International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-15

https://doi.org/10.70917/ijcisim-2026-0021
Article

Computer Vision in the Media Field and Its Contribution to Content Creation

Xiaoguang Fan [1,*], Renkexin Li [1] and Yanan Wu [2]

[1] Communication University of Zhejiang, Hangzhou 310000, Zhejiang, China
[2] Zhejiang Shuren University, Hangzhou 310015, Zhejiang, China
* Correspondence author: fff9900123@163.com

Abstract: The integration and development of computer vision and art design in the media field is an important area of ongoing research and exploration in the humanities and society. This paper takes Generative Adversarial Network (GAN) as the research entry point, proposes its specific application in AIGC, and constructs an image art stylization model based on GAN to realize the migration of image art style for content creation in the media field. On this basis, corresponding to the needs of content generation and creation in the media field, the digital media content generation and creation technology based on computer vision technology is proposed to ensure the quality of the generated content. The content generation and creation technology in the media field proposed in this paper is applied to design public art in the public resting area of Guanfeng Village in Ganbei, Jiangxi Province, China, designing a public art gallery and evaluating it in comparison with neighboring villages such as Wangkou Village, Xiaoqi Village, Likeng Village and Huangling Village. Among the 25 adjective pairs in the public semantic evaluation, except for the adjective pair "abstract - figurative", the value of positive adjective pairs in the remaining adjective pairs of Guanfeng Village is closest to 1, which brings the public a good artistic emotional experience. The value of all the remaining adjective pairs in favor of positive is close to 1.

Keywords: Generative Adversarial Networks; AIGC; Style Migration; Media Content Design

1 Introduction

As a rapidly growing cultural industry, the media industry has maintained sustained growth at a rate higher than GDP growth in recent years [1]. In 2023, the total output value of China's media industry reached 3151.823 billion yuan, representing an 8.38% year-on-year increase. However, similar to the objective reality of widening disparities in China's socio-economic development, the rapid growth of China's media industry has also exhibited uneven development. Not only do different regions exhibit varying levels of media industry development, but there are also differences in the development of various forms of media [2-3]. The causes of these development disparities include differences in economic development levels across regions and industries, as well as factors related to population, culture, institutions, and technology. With the advent of the new media era, the media has transitioned from a traditional one-to-many information dissemination model to a multi-to-many interactive communication model [4].

The development of computer technology has provided the media with more communication channels and tools, making information transmission more efficient, convenient, and precise. Through computer technology, the media can better understand audience needs and enhance the effectiveness of communication, interaction, and user experience [5-6]. Additionally, computer technology has introduced new technical methods and applications to the media industry, such as big data analysis, artificial intelligence, and virtual reality. With the rapid development of artificial intelligence technology, the media industry is undergoing a historic transformation. This transformation not only alters the operational methods and business models of the media industry but also redefines the ways information is disseminated and content is created [7-9]. In today's media landscape, the rise of short video formats has led to users uploading content on major short video platforms, with the highest daily upload volume on the Douyin platform reaching 8 billion videos. However, this has also brought about challenges in content review efficiency [10-11]. Meanwhile, the media industry is increasingly pursuing high-quality content presentation, driving content creation toward diversification, intelligence, and immersion [12-13].

Computer vision is an important branch of artificial intelligence, focusing on enabling computers to understand and interpret the content of images and videos [14]. In the past, the application of computer vision was relatively limited. However, with technological advancements, it has achieved significant breakthroughs. By leveraging image processing, pattern recognition, and machine learning technologies, it is reshaping the media industry [15]. Literature [16] summarizes the applications of computer vision technology in the media industry, including content generation, content review, visual recommendations, enhancing user experience, content retrieval, and advertising. Literature [17] describes how, in the field of news dissemination, computer vision leverages artificial intelligence to generate videos, images, and other content for targeted dissemination in response to violent ideologies. However, computer vision also combats such malicious dissemination through the detection of fake news. Literature [18] uses computer vision models to analyze user-generated content related to a specific brand on social media, thereby monitoring user attitudes toward the brand and better maintaining its image. Similarly, literature [19] employs two computer vision models to classify and aesthetically evaluate image advertisements, and introduces different computer vision models and their applications in the advertising field. Literature [20] analyzed the visual characteristics of different faces of presidential candidates using computer vision technology to explore media bias in visual depictions of candidates. Literature [21] integrates artificial intelligence and computer vision into augmented reality systems, enhancing environmental perception, narrative coherence, and character development to achieve more immersive interaction modes and promote personalized experiences for new media users. Literature [22] explores how computer vision and other analytical technologies can effectively and quickly identify and remove inappropriate content in platform user-generated content reviews, maintaining platform security and a healthy creative environment. The media industry encompasses various types of media, including broadcasting, publishing, film, digital media, social media, and mobile media, spanning fields such as news, advertising, film, and entertainment. Based on this, this paper delves into the application of computer vision in the media field and its value in content creation, aiming to provide support for the development of the media industry.

In this paper, the principle of Generative Adversarial Network (GAN) is introduced, and its specific applications in AIGC are proposed, including image generation, text generation, audio generation, video generation and so on. Taking images as the main medium for content creation in the media field, the GAN-based image art stylization model is built by combining the structural similarity index (SSIM) and the least squares generative adversarial network (LSGAN), which maintains a balance between the image content features and stylistic features required for simultaneous creation in the media field. On the basis of realizing image art stylization processing, combining deep convolutional network and recurrent neural network algorithms, digital media content generation and creation technology based on computer vision technology is proposed. In the digital media content generation technology, the generator is defined to be responsible for generating images, and the discriminator is determined to distinguish the generated data from the real data to generate image data with higher fidelity. In the digital media content creation technology, the data is cleaned by removing noise, repairing missing values, and eliminating outliers, updating the optimizer model parameters through the loss function, generating images based on the GAN algorithm and optimizing the generated images. Simulation experiments are set up to test the performance of the media content generation and creation technology proposed in this paper, and it is applied to the design of public art creation in the public resting area of Guanfeng Village in Ganbei, Jiangxi Province, China, to explore the specific application performance of the content generation and creation technology in the media field in this paper.

2. Principles of Generative Adversarial Networks and Applications in AIGC

With the rapid development of Generative Adversarial Networks (GAN) in the field of deep learning, its application in Artificial Intelligence Generated Content (AIGC) is becoming more and more widespread [23]. In this chapter, we will introduce the GAN principle and propose its specific application in AIGC.

2.1. GAN Overview

GAN is a deep learning model that is widely used to generate synthetic data that approximates the distribution of real data.GAN usually consists of two neural network models, the generator and the discriminator, where the generator learns the latent distribution of the real data from noisy data and generates realistic samples; and the discriminator discriminates whether the input samples are real data or generated data. Through adversarial training, the generator and discriminator can be continuously optimized. The generator's goal is to generate realistic data to deceive the discriminator, while the discriminator's goal is to accurately distinguish real data from generated data. Through this competition, the GAN model can be continuously enhanced in its ability to generate realistic data. The training of the generator and the discriminator can be expressed by the objective function as shown in (1):

$$\begin{aligned}\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)] \\ +E_{z\sim p_z(z)}[\log(1-D(G(z)))]\end{aligned} \tag{1}$$

In Eq. (1), x represents the target data that the system expects to simulate, $G(z)$ denotes the synthetic data generated by the generator, z is a random vector of samples from the noise distribution $p_z(z)$, $D(x)$ is the probability that the output of real samples from the discriminator is real, and $D(G(z))$ denotes the probability that the generator considers the generator data as real.

By optimizing the above functions, the discriminator can attempt to maximize the probability of discriminating the true samples and minimize the probability of discriminating the generated data as true, while the generator can minimize the discriminator's discrimination of the generated data in the optimization process.

2.2. Application of GAN in AIGC

In the process of continuously improving the generative capability of GAN model, its application has been gradually extended to many fields of AIGC. As the core technology for generating high-quality and diversified content, GAN has been widely used in AIGC for image, text, audio and video generation.GAN can realize highly realistic and diversified content by utilizing the adversarial training mechanism of generator and discriminator, so that AIGC can meet a wider range of content requirements. The following are several typical applications of GAN in AIGC.

1) Image generation: GAN generates specific style or high-resolution images by learning features from a large amount of image data.

2) Text generation: GAN is combined with natural language processing models to generate coherent dialog, articles and other text content.

3) Audio generation: through the learning of music samples or speech data, GAN can generate different styles of music clips or realistic speech.

4) Video generation: GAN generates dynamic video frame sequences that have shown advantages in virtual reality and movie production. By learning the temporal characteristics of video data, the content generated by GAN can maintain a coherent visual effect, thus meeting the needs of dynamic content creation and virtual scene generation.

In the AIGC system, GAN can not only generate images and videos that meet specific themes, but also continuously optimize the generation effect through the feedback of the discriminator, and finally generate intelligent content that meets the user's needs.The various applications of GAN in the AIGC field show its strong potential in the areas of content creation, multimedia production and virtual reality.

3. GAN-Based Stylized Model for Image art

Image art stylization is an important module of content creation in media field. Under the massive image data, the use of Generative Adversarial Networks (GAN) to realize the image art style migration is of great significance for the content creation in the media field. In this chapter, we will build an image art stylization model based on GAN, so that the image detail generation in media content creation can have better performance.

3.1. Network Setup

A residual network based on SSIM and LSGAN is constructed for image art stylization learning and image style migration [24]. The construction of the network draws on the deep residual network (ResNet) and LSGAN, in which the generator network structure consists of three parts: convolutional layer, residual block, and transposed convolutional layer with different convolutional steps and template numbers, respectively, and completes the extraction of the input image content and the target image stylistic features as well as the fusion process between the two during the adversarial training process. The discriminator network structure consists of convolutional layers, which accomplishes the distinction between the generated samples and the real samples in the process of adversarial training.

The image art stylization model in this study consists of 2 sets of generators and discriminators. During the model training and learning process, the image is generated by the generator G_{X2Y} to produce an image with sketch drawing style, and the discriminator D_{Y} learns to correctly distinguish whether the current artwork is generated by the generator or not. At the same time, the sketched image is converted into the image corresponding to the reality state by the generator G_{Y2X}, and the discriminator D_{X} learns to correctly distinguish whether the current image is converted by the generator G_{Y2X} or not. The network structure of this model generator builds a deep residual network, including three parts: the front part consists of three convolutional layers; the middle part consists of nine residual blocks; and the tail part consists of two transposed convolutional layers and one convolutional layer. Except for the first and the last convolutional layers which use 7×7 convolutional kernel and the sliding step is taken as 1, the rest of the convolutional layers use 3×3 convolutional kernel and the sliding step is taken as 2. The generator G_{Y2X} and the generator G_{X2Y} both use the same network structure.

The discriminator network structure for the image art stylization model is a deep convolutional neural network consisting of five convolutional operations. The inputs are generated samples and real samples produced by the generator, and the outputs are judgment results. Both the discriminator D_{X} and the discriminator D_{Y} use the same network structure. In the network structure of generator and discriminator, layers 1~25 are generators and layers 26~30 are discriminators. The generator uses edge-mirror padding to make up the zeros before all convolution operations except for the transpose convolution operation. An instance normalization layer and a ReLU activation function layer are added after all the convolution operations except the generator output convolution layer, which uses Tanh as the activation function. The discriminator adds an instance normalization layer and Leaky ReLU layer after all convolution operations except the output convolution layer.

Leaky ReLU is used as the activation function in the discriminator, Leaky ReLU is an improved version of the ReLU activation function:

$$f(x) = \begin{cases} 0.02x, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{2}$$

3.2. Loss Function

When determining the objective function of the model, the generator G_{X2Y} and the discriminator D_Y during training the generator G_{X2Y} try to "trick" the discriminator D_Y as much as possible to make the discriminator D_Y Misidentifying the generated sample as a real sample, i.e., learning to minimize $\left(D_Y\left(G_{X2Y}(x)\right)-1\right)^2$. At the same time, the discriminator D_Y improves the ability to distinguish between generated samples and real samples through continuous learning, so that it scores lower for generated samples and higher for real graphs, that is, learning to minimize $\left(D_Y(y)-1\right)^2$ and $\left(D_Y\left(G_{X2Y}(x)\right)\right)^2$.

The loss function of the generator G_{X2Y} is:

$$\min_{G_{X2Y}} V_{GAN}(G_{X2Y}) = \frac{1}{2} E_{x \sim P_{data}(x)} \left[(D_Y(G_{X2Y}(x)) - 1)^2 \right] \tag{3}$$

The loss function of the discriminator D_Y is:

$$\min_{D_Y} V_{GAN}(D_Y) = \frac{1}{2} E_{y\sim P_{data}(y)}\left[(D_Y(y)-1)^2 \right] + \frac{1}{2} E_{y\sim P_{data}(y)}\left[(D_Y(G_{X2Y}(x)))^2 \right] \tag{4}$$

Similarly, the loss function of the generator G_{Y2X} is:

$$\min_{G_{Y2X}} V_{GAN}(G_{Y2X}) = \frac{1}{2} E_{y\sim P_{data}(y)}\left[(D_X(G_{Y2X}(y)) - 1)^2 \right] \tag{5}$$

Similarly, the loss function of the discriminator D_{X} is:

$$\min_{D_X} V_{GAN}\left(D_X\right) = \frac{1}{2} E_{xp_{data}(x)}\left[\left(D_Y(x) - 1\right)^2 \right] + \frac{1}{2} E_{yp_{data}(y)}\left[\left(D_X\left(G_{Y2X}(y)\right)\right)^2 \right] \tag{6}$$

In order to further optimize the accuracy of the GAN, this paper adds a reconstruction loss function as part of the loss function of the generator on the basis of the loss function of the LSGAN. To measure the degree of difference between the generator input image and its reconstructed image, SSIM is introduced to quantify the similarity between the two. The structural similarity between the input image x and its reconstructed image $G_{Y2X}\left(G_{X2Y}(x)\right)$ is defined as:

$$SSIM(x,\hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_x\sigma_{\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)} \tag{7}$$

$$\hat{x} = G_{Y2X}(G_{X2Y}(x)) \tag{8}$$

Then the generator G_{X2Y} reconstructs the loss function term as:

$$L_{rec} = E_{x\sim P_{data}(x)}\left[1 - SSIM(x,\hat{x})\right] + E_{y\sim P_{data}(y)}\left[1 - SSIM(\hat{y},y)\right] \tag{9}$$

Ultimately, the loss function of the generator is:

$$\min_{G_{X2Y}} V_{GAN}(G_{X2Y}) = \lambda L_{rec} + \frac{1}{2} E_{x\sim P_{data}(x)}\left[(D_X(G_{Y2X}(y)) - 1)^2 \right] \tag{10}$$

where λ is used to control the relative importance between the confrontation loss term and the reconstruction loss term.

3.3. Algorithm Flow

Step1: Input image set X, image set Y, generator G_{X2Y} and its corresponding parameters θ_{X2Y}, discriminator D_Y and its corresponding parameters ω_Y, generator G_{YZX} and its corresponding parameter θ_{Y2X}, discriminator D_X and its corresponding parameter ω_x, weight coefficients λ, maximal cycle number N_{epoch}, the current number of iterations t, the number of rounds to start decreasing the learning rate N_{offset}, the generator initial learning rate η^g and the discriminator initial learning rate η^d.

Step2: Randomly initialize θ_{X2Y}, α_Y, θ_{Y2X} and α_X.

Step3: Update the number of iterations $t = t + 1$.

Step4: Sample an image from the sets X and Y, input it to the corresponding generator to get the corresponding generated sample, and label the real sample as well as the generated sample with 1 and 0 respectively.

Step5: Input the generation samples to the corresponding generator to get the reconstructed image of the image, and calculate the reconstruction loss for each generator.

Step6: The generated samples and the real samples are input into the discriminator respectively, calculate the discriminator objective function, minimize this objective function, and at the same time, according to the error, update the weights of the discriminator by using the error back-propagation and the improved Adam optimization; similarly, the weights of the discriminator are updated.

Step7: Optimize the generator, and at the same time minimize the generator objective function, and also take the error back-propagation and the improved Adam optimization to update the weights of the generator; similarly, optimize the generator and update the weights.

Step8: Judge the size of the current iteration number t in relation to N_{offset}, and then update the value of the learning rate.

Step9: Cycle through Step3 to Step9 until the maximum number of cycles is reached.

4. Technologies for Content Generation and Creation in the Media Field

In the above paper, this paper uses Generative Adversarial Network (GAN) to realize image artistic style migration and constructs an image artistic stylization model. On this basis, this chapter will propose the content generation and creation technology in the media field based on the generative adversarial network, combined with the images processed by artistic style migration, to realize the generation and design of images required for content creation in the media field.

4.1. GAN-Based Digital Media Content Generation Technology

Digital media content generation algorithms involve image acquisition, processing and synthesis, and their technical basis is image understanding algorithms, through which the algorithm recognizes the objects in the image, and understands the object context relations and interactions.GAN algorithm is the main algorithm for high-quality image synthesis at present, which has a better performance of image generation, and it is one of the key technologies for digital media content generation at present. The algorithm is characterized by its adversarial structure, which consists of two parts: the generator and the discriminator, the former is responsible for generating images, and the latter is responsible for distinguishing between generated data and real data. Specifically, the GAN algorithm data generation process can be decomposed into the following steps.

First, define the generator G : learn the data distribution law. Set a noise variable z according to a predefined noise distribution, and use the generator G to find the mapping $\dot{G}\left(z;\theta_g\right)$ of the noise variable z in the data space. The θ_g in $\dot{G}\left(z;\theta_g\right)$ denotes the parameters of the generator network.

Second, define the discriminator D : distinguish between real data samples and generator-generated data samples in the input data. Assuming that the data sample is x, the discriminator D will output a scalar for x indicating the probability that the sample x is real data, i.e., $D\left(x;\theta_d\right)$. where θ_d denotes the discriminator network parameters.

Third, algorithm training: regarded as a minimal maximal problem. Define the loss function $L(G,D)$ of the generator and the discriminator as:

$$x_D L(G,D) = E_{x\sim P_{data}(x)}[\log D(x)] + E_{z\sim P_z(z)}\left[\log(1 - D\{[G(z)]\}\right] \tag{11}$$

Fourth, the optimal solution D^* of the discriminator D is solved: the maximum value of $L(G,D)$ when solving G is constant. The optimal solution D^* of the discriminator D is computed with the formula expression:

$$D^*(x) = \frac{P_{data}\ (x)}{P_{data}\ (x) + p_g(x)} \tag{12}$$

Fifth, algorithm optimization: $p_g(x)$ is the data distribution generated by the generator G. The optimization objective $C(G)$ of the generator G can be obtained by substituting D^* into $\mathrm{L(G,D)}$, and the cost function of the generator $C(G)$ can be obtained by substituting the definition of D^* into the D^* expression in the above equation as:

$$C(G) = \min G - \log(4) + 2 \times JSD\left(P_{data} \middle\| P_g\right) \tag{13}$$

where: JSD is the Jensen-Shannon scatter, a measure of the similarity of probability distributions.

To optimize the generator, θ_g is updated with the gradient descent method, and the gradient of $C(G)$ with respect to θ_g is computed, and the formula is expressed as:

$$\nabla_{\theta_g} C(G) = E_{z \sim p_z(z)}[\nabla_{\theta_g} \log(1 - D\{[G(z;\theta_g)]\}] \tag{14}$$

The gradient can be continuously iterated and updated in the neural network, allowing the generator G to produce image data with a higher degree of fidelity.

4.2. GAN-Based Digital Media Content Creation Techniques

This study takes GAN algorithm as an example to construct an image generation model. According to the flow of generation technology deduced above, the creation method of digital media content mainly includes four main steps: data preprocessing, model training, image generation and post-processing.

First, data preprocessing, i.e., data cleaning, enhancement, labeling and normalization. In this study, we first cleaned the data that may interfere with the test by removing noise, repairing missing values, and eliminating outliers, enhanced the dataset by using image rotation, scaling, cropping, and color transformations, did a good job of image object recognition and classification labeling, and finally used normalization to ensure that the input data are at the same scale, so as to ensure the stability of the model training and convergence speed. Specifically, the image data is smoothed using Gaussian filter, let the original image be $I(x,y)$, Gaussian filter is $G(x,y,\sigma)$, σ is the standard deviation, the filtered image $I'(x,y)$ is denoted as [25]:

$$I'(x,y) = I(x,y) * G(x,y,\sigma) \tag{15}$$

The second is model training, i.e., initializing the network parameters, defining the loss function and the optimizer, using the loss function to measure the error value between the predicted value and the true value of the model, and updating the optimizer model parameters. Based on the algorithm iteration function to do multiple iterations of training, continuously input training data, perform forward propagation, calculate the loss, and perform backpropagation, which is used to update the model parameters. First, assuming a convolutional neural network of L layers, the output $H^{(l)}$ of each layer l can be computed from the output $H^{(l-1)}$ of the previous layer, as well as the weights $W^{(l)}$ and bias $b^{(l)}$ of that layer, which is expressed by the formula:

$$H^{(l)} = f\left(W^{(l)} * H^{(l-1)} + b^{(l)}\right) \tag{16}$$

where: f is the ReLU activation function. Let the predicted output of the model be $\hat{Y}$, the true label be Y, the number of samples be n, and the Mean Square Error (MSE) loss function L be defined as:

$$L = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \tag{17}$$

Each iteration to compute the loss function weight gradient $\nabla_W L$ updates the weights W, with η representing the learning rate, and the formula is expressed as:

$$W_{new} = W_{old} - \eta \cdot \nabla_W L \tag{18}$$

The third is image generation, which is based on the GAN algorithm to generate images. The generator network will output the input data as an image in the data space after receiving it and accurately evaluates the visual quality and similarity to the training set.

Fourthly, post-processing, i.e., optimizing the generated images so that the image quality meets the user requirements. The main functions of this step are image beautification, style conversion, resolution setting, effect addition, auto cropping, alignment and combination to make the output image more user friendly.

5. Simulation Experiments on Content Generation and Creation in the Media Field

In order to test the performance of the GAN-based digital media content generation and creation technology proposed in this paper, this chapter takes images as the main medium, and chooses face images, a common type of images in the media field, as the research sample to carry out simulation experiments of content generation and creation in the media field.

5.1. Experimental Setup

5.1.1. Data Sets

In the training phase, this chapter uses FS2K face sketch dataset and APDrawing face pen drawing dataset and performs face alignment and data enhancement in a preprocessing manner.

1) FS2K face sketching dataset

The FS2K dataset is a publicly released face sketching dataset by ETH Zurich, which includes 2250 photo-sketch image pairs from a wide range of image backgrounds, skin colors, and lighting conditions.

2) APDrawing face pen drawing dataset

The dataset used in the face pen drawing generation experiments is from Tsinghua University's publicly available dataset APDrawing, which is the most widely used face pen drawing dataset nowadays. The dataset contains 490 pairs of face photos and pen drawings with different skin colors, lighting, gestures, etc. The resolution of the original photos and pen drawings is 512 × 512.

5.1.2. Experimental Evaluation Indicators

In this experiment, FID, LPIPS, SCOOT, and FSIM are selected as the performance metrics used in the face sketch image generation experiment, and FID, SIFID, LPIPS, and CLEAN-FID are adopted as the performance metrics used in the face pen drawing image generation experiment.

1) FID: FID is a method of evaluating the generation quality by calculating the difference in the distribution of high-dimensional semantic features between the generated image and the real image. The smaller the value of FID, the higher the realism of the generated image texture.

2) LPIPS: LPIPS is an image quality assessment metric based on deep learning, which evaluates the perceptual difference between two image blocks by extracting image features and calculating the perceptual similarity. The smaller the LPIPS value, the higher the similarity between the generated image and the real image.

3) SCOOT: the structural texture perceptual metric, SCOOT describes the texture characteristics of an image by calculating the spatial relationship between pixel pairs in the image and the statistical information of the gray level. The larger the value of SCOOT, the better the sensory effect of the image.

4) FSIM: feature similarity, FSIM evaluates the quality of image generation by calculating the feature similarity between images, the larger the value of FSIM, the better the generated image features are maintained.

5) SIFID: SIFID is an index used to measure the quality of a single generated image, which uses the Inception network to extract the high-dimensional features of the image to calculate the difference between the generated image and the real image. The smaller the value of SIFID, the higher the realism of the generated image.

6) CLEAN-FID: The original image data is embedded into the model to calculate the "clean" features. The default embedding model is InceptionV3. The smaller the CLEAN-FID value, the higher the feature similarity between the images.

5.1.3. Image Preprocessing

Using the GAN-based image artistic stylization model proposed in this paper, the images in the experimental dataset are subjected to artistic style migration processing. A face pen drawing is randomly selected in the APDrawing face pen drawing dataset, and the face pen drawing is taken as an example for the artistic stylization process, and the local 2D spectrogram in the x-direction obtained through the transformation is specifically shown in Fig. 1. The original image content is more complex, color and texture, after processing its 2D image spectrum is similar to the general signal spectrum, which is convenient for the following image generation.

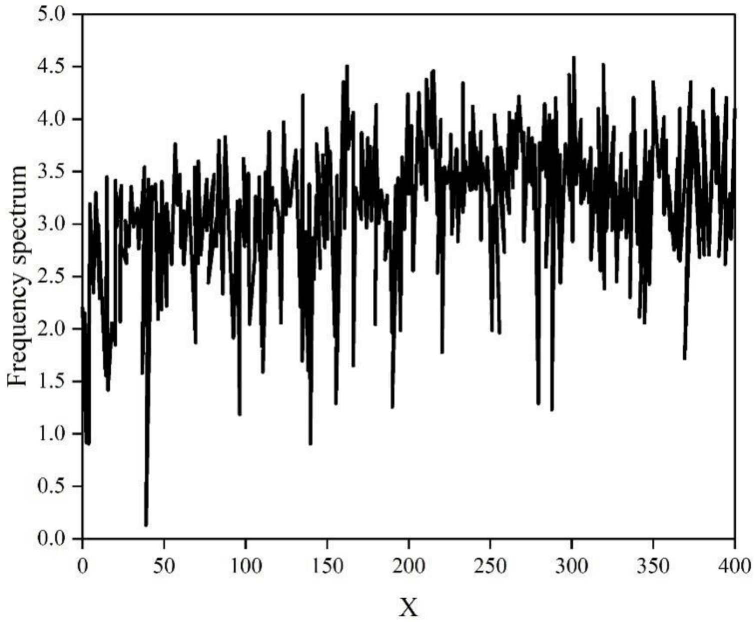

Figure 1. Local spectrum in x direction.

5.2 Analysis of Experimental Results

This section presents a quantitative analysis, user study on the FS2K dataset and the APDrawing dataset.

5.2.1. Quantitative Analysis

1) Face sketch image generation experiments

Face image generation experiments are conducted on FS2K dataset to compare the method of this paper with various state-of-the-art methods, including Pix2Pix, Pix2PixHD, CycleGAN, SCA-GAN, MDAL, FSGAN, SPADE, Genre. The evaluation metrics used include F-ID (evaluating the texture realism of the generated sketches), LPIPS (evaluating the image block similarity of the generated sketches), SCOOT (evaluating the sensory realism of the generated sketches), and FSIM (evaluating the feature similarity of the generated sketches). The results of the evaluation metrics are specifically shown in Table 1. In the table, ↓ indicates that the smaller the indicator, the better, and ↑ indicates that the larger the indicator, the better. From the table, it can be seen that the method of this paper performs the best, obtaining the best FID and LPIPS values of 14.76 and 0.224, respectively. Better SCOOT and FSIM values of 0.584 and 0.626 were also obtained, second only to SCA-GAN. The best FID and LPIPS values mean that the method in this paper produces highly realistic sketching results in terms of style and strokes. Moreover, the method in this paper obtains the second largest value of SCOOT, which is significantly better than FSGAN and Genre, and only slightly lower than SCA-GAN. Such a high SCOOT value means that the structure and texture of the sketches generated with the methods in this paper are sensually very similar to the results drawn by real artists.

Table 1. Experimental index comparison in FS2K data set.

Method	FID↓	LPIPS↓	SCOOT↑	FSIM↑
Pix2Pix	18.49	0.272	0.449	0.576
Pix2PixHD	32.49	0.453	0.324	0.581
CycleGAN	26.1	0.473	0.356	0.466
MDAL	49.98	0.532	0.374	0.572
SCA-GAN	39.27	0.318	0.642	0.815
FSGAN	35.27	0.495	0.408	0.512
SPADE	38.11	0.361	0.452	0.51
Genre	20.78	0.305	0.433	0.492
Method of this article	14.76	0.224	0.584	0.626

2) Experiments on face pen drawing image generation

In this section, the face pen drawing image generation experiments are conducted on the APDrawing dataset to compare the method of this paper with various advanced methods, including AdaIN, DRIT, DSMAP, CycleGAN, MUNIT, TUNIT, Pix2Pix, Pix2PixHD, and U2Net.The comparison results of the corresponding evaluation metrics are specifically shown in Table 2. By observing the data in the table, it can be concluded that the method of this paper performs the best as it obtains the lowest values in all the three metrics FID, CLEAN-FID and LPIPS. In particular, the FID, CLEAN-FID and LPIPS all decrease by about 7% or so when compared to the next best method, Pix2PixHD, which means that this paper's method produces highly realistic results in terms of style and strokes for the generation of pen drawing images. In addition, the relatively high SIFID may be responsible for its poor performance on individual images. Overall, the method in this paper significantly improves the quality of image-generated pen drawings compared to other existing methods, which are more similar to real drawings in terms of structure and texture.

Table 2. Experimental index comparison in APDrawing data set.

Method	FID↓	CLEAN-FID↓	SIFID↓	LPIPS↓
AdaIN	136.8	130.36	0.68	0.3355
DRIT	65.19	63.99	0.22	0.4396
DSMAP	71.71	68.46	0.23	0.4381
CycleGAN	102.7	100.06	0.27	0.4986
MUNIT	77.62	74.11	0.25	0.4995
TUNIT	92.08	89.01	0.29	0.4354
Pix2Pix	80.39	78.21	0.04	0.2454
Pix2PixHD	60.87	59.75	0.05	0.1765
U2Net	77.17	78.28	0.39	0.187
Method of this article	56.52	55.54	0.57	0.1465

5.2.2. User Studies

In order to further evaluate the visual effectiveness of sketches generated by the content generation and authoring techniques proposed in this paper for the media domain, this section will conduct a user study and statistical analysis based on qualitative analysis. In this study, a voting program for the quality of sketch generation was built using the Python language and the Tkiner standard library, and 10 non-art professionals were invited to participate in this subjective research experiment, and they were presented with a total of 1,000 samples randomly selected from the test set of the FS2K dataset. For each sample, the participants were shown the generation results obtained from the original face photographs, the real sketches drawn by the artist, the sketches generated by the technology migration in this paper, and the seven existing models (Pix2Pix, Pix2PixHD, FSGAN, SCA-GAN, MDAL, CycleGAN, and Genre). Participants were asked to choose the sketch portrait that they thought was the best based on the similarity between the model-generated sketches and the real sketches, as well as the quality of the generated sketches, following their own preferences. A total of 10,000 preference labels were collected for this user study. The average preference percentage for each model, and the standard deviation between different participants are specifically shown in Table 3. It is clear that the method of this paper significantly outperforms all other methods in terms of user preferences. From the average preference percentage of each model, this paper's method achieves the best generation results on roughly 70% of face photos. In addition, from the statistics of standard deviation, it can be found that there is also a slight difference between different subjective participants. The results of this user study amply demonstrate that the sketch portraits generated by this paper's technique significantly outperform other existing methods in terms of visualization.

Table 3. Statistical results of subjective voting results.

Method	Average	Standard deviation
Method of this article	70.91	1.69
Genre	8.01	0.15
Pix2Pix	9.04	0.68
Pix2PixHD	4.01	0.49
SCA-GAN	2.31	0.37
CycleGAN	3.11	0.33
FSGAN	1.38	0.27
MDAL	1.23	0.16

6. Evaluation of Public Art Creation and Design in the Field of Media

In this chapter, the content generation and creation technology in the media field proposed in this paper will be applied to the public art creation in the media field in Guanfeng Village in Ganbei, Jiangxi Province, China, and the public art gallery will be designed in the public resting area of the village. The neighboring villages of Wangkou, Xiaoqi, Likeng, and Huangling were selected as comparative objects to evaluate the design of public art creation in the public resting areas of the villages, and to explore the performance of the content generation and creation technology in the media field proposed in this paper.

6.1. Questionnaire Design

The questionnaire mainly collects information about respondents' purpose of traveling, satisfaction and demand for public art in public rest areas, the nature and art form of public art, art forms and existing problems of public art. For the public satisfaction survey, a five-level evaluation scale was adopted to facilitate respondents' understanding of the issues and to prevent the evaluation scale from being too rough to reduce the accuracy of the evaluation. According to the different questions, the scale is set as very poor, average, medium, relatively good and very good.

The questionnaire design was combined with the semantic differential method to determine that there are 25 adjective pairs that best describe the design of public art creations in the media field to describe the subjective satisfaction and feelings of the respondents. The specific semantic word pairs are shown in Table 4. It covers not only adjectives such as "complex-simple" and "natural-artificial" related to design modeling, but also adjectives related to design style such as "streamlined-geometric" and "dynamic-static".

Table 4. Adjective pair.

Adjective pair			
Bias positive adjectives	Serial number	Biased negative adjectives	Serial number
Adapted to the environment of the rest area	P1	Not adapted to the environment of the rest area	N1
Suitable distribution location	P2	Location distribution is not suitable	N2
Proportionally coordinated	P3	Incongruous proportion	N3
Can relieve fatigue	P4	can 't relieve fatigue	N4
Aesthetically pleasing	P5	Beauty-free	N5
Colorful	P6	Single color	N6
Warm	P7	Cold	N7
Figurative	P8	Abstract	N8
Interesting	P9	Uninteresting	N9
Interactive	P10	Isolated	N10
Natural	P11	Artificial	N11
The shape is changeable	P12	Single shape	N12
Complicated	P13	Simple	N13
Contactable	P14	Untouchable	N14
The content is traditional	P15	The content is modern	N15
Lovely	P16	Disgusting	N16
Attractive	P17	Non-attractive	N17
The atmosphere is relaxed	P18	The depressing atmosphere	N18
Innovative	P19	Antique	N19
Coordinated with the local landscape	P20	Incongruous with the local landscape	N20
Rich cultural connotation	P21	Lack of cultural connotation	N21
There is resonance	P22	There is no resonance	N22
Vital	P23	Lifeless	N23

Detail processing rich	P24	Detail processing is simple	N24
Overall comfortable	P25	Overall discomfort	N25

In this paper, 50 people were selected in each public rest area in each village to distribute questionnaires on the spot, so as to obtain more direct and immediate data, and then analyze them in the next step. A total of 250 questionnaires were distributed, excluding 18 invalid questionnaires, resulting in 232 valid questionnaires, with a validity rate of 92.8%.

6.2. Reliability Analysis

After finishing the recovered questionnaires, the reliability of the recovered samples needs to be analyzed, and the results have a certain high degree of internal consistency to make the next measurement analysis feasible. The reliability analysis of this paper is carried out in SPSS17.0 statistical software, and the results of the reliability analysis are specifically shown in Table 5. According to the test standard of Cronbrch α coefficient for 232 questionnaires in five villages for SPSS reliability analysis, we get Cronbrch $\alpha>0.8$, which proves that the questionnaire's questions are set with high confidence, and the consistency and stability of the measurement results are high, which can provide the basis for the next data analysis.

Table 5. Reliability test.

Village	sample size	Cronbrch α coefficient
Gaofeng Village	49	0.956
Wangkou Village	45	0.972
Xiaoqi Village	46	0.914
Likeng Village	45	0.948
Huangling Village	47	0.909

The reliability of the design of the variables in the questionnaire and the ability of the researcher to measure the variables correctly in the questionnaire are the basic basis of whether the experiment can be conducted. In this paper, by using the factor analysis function in SPSS17.0 statistical software, by analyzing 232 questionnaires from five villages, the results of validity analysis are specifically shown in Table 6. It can be seen that the KMO value of the questionnaires in each village is greater than 0.5, and the significance is less than 0.05, so that the next step of statistical analysis can be carried out.

Table 6. Validity test.

Village	-	Cronbrch α coefficient
Gaofeng Village	KMO value	0.748
	Approximate chi-square	1452.802
	Degree of freedom	300
	Sig.	0.001
Wangkou Village	KMO value	0.53
	Approximate chi-square	801.154
	Degree of freedom	300
	Sig.	0.001
Xiaoqi Village	KMO value	0.585
	Approximate chi-square	701.32
	Degree of freedom	300
	Sig.	0.002
Likeng Village	KMO value	0.814
	Approximate chi-square	1042.311
	Degree of freedom	300
	Sig.	0.000
Huangling Village	KMO value	0.732
	Approximate chi-square	1061.08
	Degree of freedom	300
	Sig.	0.001

6.3. Analysis of Evaluation Results

The collected questionnaires were data organized, and the valid samples were screened to derive the average value of 25 adjective pairs, resulting in the SD method evaluation scale scores as shown in Figure 2. Among them, the adjective pairs with more positive perceptions are closer to the value of 1, and the adjective pairs with more negative perceptions are closer to the value of 5. From the figure, we can see that the public evaluation of the village of Guanfeng, which applies this paper's content generation and creation technology in the media field to make public art in the public resting area, has the highest semantic evaluation of the whole among the five villages. Among the 25 adjective pairs, except for “abstract” in the adjective pair “abstract-figurative”, which has a slightly higher score than Wangkou, all the other adjective pairs with a positive bias have a higher score than Wangkou. It is clear that the public art creation in the public resting area of Guanfeng Village by applying this paper's content generation and creation technology in the media domain brings a good artistic and emotional experience to the public and is favored by the public. Contrary to the village of Grant Peak, Huangling Village public evaluation for all the villages in the lowest evaluation, the gap is more obvious, the main reason is that the service area of Huangling Village public art design components at least, can not be better to meet the public on the deep spiritual level of the needs of the public.

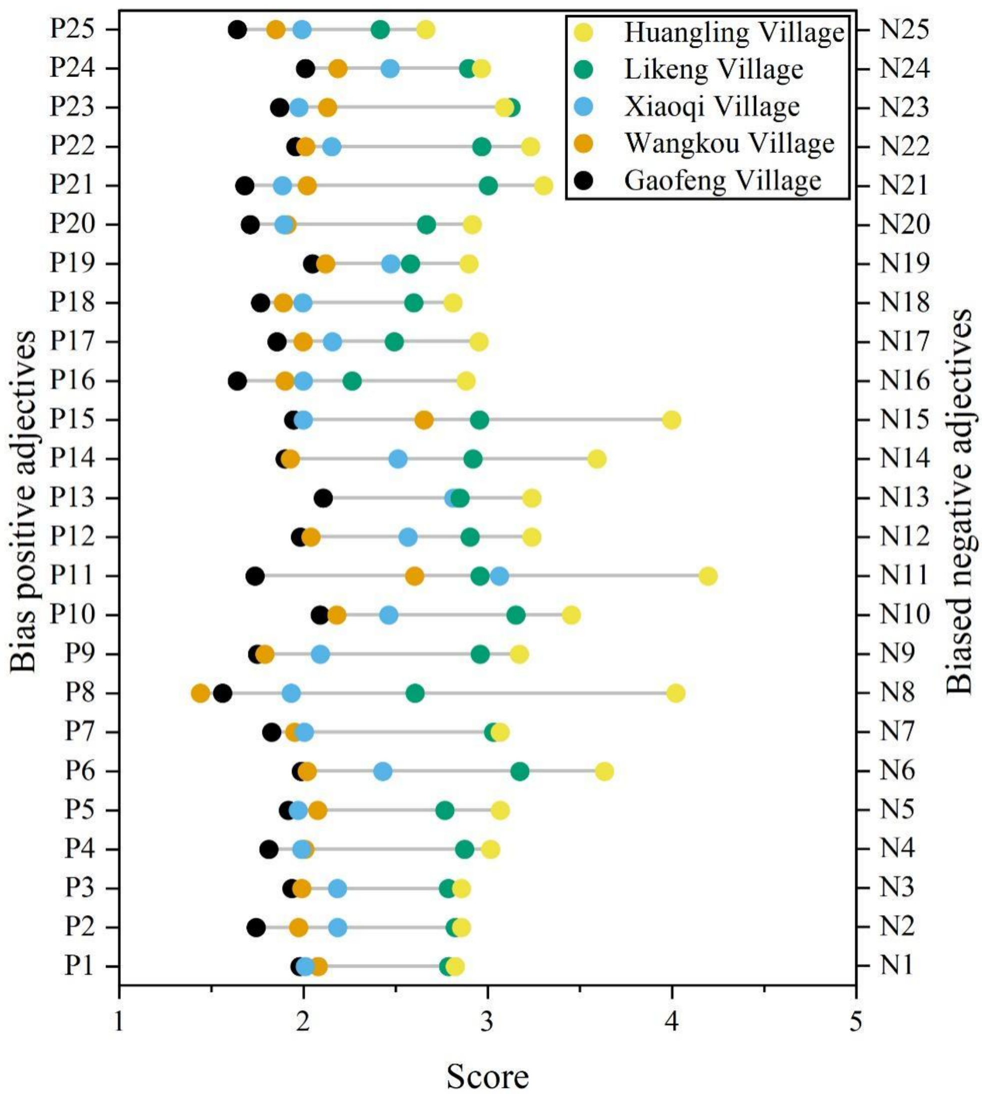

Figure 2. Public art creation design evaluation.

7. Conclusion

This paper proposes the way of generative adversarial network (GAN) application in AIGC, and focuses the research on the way of image generation application, and constructs the image art stylization model based on GAN by using image as the main medium of content creation in media field. On the basis of completing the image art style migration processing, the digital media content generation and creation technology based on computer vision technology is proposed.

In the face sketch image generation experiments on the FS2K dataset, this paper's method performs the best, with FID and LPIPS values of 14.76 and 0.224, respectively, and also obtains better SCOOT (0.584) and FSIM values (0.626), second only to SCA-GAN, and the structure and texture of the sketches it generates are very much sensory similar to the results drawn by the real artist similar. In the face pen drawing image generation experiments on the APDrawing dataset, this paper's method obtains the lowest values in the three metrics of FID, CLEAN-FID and LPIPS, which is the best performance, and generates highly realistic pen drawing images. Ten non-art professionals were invited to participate in the subjective research experiments, and the best generation results were achieved on about 70% of the face photos generated by this paper's method, the performance of the digital media content generation and creation techniques proposed in this paper was excellent, and the visual effects of the generated images were better than other methods.

The digital media content generation and creation techniques proposed in this paper were applied to the design of public art creation in the public resting area of Guanfeng Village in northern Ganbei, Jiangxi Province, China, and evaluated in comparison with neighboring villages such as Wangkou Village, Xiaoqi Village, Likeng Village, and Huangling Village. Except for the “abstract” in the “abstract-figurative” adjective pairs, which scores slightly higher than Wangkou Village, the positive adjective pairs in the remaining adjective pairs in Guanfeng Village are all closest to 1, which is better for public art creation in public rest areas. In the remaining adjective pairs, the values of positively-oriented adjective pairs are all closest to 1, which can better satisfy the public's emotional experience of art and is more favored by the public. This proves the utility of the content generation and creation techniques proposed in this paper in media content creation, and provides a feasible methodology for content creation design.

Acknowledgements
2025 Zhejiang Social Science Planning Project: "Research on the Interactive Mechanism of Elderly identity Construction and Social Integration in the Context of Short Video" (25NDJC020Z) and 2024 Teaching Model Innovation Project of Communication University of Zhejiang: "Exploring Pathways of New Media Content Creation" (KGXM202403).

References

1. Jantunen, A., Tarkiainen, A., Chari, S., & Oghazi, P. (2018). Dynamic capabilities, operational changes, and performance outcomes in the media industry. Journal of Business Research, 89, 251-257.
2. Zhang, L., & Liu, Z. (2024). Beijing's Global Media Offensive: China's Uneven Campaign to Influence Asia and the World, Joshua Kurlantzick (2022). International Journal of Media & Cultural Politics, 20(1).
3. Hu, Z., Ji, D., Xu, P., & Bhuju, K. (2020). Beyond convergence: Rethinking China’s media system in a global context. In BRICS Media (pp. 124-137). Routledge.
4. Tang, X., Huang, C., & Liu, R. (2017). National Strategy: A New Stage in the Development of China’s New Media. Development Report on China’s New Media, 3-33.
5. Wang, R. (2021). Computer-aided interaction of visual communication technology and art in new media scenes. Computer-Aided Design and Applications, 19(S3), 75-84.
6. Shan, D. (2023). New Media Development Model Based on Computer Network Technology. Procedia Computer Science, 228, 655-662.
7. Al Adwan, M. N., Mahmoud, M. A. A., Abdallah, R., Abokhoza, R., & Taha, S. (2023). The impact of artificial intelligence applications on media industries: a prospective study. J. Namibian Stud. Hist. Polit. Cult, 33, 721-734.
8. Prakash, G., & Sabharwal, D. (2024). AI Revolution in Online Media: Transforming Content Creation, Distribution, and Consumption. Media and Al: Navigating, 179.
9. Gentzkow, M. (2018). Media and artificial intelligence. Toulouse Network for Information Technology Whitepaper.
10. Deshmane, A., & Barriola, X. (2025). Frame by Fame: Content Creation on Short Video-Format Platforms. Manufacturing & Service Operations Management, 27(2), 479-495.
11. He, R., & Tian, H. (2023). Social media influencer and source credibility: endorsing content moderation on Douyin. International Journal of Communication, 17, 21.
12. Sun, Y. (2022). Research on the method of digital media content creation based on the internet of things. Computational Intelligence and Neuroscience, 2022(1), 8529875.
13. Wang, L., & Jiang, Y. (2025). New Media Content Creation and Recommendation Based on Artificial. Frontier Computing: Vol 3: Proceedings of FC 2024, 1357, 118.
14. Feng, X., Jiang, Y., Yang, X., Du, M., & Li, X. (2019). Computer vision algorithms and hardware implementations: A survey. Integration, 69, 309-320.
15. Hassaballah, M., & Hosny, K. M. (2019). Recent advances in computer vision. Studies in computational intelligence, 804(1-84), 3.
16. Hasan, M., Athrey, K. S., Khalid, A., Xie, D., Younessian, E., & Braskich, T. (2024). Applications of computer vision in entertainment and media industry. In Computer Vision (pp. 205-238). Chapman and Hall/CRC.

17. Ong'eta, W. (2024). Computer Vision and Building Resilience Against Online Violent Extremism in Africa: Computer Vision for Fake News Detection. Global Perspectives on the Applications of Computer Vision in Cybersecurity, 89-101.
18. Nanne, A. J., Antheunis, M. L., Van Der Lee, C. G., Postma, E. O., Wubben, S., & Van Noort, G. (2020). The use of computer vision to analyze brand-related user generated image content. Journal of Interactive Marketing, 50(1), 156-167.
19. Li, H., & Zhang, N. (2024). Computer Vision Models for Image Analysis in Advertising Research. Journal of Advertising, 53(5), 771-790.
20. Peng, Y. (2018). Same candidates, different faces: Uncovering media bias in visual portrayals of presidential candidates with computer vision. Journal of Communication, 68(5), 920-941.
21. Shi, R. (2024). Integrating computer vision and AI for interactive augmented reality experiences in new media. Applied and Computational Engineering, 102, 49-54.
22. Mantri, A. (2021). Real-Time Content Moderation Using Artificial Intelligence and Machine Learning. International Journal of Scientific and Engineering Research, 10, 1682-1684.
23. Pilsoo Lee. (2025). Inferring particle distributions in two-dimensional space with numerical features based on generative adversarial networks (GANs). Nuclear Engineering and Technology,57(10),103681-103681.
24. Ponamaryov V. V. & Kitov V. V. (2024). Automatic Image Style Transfer using an Augmented Style Set. Programming and Computer Software,50(3),231-237.
25. Aarthi D,Panimalar A,Santhosh Kumar S & Anitha K. (2024). Development of Adaptive Gaussian Filter Based Denoising as an Image Enhancement Technique. Mathematical Modelling of Engineering Problems,11(10).