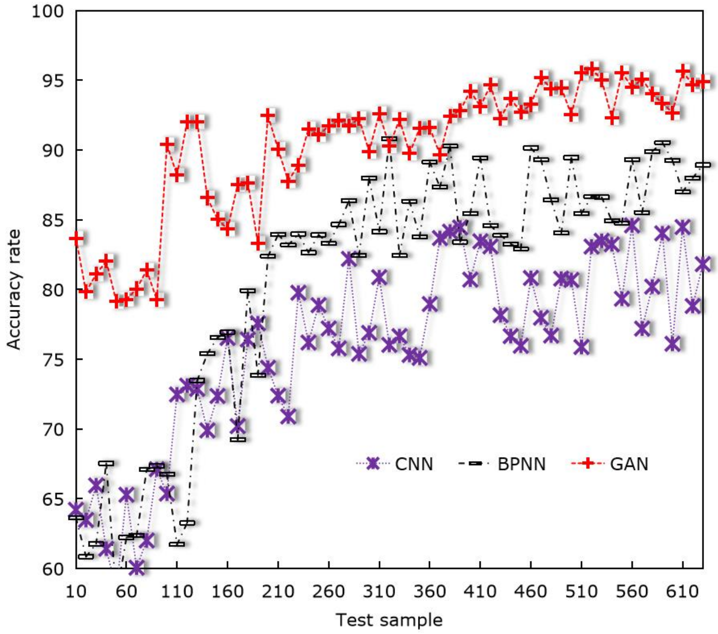

Figure 7: The accuracy of the model in the test sample.

The accuracy of the classic CNN model is not ideal, which is about 82%. The accuracy of classical BPNN model is higher than that of CNN model, which is about 89%. Among them, the accuracy of GAN model is the best, which can reach more than 90% and the highest can reach 95.24%. On the whole, the accuracy of GAN model is 6~13% higher than that of the other two neural network models.

5 CONCLUSIONS

Digital animation is an important presentation mode of AR, which has the most intuitive impact on the user experience, interacts with the real world and produces action response. AR 3D animation makes animation no longer limited to limited screen display, and it expands the possibilities of digital animation in real scenes. This article summarizes the application of 3D animation based on AR technology. On this basis, the GAN model in AI is applied to the information extraction of AR animation images, and the modeling process of 3D animation CAD is optimized. Using existing CAD data to build 3D model and 2D electronic map of the scene has the advantages of fast and accurate modeling. In addition, in the training of local discriminators, Euler distance based on feature variance is introduced as the loss function of perceptual feature matching, which facilitates the extraction of image edge information and enhances the structural consistency of 3D animation images. Moreover, the risk penalty term is introduced into the anti-loss function to satisfy the Lipschitz continuity condition, so that the network can converge quickly and stably during training. Finally, the simulation shows that the accuracy of this network is the best, which can reach more than 90% and the highest can reach 95.24%. Generally speaking, the accuracy of GAN model is 6~13% higher than that of the other two neural network models. This shows that the research on image information extraction and modeling of AR 3D animation has achieved good results in order to further promote the development of AR 3D animation.

Xiaoxuan Tian, https://orcid.org/0009-0008-5543-9631
Chengkun Li, https://orcid.org/0009-0000-7894-3680

REFERENCES

[1] Aberman, K.; Weng, Y.; Lischinski, D.; Cohen, O.-D.; Chen, B.: Unpaired motion style transfer from video to animation, ACM Transactions on Graphics (TOG), 39(4), 2020, 64:1-64:12. https://doi.org/10.1145/3386569.3392469

[2] Bao, W.: The application of intelligent algorithms in the animation design of 3D graphics engines, International Journal of Gaming and Computer-Mediated Simulations, 13(2), 2021, 26-37. https://doi.org/10.4018/IJGCMS.2021040103

[3] Ben, A.-H.; Jovančević, I.; Orteu, J.-J.; Brèthes, L.: Automatic inspection of aeronautical mechanical assemblies by matching the 3D CAD model and real 2D images, Journal of Imaging, 5(10), 2019, 81. https://doi.org/10.3390/jimaging5100081

[4] Bodini, M.: A review of facial landmark extraction in 2D images and videos using deep learning, Big Data and Cognitive Computing, 3(1), 2019, 14. https://doi.org/10.3390/bdcc3010014

[5] Cao, Q.; Zhang, W.; Zhu, Y.: Deep learning-based classification of the polar emotions of "moe"-style cartoon pictures, Tsinghua Science and Technology, 26(3), 2020, 275-286. https://doi.org/10.26599/TST.2019.9010035

[6] Ding, W.; Li, W.: High speed and accuracy of animation 3d pose recognition based on an improved deep convolution neural network, Applied Sciences, 13(13), 2023, 7566. https://doi.org/10.3390/app13137566

[7] Feng, Y.; Feng, H.; Black, M.-J.; Bolkart, T.: Learning an animatable detailed 3D face model from in-the-wild images, ACM Transactions on Graphics (ToG), 40(4), 2021, 1-13. https://doi.org/10.1145/3450626.3459936

[8] Holland, S.; Mcinnis, C.; Radkowski, R.: NDE data analysis and modeling in 3D CAD context, Materials Evaluation, 78(1), 2020, 95-103. https://doi.org/10.32548/2020.me-04095

[9] Jin, H.; Yang, J.: Using computer-aided design software in teaching environmental art design, Computer-Aided Design and Applications, 19(S1), 2021, 173-183. https://doi.org/10.14733/cadaps.2022.S1.173-183

[10] Jing, Y.; Song, Y.: Application of 3D reality technology combined with cad in animation modeling design, Computer-Aided Design and Applications, 18(S3), 2020, 164-175. https://doi.org/10.14733/cadaps.2021.S3.164-175

[11] Li, L.: Application of cubic b-spline curve in computer-aided animation design, Computer-Aided Design and Applications, 18(S1), 2020, 43-52. https://doi.org/10.14733/cadaps.2021.S1.43-52

[12] Liu, F.; Yang, K.: Exploration on the teaching mode of contemporary art computer aided design centered on creativity, Computer-Aided Design and Applications, 19(S1), 2021, 105-116. https://doi.org/10.14733/cadaps.2022.S1.105-116

[13] Lu, Y.; Chai, J.; Cao, X.: Live speech portraits: real-time photorealistic talking-head animation, ACM Transactions on Graphics (TOG), 40(6), 2021, 1-17. https://doi.org/10.1145/3478513.3480484

[14] Paier, W.; Hilsmann, A.; Eisert, P.: Interactive facial animation with deep neural networks, IET Computer Vision, 14(6), 2020, 359-369. https://doi.org/10.1049/iet-cvi.2019.0790

[15] Rathore, R.; Leggon, Z.; Lessard, L.; Schloss, K.-B.: Estimating color-concept associations from image statistics, IEEE Transactions on Visualization and Computer Graphics, 26(1), 2019, 1226-1235. https://doi.org/10.1109/TVCG.2019.2934536

[16] Wang, H.; Zhang, Z.; Zhao, J.: Three-dimensional computer-aided design modeling and printing for accurate toe-to-hand transplantation, The Journal of Hand Surgery, 48(2), 2023, 198.E1-198.E11. https://doi.org/10.1016/j.jhsa.2021.09.034

[17] Zhang, P.; Li, C.; Wang, C.: Viscode: Embedding information in visualization images using encoder-decoder network, IEEE Transactions on Visualization and Computer Graphics, 27(2), 2020, 326-336. https://doi.org/10.1109/TVCG.2020.3030343