Journal of Engineering, Project, and Production Management
2026, 16(7), 2025-308

Artistic Style Transfer Based on Generative Adversarial Networks and Its Application in Film and Animation Production

Shifeng Wang
Lecturer, School of Film, Jilin Animation Institute, Changchun 130000, China, E-mail: wangshifeng2025@163.com

Production Management
Received December 7, 2025; revised January 27, 2026; accepted July 2, 2026
Available online July 12, 2026

Abstract: Artistic style transfer technology plays an important role in film and animation production. High-quality artistic style transfer can significantly enhance the production value of animation. However, traditional methods for artistic style transfer face limitations, including low efficiency and poor results. Therefore, this study proposes a model for artistic style transfer in film and animation production based on Deep Convolutional Generative Adversarial Networks (DCGAN). The model utilizes the Feature Pyramid Network (FPN) with a skip connection structure and introduces sub-pixel convolution blocks and Batch Normalization to further improve the generator architecture. To further reduce interference from abnormal data, an improved multi-window spectral estimation and spectral subtraction method is introduced to enhance data representation, thereby enabling high-quality film and animation production. Experimental results show that the proposed algorithm achieves high adaptability and strong fitting performance, with content and style losses of 0.09 and 0.12, respectively and outperforms the comparison algorithms. The constructed model also demonstrates strong performance in artistic style classification and transfer. In two datasets, the matching degree of style features was 89.14% and 91.20%, respectively. The proposed model can maintain high-quality style transfer under various forms of noise, and the ablation study demonstrates its validity. These results indicate that the artistic style transfer model can enhance the efficiency and accuracy of style recognition while improving animation quality. This study contributes to the precise construction of artistic style transfer models for film and animation production and promotes the development of the animation industry.

Keywords: Deep convolutional generative adversarial networks; generator; artistic style transfer; animation production; multi-window spectral subtraction.

DOI 10.32738/JEPPM-2025-308

1. Introduction

With the rapid development of film and animation production, intelligent computer technology has become an important research direction in computer vision and digital image processing (Mokayed et al., 2023). Among them, artistic style transfer extracts features from content and style images to generate detailed, vivid image effects. It is efficient, diverse, and widely applied in film and animation production (Hu et al., 2022; Kang and Kim, 2024). However, traditional methods for artistic style transfer produce stylized images only by minimizing both content and style loss. They face limitations such as high computational difficulty, slow generation speed, and unnatural image effects (Ahmad et al., 2025; Aliet et al., 2025). Therefore, there is an urgent need for an artistic style transfer model that can be applied accurately and efficiently in film and animation production. With rapid advances in computer technology, intelligent, algorithm-based models for artistic style transfer have become a recent focus of research. Intelligent algorithms offer high efficiency, strong objectivity, and reduced dependence on manual intervention. Among them, Generative Adversarial Networks (GANs) learn image features through adversarial training between a generator and a discriminator, generating more detailed and vivid images. GANs provide efficient solutions for applications such as image editing and video processing. Therefore, this study optimizes the traditional GAN algorithm by incorporating deep convolutional neural networks into both the generator and the discriminator, forming a Deep Convolutional Generative Adversarial Network (DCGAN). The model also applies the Feature Pyramid Network (FPN) algorithm with skip connection structure and introduces sub-pixel convolution blocks and Batch Normalization (BN) to further improve DCGAN. In addition, Multi-Window Spectral Subtraction (MSS) is applied to enhance spectral estimation, ultimately constructing a model for artistic style transfer in film and animation production. The innovation points of this study are as follows:

1. Research and propose a DCGAN framework, and introduce the FPN algorithm, sub-pixel convolution block, and BN to improve the framework generator to enhance the efficiency and accuracy of the algorithm recognition.

2. Integrate KCF and MSS algorithms, suppress abnormal data interference, strengthen the data expression ability of artistic style features, and ensure the stability of the migration process.

3. This study will effectively integrate the improved DCGAN, KCF, and MSS feature processing modules, build a transfer model of artistic style of film and television animation production, and provide a new method for research in related fields.

2. Literature Review

GANs automatically learned deep features through adversarial training. They exhibited superior feature-analysis performance and strong capability for expressing complex signals, providing solutions to problems across multiple fields. Many scholars have conducted research on GANs. (Zhang et al., 2024) proposed a GAN for image optimization. They classified images using class-aware generative adversarial networks, restored image quality with DCGAN, and filled in missing parts using image inpainting methods, thereby optimizing image quality. (Njima et al., 2022) aimed to improve indoor positioning accuracy. They enhanced positioning precision using a weighted deep neural network and collected real and synthetic samples via a GAN to form a dataset. Weighted techniques were applied to prevent model overfitting, resulting in a GAN-based indoor positioning model. (Kiranyaz et al., 2022) addressed artifacts in Electrocardiogram (ECG) recordings and proposed a denoising model based on a cycle-consistent GAN. They incorporated a one-dimensional deep neural network into the cycle-consistent GAN, improving ECG signal quality and achieving high accuracy in ECG diagnosis. (Christophe, 2022) applied a GAN architecture to identify visual salient features and topological arrangements, transferring visual appearance and image style to other images. Using maps from multiple geospatial regions, simple planar maps were upgraded to complex-style maps, thereby constructing a neural pattern transfer model based on GANs. (Soleymanzadeh et al., 2023) tackled data imbalance in intrusion detection systems by using GANs to reduce overfitting. They introduced ensemble deep learning algorithms into GAN, constructing a generator-discriminator architecture that mitigated model collapse and improved detection accuracy.

Artistic style transfer technology plays an important role in film and animation production. Theories and applications of transfer methods have become relatively mature, and many scholars have conducted in-depth studies. (Chen et al., 2024) addressed 3D scene style transformation by constructing a neural radiance field-based 3D scene style transfer model. They improved video rendering, trained data using a 2D realistic style transfer network, and optimized hypernetwork parameters to enhance the quality and consistency of visual style transfer. At present, GAN-based style transfer technology has made significant progress in specific modes or single-task scenarios. (Krishnan, 2023) parameterized text images to obtain vector parameters and used a style-consistent GAN to classify style content. They improved image resolution and applied self-supervised training standards to select fonts and text styles, ultimately enhancing the quality of text image content and style transfer. (Garg, 2023) proposed a conditional GAN based on neural style transfer to address the reliable transfer of sensitive data. The model transferred style to image content, restored secret content, and generated steganographic images via neural style transfer, improving image quality. (Chen et al., 2022) constructed a cycle-consistent GAN to address 3D style transfer in indoor positioning. They extracted 3D visual information via style conversion, determined camera poses from it, and obtained indoor positioning coordinates using 3D style-transferred building information. (Sun et al., 2023) proposed a generative steganography model based on style transfer and deep feature fusion. They embedded information using a channel reduction network, preliminarily extracted structural features via a fusion network, and further extracted secret information via a residual network, thereby improving both quality and security.

Although the current research has made some progress in general image style migration and specific scene style optimization, there are still limitations, such as insufficient scene adaptation, limited accuracy in style feature extraction, and a migration process that is vulnerable to interference from abnormal data. Therefore, in view of the difficulties in capturing the exclusive style features of film and television animation and the unnatural fusion of style and content, this study investigates how to integrate a deep convolutional neural network into the GAN algorithm, and proposes a DCGAN algorithm, which uses FPN algorithm and sub-pixel convolution blocks to improve the DCGAN algorithm, in order to improve the ability of feature extraction and the accuracy of style migration. At the same time, to suppress abnormal data and noise interference and strengthen the data's expressive ability for artistic style features, this study introduces an improved MSS algorithm, cooperatively optimizes feature processing and migration paths, and finally constructs a film and television animation production artistic style migration model. The model aims to achieve efficient artistic style transfer, improve animation quality, and meet the demands of artistic style transfer in film and animation production.

3. Film and Animation Production Artistic Style Transfer Model Based on DCGAN

3.1. Design of Improved DCGAN Algorithm Based on Structural Optimization

With the rapid development of artificial intelligence, intelligent computer algorithms provide new technical pathways for artistic style transfer in film and animation production (Zhang and Pu, 2024). Traditional artistic style transfer algorithms have limitations, including high computational cost, loss of image details, and unnatural style (Kang et al., 2023). DCGAN reduces computational complexity through end-to-end contrastive learning and retains more texture details and a more natural style when generating images. To address these issues, this study proposes an improved DCGAN algorithm applied to artistic style transfer in film and animation production. The algorithm enhances recognition of image style and content and improves image detail. The proposed DCGAN algorithm learns image features through adversarial training between the generator and the discriminator, generates more image details, improves style transfer efficiency, and exhibits strong adaptability. The operation flow of DCGAN is shown in Fig. 1.

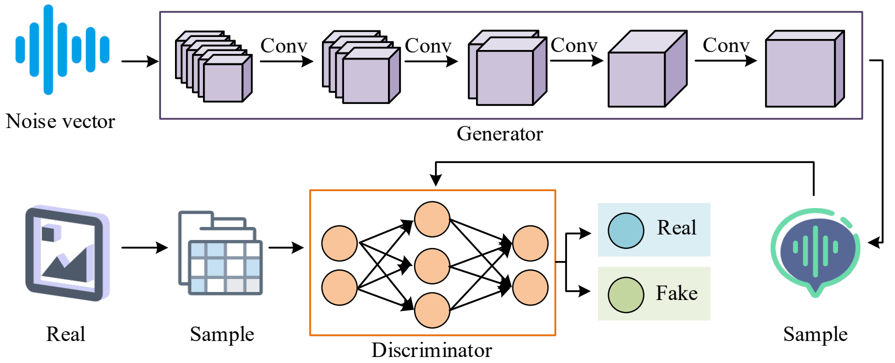


Fig. 1. Operation flow chart of DCGAN algorithm (Icon source: www.iconfont.cn)

As shown in Fig. 1, DCGAN first takes a noise vector and passes it to the generator. Four convolutional layers extract features of both image content and style to generate image samples that resemble real samples. Real data are preprocessed into samples and fed into the discriminator along with the generated image samples. The hierarchical structure of convolution + BN + activation function is adopted in the generator, and the size of a convolution kernel is set to 4×4, and the convolution step is set to 2. The discriminator distinguishes real samples from generated samples using convolutional operations and feeds its output back to the generator, further optimizing the generator's learning. By dynamically balancing the discriminator and the generator, the algorithm minimizes the objective function. The discriminator uses a hierarchical structure of convolution + LeakyReLU + BN, with a 4×4 convolutional kernel and a stride of 2. The convolution operation is mathematically expressed in Eq. (1).

$$x_i{}^l = f(W_i{}^l * X^{(l-1)} + b_i{}^l) \tag{1}$$

In Eq. (1), $x_i{}^l$ represents the i feature of the l layer, $W_i{}^l$ represents the i-th weight matrix of the l layer, f represents the activation function, X is the output, and $b_i{}^l$ represents the bias term. The adversarial loss function of DCGAN is shown in Eq. (2).

$$\underset{D}{max}\,\underset{G}{min}\,\ell(\theta_D, \theta_G) = E_{x\sim p_{data}}[log(D(x))] + E_{x\sim p_z}[log(1 - D(z))] \tag{2}$$

In Eq. (2), E represents the expected optimization value, p_{data} and p_z represent the real samples and generator-generated samples, $D(x)$ is real data, and $D(z)$ is the input data to the generator. By calculating the loss function, the generator parameters are fixed, and the optimal discriminator parameters are obtained, expressed in Eq. (3).

$$D* = \frac{p_{data}}{p_{data}+p_g} \tag{3}$$

In Eq. (3), p_g represents the generated samples. The expression for introducing the Jensen-Shannon divergence during optimization is shown in Eq. (4).

$$V(D_G^*, G) = 2JS(p_g \| p_{data}) - 2\,log\,2 \tag{4}$$

In Eq. (4), JS represents the Jensen-Shannon divergence. DCGAN updates alternately through fixed iterations to generate more realistic samples. However, traditional DCGAN still suffers from regional artifacts and instability (Rahman et al., 2025; Wei et al., 2023). To improve the algorithm, this study applies FPN to optimize the encoder-decoder structure in the generator, reducing information loss, and introduces sub-pixel convolution blocks and BN in the decoder to further improve the DCGAN structure and enhance recognition efficiency and accuracy. The optimized encoder-decoder structure is shown in Fig. 2.

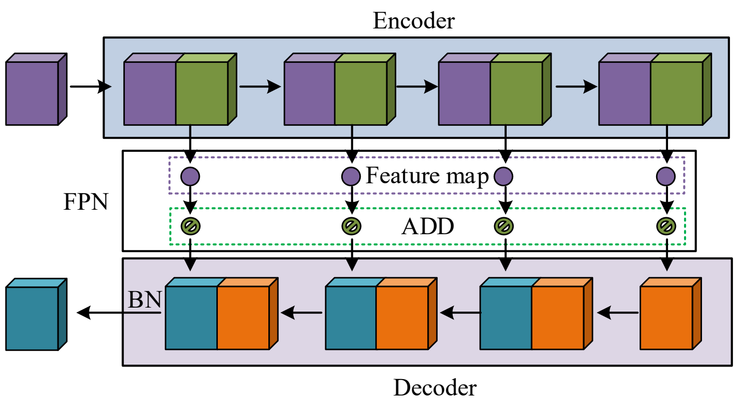


Fig. 2. Optimized encoder-decoder structure diagram

As shown in Fig. 2, the encoder extracts image content and style features via multiple convolutional layers, yielding shallow and deep features and details. FPN uses skip connections between the encoder and decoder to capture location information of blurred regions, reducing data loss and achieving comprehensive feature extraction. Sub-pixel convolution layers are added to the decoder to upsample feature maps and amplify details, improving output image quality. Finally, BN modules are added to the decoder to accelerate training. Among them, the upper Encoder module is a 4-layer convolutional structure, with a 4×4 convolution kernel and a stride of 2. The multi-scale style features are extracted layer by layer. Decoder module is a 4-layer transposed convolution structure with 4×4 convolution kernels and 2 steps. The computation of FPN is expressed in Eq. (5) (Tulasiram et al., 2025).

$$F_i \in \mathbb{R}^{C_i \times H_i \times W_i}, i = 1,2,3,4 \tag{5}$$

In Eq. (5), C_i, H_i, and W_i are the number of channels, feature map height, and width. Residual modules fuse features and reduce channel dimensions, expressed in Eq. (6).

$$M_4 = Conv2d_{1\times1}(F_4) \tag{6}$$

In Eq. (6), M_4 represents 4-layer convolutional fusion, $Conv2d_{1\times1}$ represents convolution operations, and F_4 represents the 4-layer feature maps. The remaining upsampled features are fused, expressed in Eq. (7).

$$M_i = Upsample(M_{i+1}) \oplus conv2d_{1\times1}(F_i) \tag{7}$$

In Eq. (7), $Upsample$ represents the upsampling operation. The improved generator structure is integrated into DCGAN to obtain the Improved DCGAN (IDCGAN), with its operation flow shown in Fig. 3.

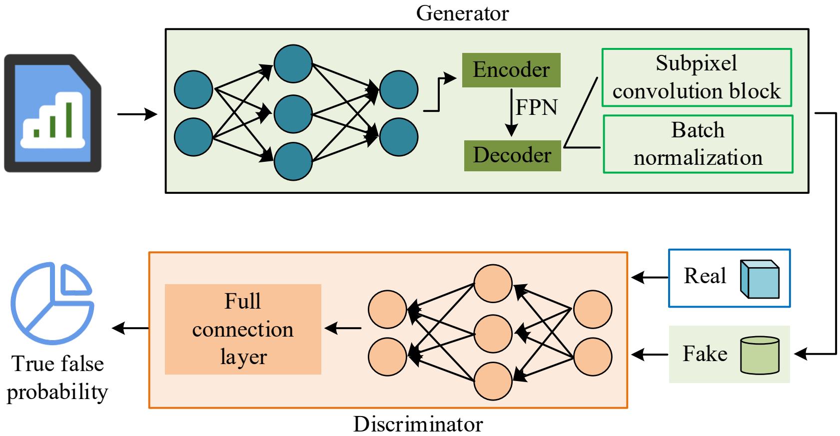


Fig. 3. Flowchart of the IDCGAN algorithm (Icon source: www.iconfont.cn)

As shown in Fig. 3, IDCGAN first inputs data into the generator. The encoder extracts features and details through convolution operations. Skip connections link the encoder and decoder to capture the location information of blurred regions. Sub-pixel convolution layers in the decoder amplify feature maps to enhance details. BN modules improve output speed. The generated image samples are fed to the discriminator along with preprocessed real samples. The discriminator distinguishes real samples from generated samples and feeds back to the generator. Through interactive feedback between the discriminator and the generator, the algorithm ultimately converges to optimal parameters.

3.2. Construction of Artistic Style Transfer Model in Film and Animation Production

The IDCGAN algorithm improves the saliency of content and style features in animated images, but it suffers from algorithmic noise, low robustness, and may reduce recognition accuracy. To address data fluctuations in IDCGAN, this study combines KCF and MSS algorithms to develop the Improved MSS (IMSS) algorithm and integrates it into IDCGAN. This approach resolves interference from abnormal data, ultimately constructing an artistic style transfer model based on IDCGAN and IMSS. Among them, the MSS algorithm can accurately distinguish the core features of animation art style from noise, abnormal textures, and other interference data by dynamically adjusting the size and weight distribution of multiple windows, thereby minimizing the impact of non-target interference information on the style migration path. The calculation of multi-window spectra is shown in Eq. (8).

$$S^{mt}(\omega) = \frac{1}{L}\sum_{k=0}^{L-1} S_k^{mt}(\omega) \tag{8}$$

In Eq. (8), S^{mt} represents the spectrum, L represents the number of data windows, and k represents the order of data windows. The windowed framed data undergo Fourier transform, and the spectral subtraction amplitude is calculated as shown in Eq. (9).

$$|X_i(k)| = \frac{1}{2M+1}\sum_{i=-M}^{M}\left|X_i+_j(k)\right| \tag{9}$$

In Eq. (9), M represents the number of frames, and i represents the frame center. The workflow of the MSS algorithm is shown in Fig. 4.

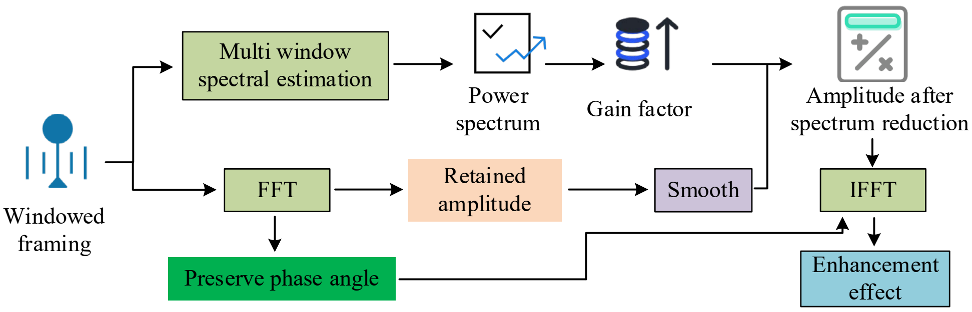


Fig. 4. Specific flow chart of MSS algorithm (Icon source: www.iconfont.cn)

As shown in Fig. 4, the MSS algorithm improves on the traditional periodogram method by performing multi-window framing. To match the animation frame's local style, the multi-window framing size is set to 32×32, and the window repetition rate is 50%. Pre-processing abnormal texture and noise data frame-by-frame can effectively avoid the influence of non-target interference on style feature extraction. Then, calculate the amplitude via the Fourier transform, accurately retain the phase-angle correlation features, such as key line trends and color transitions, and avoid the loss of style details. To balance the calculation accuracy and efficiency, set the number of Fourier transform points to 256. At the same time, the animation data are smoothed over three iterations to calculate the amplitude after spectral subtraction. The average power spectrum after smoothing and the average noise power spectrum are estimated using the multi-window spectrum estimation method, and the spectral subtraction amplitude and gain factor are determined from the power spectrum to efficiently separate noise and style features. Fourier transform operations enhance data quality, removing noise and improving feature and style quality. The power spectrum calculation is shown in Eq. (10).

$$P_y(k,i) = \frac{1}{2M+1}\sum_{i=-M}^{M} P(k,i+j) \tag{10}$$

In Eq. (10), j represents the center of the Fourier-transformed subtracted spectra, and P represents the smoothed power spectrum. Enhanced information is obtained through the Fourier transform, as expressed in Eq. (11).

$$\left|\hat{X}_i(k)\right| = g(k,i) \times |X_i(k)| \tag{11}$$

In Eq. (11), $g(k,i)$ represents the gain factor, and $X_i(k)$ represents the spectral subtraction amplitude. KCF is introduced into MSS to improve the matching mechanism, achieving more accurate matching in higher-dimensional spaces. The workflow of the IMSS integration is shown in Fig. 5.

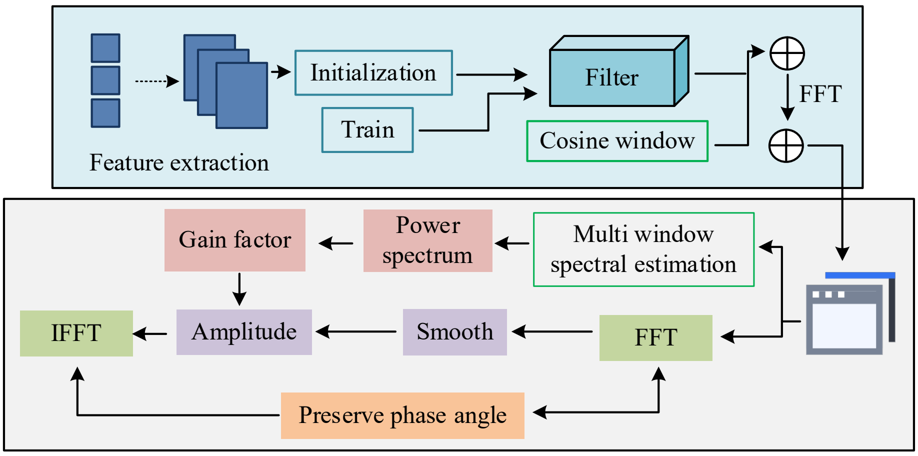


Fig. 5. Specific flow chart of IMSS fusion algorithm (Icon source: www.iconfont.cn)

As shown in Fig. 5, KCF first selects the initial frame and extracts region features to initialize the filter. For each current frame, it extracts features and applies a cosine window to the convolution operations. Fourier transform switches the computation mode. The filter denoises Fourier-transformed features, thereby highlighting target-feature recognition. Fourier transform converts the output to a response map, predicts positions, and updates filter parameters. MSS then computes the amplitude via the Fourier transform, applies smoothing, calculates the spectral-subtraction amplitude and gain factor via multi-window spectral estimation, and enhances signals using the Fourier transform. The Gaussian classifier parameters are calculated as shown in Eq. (12).

$$\omega = \sum_{i=1}^{m} \alpha_i\, \varphi(x_i) \tag{12}$$

In Eq. (12), ω represents filter coefficients, α represents column vectors, and $\varphi(x_i)$ represents input data. The Gaussian kernel function in KCF is expressed in Eq. (13).

$$k(x,x_1) = \langle \varphi(x), \varphi(x_1) \rangle \tag{13}$$

In Eq. (13), k represents the Gaussian kernel function. The classification function derived from the Gaussian kernel is expressed in Eq. (14).

$$f(c) = \omega^T c = \sum_{i=1}^{m} \alpha_i k(c, x_i) \tag{14}$$

In Eq. (14), $f(c)$ represents the classification function. The ridge regression vector is then Fourier-transformed, as shown in Eq. (15).

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx}+\lambda} \tag{15}$$

In Eq. (15), $\hat{\alpha}$ represents the Fourier-transformed ridge regression vector, $\hat{y}$ represents the regression target, $\hat{k}^{xx}$ represents the first row vector of the kernel matrix, and λ represents the regularization parameter. IMSS improves recognition accuracy in artistic style transfer, reduces computational load, optimizes overall structure, enhances saliency in feature extraction, and effectively reduces multi-level feature loss. Therefore, IDCGAN and IMSS are integrated to construct a film and animation production-style artistic-style transfer model (IDCGAN-IMSS), with the operational flow shown in Fig. 6.

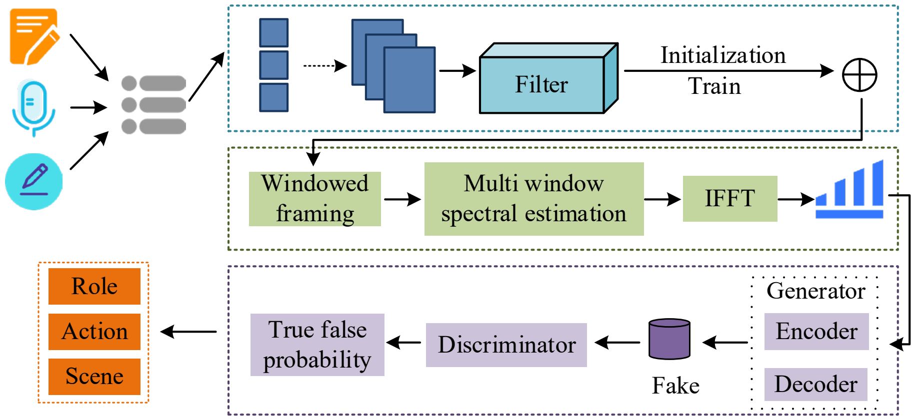


Fig. 6. Operation process of IDCGAN-IMSS model (Icon source: www.iconfont.cn)

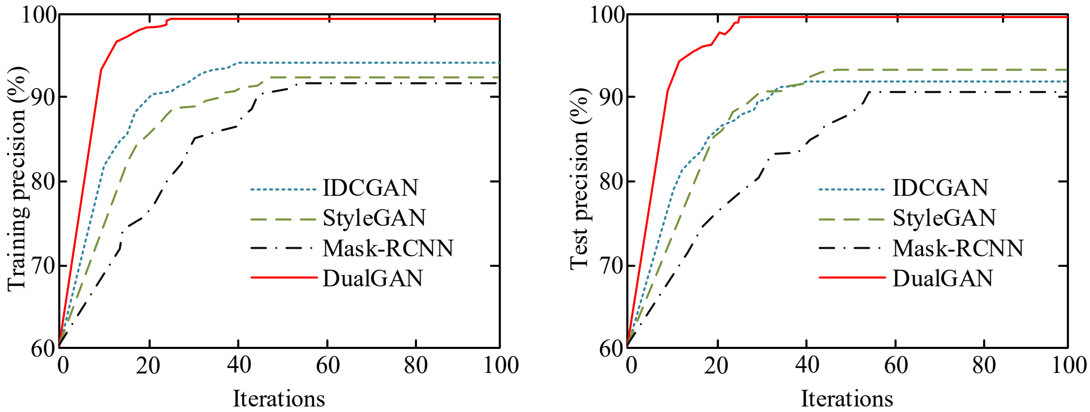
As shown in Fig. 6, the model takes image, audio, and text data as input for recognition and converts them into animation sequences. IMSS initializes the filter, applies a cosine window and Fourier transform for denoising, predicts positions, and updates filter parameters. Fourier transform and multi-window spectral estimation compute spectral subtraction amplitude and gain factor, and the Fourier transform further enhances the signal. The generator extracts feature information using skip connections, sub-pixel convolution, and BN modules to improve data quality and output speed. The discriminator distinguishes between samples and feeds back to the generator, jointly optimizing the parameters. The optimized model performs artistic style transfer, automatically generating animation characters, actions, and scenes, improving animation production speed and quality.

4. Performance of IDCGAN Film and Animation Production Artistic Style Transfer Model

4.1. Performance Validation of IDCGAN Algorithm

To verify the effectiveness of the IDCGAN artistic style transfer algorithm, the study compared it with the Style-Based Generative Adversarial Network (StyleGAN), the Mask-Region-based Convolutional Neural Network (Mask-RCNN), and the Dual Learning GAN (DualGAN) artistic style transfer algorithms. The experiments ran on Windows 11 and Ubuntu 18.04, using Intel(R) Xeon(R) Platinum processors and Adam as the optimizer. The learning rate was set to 0.0005, and the number of iterations was 100. The WikiArt dataset was used as the style image dataset, and the Microsoft Common Objects in Context (MS-COCO) dataset was used as the content image dataset. To analyze the performance of the IDCGAN algorithm, this study evaluated the training and testing accuracies of IDCGAN, StyleGAN, Mask R-CNN, and DualGAN for artistic style migration. The results are shown in Fig. 7.

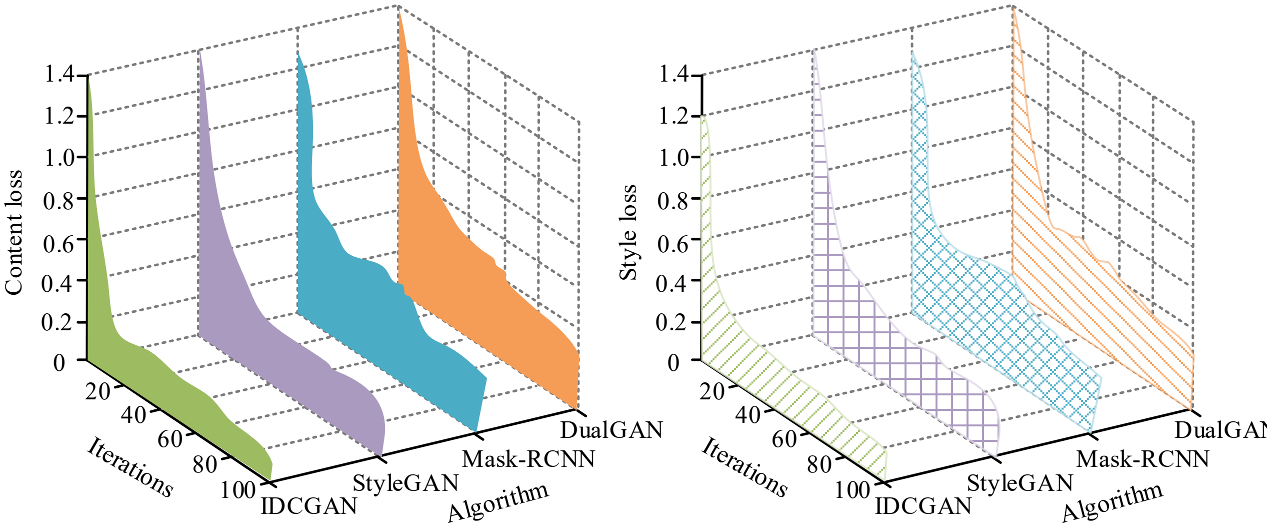
As shown in Fig. 7(a), the training accuracy of the IDCGAN algorithm increased rapidly with increasing number of iterations. After the iteration, the training accuracy rose to 98.93%, and the change trend and accuracy value are significantly higher than those of the comparison algorithm. As shown in Fig. 7(b), the test accuracy of the IDCGAN algorithm rose rapidly. When the number of iterations was 24, the test accuracy reached a high of 99.19%, outperforming the comparison algorithm. To sum up, the training and test accuracies of the IDCGAN algorithm increased simultaneously to high levels, indicating that the model exhibits good generalization performance. At the same time, from the perspective of convergence characteristics, the training accuracy and test accuracy of the IDCGAN algorithm show a rapid upward trend at the beginning of the iteration, indicating that the algorithm could achieve effective convergence within a limited number of iterations, and its network structure has an efficient feature learning ability, which could quickly capture the core features in the data. To further verify the convergence performance of the IDCGAN algorithm in artistic style transfer, the study compared the content and style losses for IDCGAN, StyleGAN, Mask-RCNN, and DualGAN. The test results are shown in Fig. 8.



(a) Comparison of training precision of each algorithm

(b) Comparison of test precision of each algorithm

Fig. 7. Results of training precision and test precision for each model

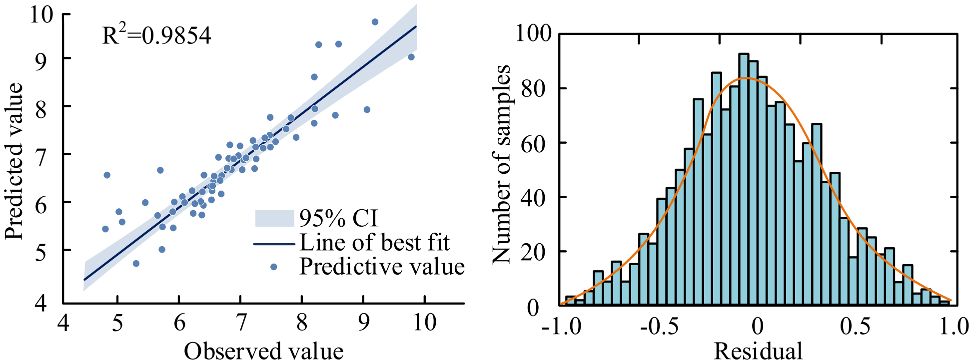


(a) Content loss value of each algorithm

(b) Style loss value of each algorithm

Fig. 8. Comparison of content loss and style loss results

In Fig. 8(a), the content loss of IDCGAN decreased rapidly to 0.18 within 0–20 iterations. From iteration 20 to 80, the rate of decrease slowed, and the loss reached 0.09 after 100 iterations. The trend and loss values of IDCGAN were superior to those of other comparison models. Fig. 8(b) showed that the style loss of IDCGAN decreased rapidly in the first 10 iterations. Between 20 and 80 iterations, the rate of decrease slowed, and the final loss was 0.12 after 100 iterations. The trend and values were better than those of other models. These results indicated that IDCGAN significantly reduced data loss and exhibited better convergence, as its upsampling and downsampling mechanisms made feature extraction and image generation more efficient and stable, minimizing information loss. To evaluate the fitting performance of IDCGAN for artistic style transfer, the study analyzed predicted correlation and residuals. The results are shown in Fig. 9.



(a) Correlation of IDCGAN algorithm

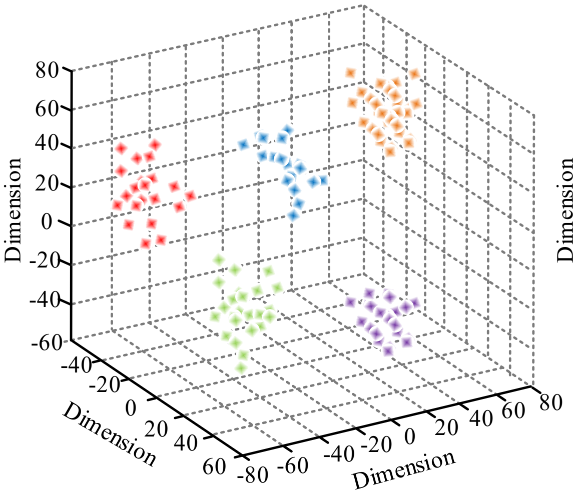
(b) Residual of IDCGAN algorithm

Fig. 9. Comparison of prediction correlation and residual results of IDCGAN algorithm

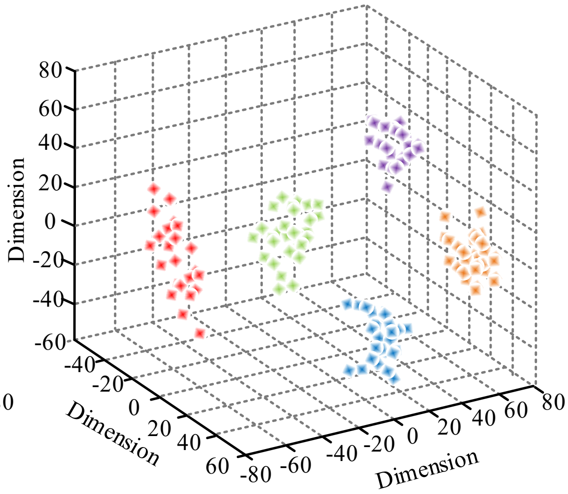
As shown in Fig. 9(a), the predicted values and real values exhibited a linear correlation. The Coefficient of Determination (R2) calculated from the best-fit line was 0.9854, approaching 1, indicating strong explanatory ability. The 95% confidence interval was narrow and close to the best-fit line, indicating high stability, as the algorithm prevents gradient vanishing or explosion during backpropagation, thereby enhancing reliability. Fig. 9(b) showed that the predicted results followed a significant normal distribution, indicating that convolutional layers reduced redundant parameter computation and lowered the risk of underfitting and overfitting. These results confirmed that IDCGAN had strong stability and reliability.

4.2. Application Effects of Artistic Style Transfer Model in Film and Animation Production

After verifying the IDCGAN's performance, the study further evaluated its practical application by comparing it with StyleGAN, Mask R-CNN, and DualGAN. The optimal running environment for the model is an Intel (R) Xeon (R) Platinum processor, an RTX 4090 graphics card, and 20 G of memory. When the hardware configuration falls short of this standard, the model's reasoning delay will increase sharply, and video memory overflow is more likely to occur when processing high-resolution animation frames. The set learning rate is 0.00001, the number of iterations is 100, the batch size is 64, and the weight attenuation coefficient is 0.005. The scene adaptation is suitable for 256×256 resolution-style conversion but not for 8K UHD frame migration, mobile terminal real-time interactive animation creation, or other scenes. The migration effect for solid-background animation frames without obvious style characteristics is poor. Study RGB video animation frame images with 256×256×3 inputs. WikiArt and MS-COCO datasets were used. To assess artistic style classification, scatter plots were used to visualize the image style classification results of IDCGAN-IMSS, as shown in Fig. 10.



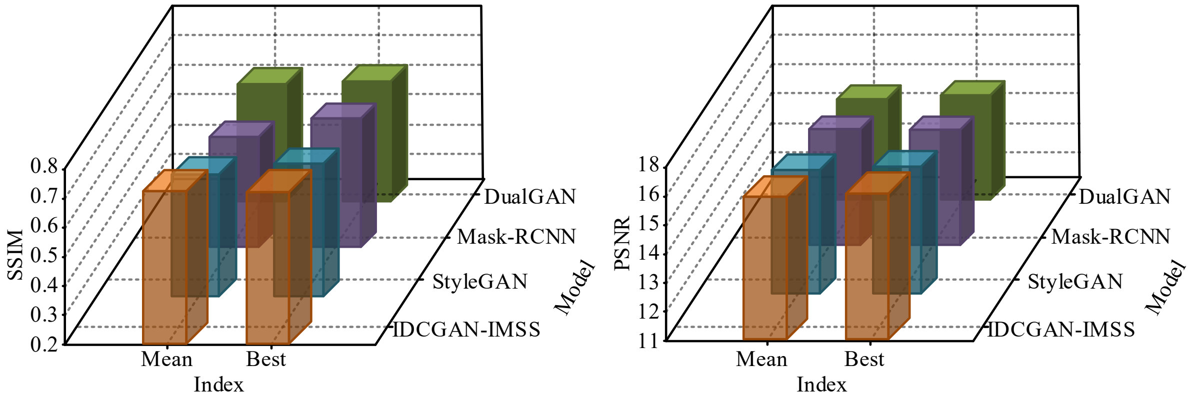
(a) Distribution of distribution points in WikiArt



(b) Distribution of distribution points in MS-COCO

Fig. 10. Visualization results of artistic style classification

The two-dimensional distribution in Fig. 10 is the visual result obtained by mapping the high-dimensional artistic style features produced by the model into a two-dimensional space using the t-SNE (t-distributed Stochastic Neighbor Embedding) dimensionality reduction algorithm. Fig. 10(a) showed that IDCGAN-IMSS classified images in the WikiArt dataset into five visual styles: realistic, cartoon, abstract, ink painting, and cyberpunk. The classification boundaries were clear, and the clustering was well-defined. Fig. 10(b) shows that images in MS-COCO were similarly divided into five clusters, with good intra-cluster aggregation and clear inter-cluster separation. These results indicated that IDCGAN-IMSS had excellent classification accuracy. The improvement was mainly due to IMSS reducing noise interference and enhancing image representation, allowing the generator to extract features accurately and improve classification accuracy. To further assess artistic style transfer performance, IDCGAN-IMSS was compared with StyleGAN, Mask-RCNN, and DualGAN using the Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR). The results are shown in Fig. 11.



(a) SSIM values of each model (b) PSNR values of each model

Fig. 11. Comparison of SSIM and PSNR test results

As shown in Fig. 11(a), the average SSIM of IDCGAN-IMSS was 0.75, with a maximum of 0.76. In comparison, the SSIM averages of StyleGAN, Mask-RCNN, and DualGAN were 0.52, 0.61, and 0.71, respectively. Fig. 11(b) shows that the average and maximum PSNR values for IDCGAN-IMSS were 15.81 and 16.79, respectively, outperforming the other three models. This was because IDCGAN-IMSS quickly captured shallow and deep features, accurately learned artistic styles, and reduced content loss, achieving superior animation quality. Overall, IDCGAN-IMSS exhibited excellent style transfer effects, high perceptual quality, and visual performance. To further analyze artistic style transfer, IDCGAN-IMSS results were compared with the transferred images of the three other models, shown in Fig. 12.

(a) Style (b) Content (c) IDCGAN-IMSS (d) StyleGAN (e) Mask-RCNN (f) DualGAN

Fig. 12. Comparison of style transfer effects

Using Figs. 12(a) and (b) as style and content templates, Fig. 12(c) showed that IDCGAN-IMSS produced clear images after style transfer, preserving style quality and content structure. Fig. 12(d) showed that StyleGAN partially completed the style transfer, but the content structure was blurry. Figs. 12(e) and (f) showed that Mask-RCNN had poor style transfer, with color loss and incomplete content, and DualGAN produced locally blurry images with artifacts. These results indicated that IDCGAN-IMSS had superior artistic style transfer performance. To further evaluate comprehensive performance in film and animation production, IDCGAN-IMSS was compared with StyleGAN, Mask-RCNN, and DualGAN in Fréchet Inception Distance (FID), ArtFID, Floating-point Operations Per Second (FLOPs), and style feature matching. The FLOPs calculation formula is as follows (16)

$$FLOPs = 2 \times C_{in} \times C_{out} \times K_h \times K_w \times H_{out} \times W_{out} \tag{16}$$

In formula (16), C_{in} represents the number of input channels, C_{out} represents the number of output channels, K_h and K_w represents the size of convolution kernel, H_{out} and W_{out} represents the size of output graph. The results are shown in Table 1.

Table 1. Comparison of multiple comprehensive indicator results

Dataset	Model	FID	ArtFID	FLOPs (G)	Matching degree (%)
WikiArt	StyleGAN	17.65	28.45	69.21	74.36
	Mask-RCNN	18.54	29.66	70.25	70.25
	DualGAN	16.62	30.21	69.01	79.62
	IDCGAN-IMSS	19.24	31.01	70.22	89.14
MS-COCO	StyleGAN	21.58	30.45	54.22	74.51
	Mask-RCNN	22.14	28.60	56.87	78.22
	DualGAN	26.94	30.19	55.88	76.38
	IDCGAN-IMSS	22.55	31.69	54.78	91.20

As shown in Table 1, through the WikiArt dataset test, the style feature matching degree is 89.14%, the highest among all models.The matching degree of style features is the core evaluation goal of this task, and the results above verify its advantages for artistic style expression and classification. While FID and ArtFID are higher than the comparison model, the gap is small. This is because the model's reasonable balance of the global pixel distribution does not affect its superiority in core tasks when learning focus style features. Testing with MS-COCO, IDCGAN-IMSS achieved FID of 22.55, ArtFID of 31.69, FLOPs of 54.78, and style feature matching of 91.20%. ArtFID and style feature matching were significantly higher than those of the other models because upsampling and convolution increased image resolution, and IMSS enhanced data representation in animation style transfer, thereby maximizing content and style preservation and improving overall visual performance. These results indicated that IDCGAN-IMSS improved artistic style transfer quality and enhanced animation production. In order to further explore the integration effectiveness of the IDCGAN-IMSS model, the study set DCGAN as model A, IDCGAN as model B, IDCGAN+MSS as model C, and the complete IDCGAN-IMSS model as model D. At the same time, based on the WikiArt and MS-COCO data sets, continued to mix Studio Ghibli to open the line draft material data, and under the interference of low-intensity Gaussian noise (standard deviation is 0.1) and high-intensity noise (standard deviation is 0.3), test the ArtFID and Matching degree of ablation models, and the results are shown in Table 2.

As shown in Table 2, as modules are gradually superimposed, the model's ArtFID continues to decrease, while the matching degree continues to increase. In the noiseless scene, the ArtFID of complete model D is 29.63, which is 29.3% lower than that of model A, and the matching degree of complete model D is 90.69%, which is higher than that of other models. Under the interference of low- and high-intensity noise, the ArtFID and the matching degree of the complete model D show no obvious change compared with the environment without noise. The ArtFID of model D is 32.01 and 33.51, respectively, and the matching degree is 89.50% and 88.32%, respectively, which indicates that the DCGAN structure optimization and the introduction of IMSS module effectively strengthen the model's resistance to noise interference, and can maintain stable style migration accuracy even under the input of degraded data. In addition, the anti-interference ability of model D is better than that of each ablation module under varying levels of noise, indicating that IDCGAN and IMSS exhibit a synergistic effect that enhances the model's robustness in complex data scenarios. In conclusion, the integrated design of IDCGAN-IMSS was effective and reliable.

Table 2. Results of ablation test

Interfere	Model	ArtFID	Matching degree (%)
No noise	A	41.88	69.29
	B	35.88	78.50
	C	32.91	85.39
	D	29.63	90.63
Low intensity noise	A	56.77	57.09
	B	49.13	69.74
	C	40.61	79.28
	D	32.01	89.50
High intensity noise	A	81.12	40.98
	B	63.21	55.51
	C	45.45	67.52
	D	33.51	88.32

5. Conclusion and Recommendations

To address the issues of slow generation speed, unnatural image effects, and high computational load in artistic style transfer for film and animation production, this study developed an artistic style transfer model that combines the IDCGAN and IMSS algorithms. The proposed model added deep convolutional neural networks to the traditional GANs algorithm, resulting in the optimized DCGAN. FPN, sub-pixel convolution blocks, and BN improve the DCGAN generator architecture. IMSS was then introduced to construct the artistic style transfer model, further enhancing feature representation and improving animation production quality. Experimental results demonstrated that the algorithm outperformed comparative algorithms in fitness curve, content loss, and style loss. The optimal IDCGAN value was 0.9854, indicating good operational stability. The final constructed artistic style transfer model achieved accurate style classification, with an average SSIM of 0.75 and an average PSNR of 15.81, both higher than those of the comparative models. In the WikiArt and MS-COCO datasets, the style feature matching degree is 89.14% and 91.20%, respectively, both higher than those of the StyleGAN, Mask RCNN, and DualGAN models. In addition, the proposed model can maintain stable style transfer accuracy under varying levels of noise interference and is significantly superior to each ablation model. Overall, the IDCGAN-IMSS model efficiently captured features, performed fine-grained artistic-style classification, and improved animation production quality. Although this study has achieved the improvement of the style migration effect of film and television animation, it has not clearly defined the model failure scenarios, lacks quantitative analysis of computing resource constraints, and lacks research on generalization capability under the unbalanced style distribution scenarios, so in the future, in-depth research can be conducted on computing load optimization and multi-dimensional limitation verification.

Funding

The research is supported by Interim research results of the general scientific research project of the Education Department of Jilin Province in 2024, titled "Research on the Exploration and Development Strategies of Jilin Folk Literature Resources in the Digital Media Era" (Project No. JJKH20241454SK); Scientific Research Program of Jilin Animation Institute, 2025, Interim Research Results of the General Project Research on the Interactive Mechanism of AI + Ideological and Political Education' in Film and Television Teaching under the OBE Concept, (Project No. KY25DY01).

Institutional Review Board Statement

Not applicable.

Declaration of Artificial Intelligence (AI) Tools

The author used AI tools solely for language editing and to improve readability. The author reviewed and verified all content and takes full responsibility for the accuracy and integrity of the manuscript.

References

Ahmad, Z., Jaffri, Z. A., Chen, M., and Bao, S. (2025). Understanding GANs: Fundamentals, variants, training challenges, applications, and open problems. *Multimedia Tools and Applications*, 84(12), 10347-10423. doi: 10.1007/s11042-024-19361-y.

Ali, M., Ali, M., Hussain, M., and Koundal, D. (2025). Generative adversarial networks (GANs) for medical image processing: recent advancements. *Archives of Computational Methods in Engineering*, 32(2), 1185-1198. doi: 10.1007/s11831-024-10174-8.

Chen, J., Li, S., Liu, D., and Lu, W. (2022). Indoor camera pose estimation via style‐transfer 3D models. *Computer‐Aided Civil and Infrastructure Engineering*, 37(3), 335-353. doi: 10.1111/mice.12714.

Chen, Y., Yuan, Q., Li, Z., Liu, Y., Wang, W., and Xie, C. (2024). Upst-nerf: Universal photorealistic style transfer of neural radiance fields for 3d scene. *IEEE Transactions on Visualization and Computer Graphics*, 31(4), 2045-2057. doi: 10.1109/TVCG.2024.3378692.

Christophe, S., Mermet, S., Laurent, M., and Touya, G. (2022). Neural map style transfer exploration with GANs. *International Journal of Cartography*, 8(1), 18-36. doi: 10.1080/23729333.2022.2031554.

Garg, M., Ubhi, J. S., and Aggarwal, A. K. (2023). Neural style transfer for image steganography and destylization with supervised image to image translation. *Multimedia Tools and Applications*, 82(4), 6271-6288. doi: 10.1007/s11042-022-13596-3.

Hu, Z., Lee, R. K. W., Aggarwal, C. C., and Zhang, A. (2022). Text style transfer: A review and experimental evaluation. *ACM SIGKDD Explorations Newsletter*, 24(1), 14-45. doi: 10.1145/3544903.354490.

Kang, M., Shin, J., and Park, J. (2023). StudioGAN: A taxonomy and benchmark of GANs for image synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12), 15725-15742. doi: 10.1109/TPAMI.2023.3306436.

Kang, Y., and Kim, J. (2024). Animation character generation and optimization algorithm based on computer aided design and virtual reality. *Computer-Aided Design and Applications*, 21(S14), 46-62. doi: 10.14733/cadaps.2024.S14.46-62

Kiranyaz, S., Devecioglu, O. C., Ince, T., Malik, J., Chowdhury, M., and Hamid, T. (2022). Blind ECG restoration by operational cycle-GANs. *IEEE Transactions on Biomedical Engineering*, 69(12), 3572-3581. doi: 10.1109/TBME.2022.3172125.

Krishnan, P., Kovvuri, R., Pang, G., Vassilev, B., and Hassner, T. (2023). Textstylebrush: Transfer of text aesthetics from a single example. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7), 9122-9134. doi: 10.1109/TPAMI.2023.3239736.

Mokayed, H., Quan, T. Z., Alkhaled, L., and Sivakumar, V. (2023). Real-time human detection and counting system using deep learning computer vision techniques. *Artificial Intelligence and Applications*, 1(4), 221-229. doi: 10.47852/bonviewAIA2202391.

Njima, W., Bazzi, A., and Chafii, M. (2022). DNN-based indoor localization under limited dataset using GANs and semi-supervised learning. *IEEE Access*, 10, 69896-69909. doi: 10.1109/ACCESS.2022.3187837.

Rahman, M. A., Francia, G. A., and Shahriar, H. (2025). Leveraging GANs for synthetic data generation to improve intrusion detection systems. *Journal of Future Artificial Intelligence and Technologies*, 1(4), 429-439. doi: 10.62411/faith.3048-3719-52.

Soleymanzadeh, R. and Kashef, R. (2023). Efficient intrusion detection using multi-player generative adversarial networks (GANs): an ensemble-based deep learning architecture. *Neural Computing and Applications*, 35(17), 12545-12563. doi: 10.1007/s00521-023-08398-z.

Sun, Y., Liu, J., and Zhang, R. (2023). Large capacity generative image steganography via image style transfer and feature-wise deep fusion. *Applied Intelligence*, 53(23), 28675-28693. doi: 10.1007/s10489-023-04993-8.

Tulasiram, J., Banothu, B., and Nickolas, S. (2025). Monkeypox data enhancement and diagnosis using improved DCGAN. *Soft Computing*, 29(5), 2497-2513. doi: 10.1007/s00500-025-10509-y.

Wei, C., Liang, J., Liu, H., Hou, Z., and Huan, Z. (2023). Multi-stage unsupervised fabric defect detection based on DCGAN. *The Visual Computer*, 39(12), 6655-6671. doi: 10.1007/s00371-022-02754-1.

Zhang, N. and Pu, B. (2024). Film and television animation production technology based on expression transfer and virtual digital human. *Scalable Computing: Practice and Experience*, 25(6), 5560–5567. doi: 10.12694/scpe.v25i6.3351.

Zhang, Y., Xie, H., Zhuang, S., and Zhan, X. (2024). Image Processing and Optimization Using Deep Learning-Based Generative Adversarial Networks (GANs). *Journal of Artificial Intelligence General science*, 5(1), 50-62. doi: 10.60087/jaigs.v5i1.163.

Shifeng Wang earned his Bachelor of Arts in Radio and Television Directing from Changchun Normal University in 2013 and his Master of Arts in Radio and Television from the same university in 2021. His research interests focus on narratology and film and television technology. He served as a news reporter at Jilin TV Station (2013-2015). Since 2015, he has served as a lecturer at the School of Film, Jilin Animation Institute. He has published one academic paper, one academic monograph and textbook, participated in five scientific research projects, and won one academic award.