JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 39, 55-66 (2023)
DOI: 10.6688/JISE.202301_39(1).0004

Research on Digital Media Intelligent Art Creation Based on the Fusion of Virtual Reality and Semantic Features*

YANG LIU[1], CUN SHANG[1,+] AND PENG-XIN LIU[2]
[1]*College of Planning and Design*
Xinyang Agriculture and Forestry University
Xinyang, 46400 P.R. China
[2]*School of Humanities and Law*
Zhengzhou Institute of Aeronautical Industry Management
Zhengzhou, 450046 P.R. China
E-mail: leadnat@163.com

With the rapid development of computer science, technologies such as virtual reality and semantic feature fusion have been greatly developed in recent years. In order to construct a digital media intelligent art creation model, this paper first designs a digital media intelligent creation framework based on virtual reality and maintains the original pattern and color distribution of natural images as accurately as possible. This method first performs automatic color segmentation and clustering on the image to extract key lines; then, re-colors the different segmented regions; and finally completes the detail enhancement. For the optimization of color matching, the algorithm integrates the contrast of the color block, and uses the differential evolution strategy to optimize the color configuration scheme. In addition, this paper also proposes a generative adversarial network-based on semantic feature fusion for digital media intelligent art creation. The network introduces multi-scale feature fusion by embedding the residual block feature pyramid structure and directly generates the final fine image by adaptively fusing these features. Only one discriminator can generate a realistic image of $256px \times 256px$. The experimental results have shown that this paper makes full use of virtual reality and semantic feature fusion technology, and obtains better results and better performance of intelligent art creation results on digital media than existing methods.

Keywords: virtual reality, semantic feature fusion, digital media, intelligent art creation, computer vision

1. INTRODUCTION

With the rapid development of current science and technology, virtual reality technology emerged at the historic moment. It is widely used as a new type of technology. This technology has been involved in the field of digital media art creation [1]. With the rapid progress of computer science, technologies such as virtual reality and semantic feature fusion have been considerably developed in recent years, in order to establish a digital media intelligent art creation model. In the process of the development of virtual reality technology, the application fields of digital media art continue to expand, and its creative

Received August 24, 2021; revised January 28, 2022; accepted March 5, 2022.
Communicated by Ching-Hsien Hsu.
+ Corresponding author.
* Xinyang Agriculture and Forestry University, Teacher Fund Project (201701002); The Education Department of Henan Province, Teacher Fund Project (2020-ZZJH-389).

methods are increasingly diversified in recent years [2]. In the environment of rapid development of virtual reality technology, it is very important to study the characteristics and concepts of virtual reality technology in order to better promote the development of digital media art [3]. This paper starts from the creation of digital media art based on virtual reality technology, and analyzes how to further advance the progress of digital media art creation. In the context of virtual reality art and semantic feature fusion algorithm, this paper aims to study the intelligent art creation of digital media. This article first designs a digital media intelligent creation framework based on virtual reality, and retains the original pattern and color distribution of natural images as faithfully as feasible.

In the context of virtual reality technology, computer-based stylization of natural images currently has many different types, such as cartoon-style [4], ink painting style [5], oil painting style [6], pen sketch style [7], *etc.* It is a fundamental issue of digital media intelligent art creation [8]. Based on semantic fusion features, generating realistic images from text is an important issue in digital media intelligent art creation, that is, inputting a text description to output an image containing the semantic information of the text has a wide range of applications [9]. Generative adversarial networks are widely used in text-to-image synthesis based on semantic feature fusion [10-12]. Since the generator-discriminator in the rough state generates defective results due to the lack of detailed information. This result will be used as input to promote the generation of the final result, resulting in the generation of unnatural local details and distorted global images [13].

In response to the above problems, we propose a new end-to-end framework, which is a generative confrontation network based on feature fusion (FF-GAN). The FF-GAN model introduces multi-scale feature fusion by embedding the feature pyramid structure of residual blocks, and through fusion these features to directly generate the final fine image, making the generated image more realistic and with higher semantic consistency. On the basis of semantic fusion features, the generation of realistic visuals from text is an essential topic in digital media intelligent art production.

2. DIGITAL MEDIA INTELLIGENT ART CREATION BASED ON VIRTUAL REALITY

2.1 Non-Photorealistic Rendering Based on Virtual Reality

Help the police find suspects by creating pictures of different bird types, *etc.* A wide range of text-to-image synthesis techniques uses adversarial networks. In the context of virtual reality technology, non-photorealistic rendering mainly refers to the simulating artistic drawing style, and the result does not have the reality like photos. It mainly highlights the personalized and abstract feelings of its artistic aspects, which can be effectively added as realistic graphics. It has a wide range of uses and frequently appears in the fields of art, industry, medicine, and entertainment. The main types of simulated paintings include oil painting, watercolor painting, pen drawing, pencil drawing, ink painting, and cartoon animation. Non-photorealistic rendering first appeared in the 1980s. [14] proposed an algorithm for the two-dimensionalization of a three-dimensional model, which depicts realistic objects with lines that can reflect the spatial characteristics of the original objects. This method mainly uses two-dimensional image operations to extract lines, corners, and curves with a sense of space. Compared with three-dimensional boundary tracking, this method has advantages in the real-time processing of the model. Virtual reality technology came at a critical time in science and technology progress. As a novel sort of technology, it is widely employed. In the subject of digital media art, this technology has been used extensively.

[15] researched the stylization of pen sketches at the SIGGRAPH meeting and proposed strokes and textures that mimic the style of pen sketches. The paper proposes an automated algorithm to construct sketch textures by simulating the stroke path, and adding sketch-specific shadows and occlusions on this basis, so as to convert the image into a sketch style. [16] introduced an interactive method to generate the final result according to the user's strokes and introduced a direction field in [17] so that the result not only reflects the light-dark relationship but also adds surface textures for users. The above papers focused on the generation of strokes and textures. For sketch works, only the distribution of light and shade is not added, so the color is not studied for the above research. In which genuine objects are depicted with lines that can reflect their spatial qualities in a two-dimensional format. Lines, corners, and curves with a sense of space are extracted using two-dimensional picture processes. It provides advantages in real-time processing over a three-dimensional boundary tracking technique.

2.2 Color Match Mapping Based on the Optimized Digital Media Artwork

Suppose that the total number of color blocks obtained by color clustering of the original natural image is m, and the number of selectable preset colors determined by the artwork image of a certain origin is k. The problem for color matching of art creation works is to determine a color scheme, replacing the colors ($P_1, P_2, \ldots, P_n$) of each color block on the original image with a preset color label ($C_1, C_2, \ldots, C_n$), where the value of $\{C_1, C_2, \ldots, C_m\}$ is from 1 to k. This paper defines the following color match mapping optimization energy equation:

$$(C_{t1}, C_{t2}, \ldots, C_{tn}) = \text{argmin}_{(C_{i1}, C_{i2}, \ldots, C_{in})}[E_1 + \varphi E_2] \tag{1}$$

where ($C_{t1}, C_{t2}, \ldots, C_{tn}$) is the optimal color scheme; φ is the contrast adjustment coefficient, and $\varphi = 0.5$ in the actual experiment; E_1 and E_2 represent the two aspects of the color mapping of the considered art creation work: Assuming that the new stylized image should maintain the original natural image's colors as closely as possible. Flower petals were once pink, and leaves were once dark green. This chromaticity should be maintained in the generated image of the art creation endeavor.

This means that even if the mutation operation is done first in repeated iterations of a search space, it will span the complete range of search space. Here, the resulting solution must be evaluated in terms of restrictions. It's necessary to discard solutions that are out of boundaries or do not meet specific requirements. This paper selects to use the differential evolution (DE) algorithm to find the optimal solution of the energy Eq. (1) [18]. For a given parameter carrier $x_{i,G}$, $x_{r1,G}$, $x_{r2,G}$ and $x_{r3,G}$ are randomly selected by the DE algorithm. In which the term of F is a constant, so the process is shown in Eq. (2):

$$v_{i,G+1} = x_{i,G} + K(x_{r1,G} - x_{i,G}) + F(x_{r2,G} - x_{r3,G}) \tag{2}$$

where $v_{i,G+1}$ represents a new solution derived from a certain mutation; i, $r1$, $r2$, and $r3$ are randomly selected and are different from each other; K and F are the combination constant and the scale constant of the difference, such two parameters are usually set as $K, F \in [0,2]$.

In the process of reorganization, a new solution $u_{i,G+1}$ is generated by $x_{i,G}$ and $v_{i,G+1}$. There is a certain probability that $v_{i,G+1}$ can be reorganized into $u_{i,G+1}$. This probability is determined by Cramer-Rao (CR) decisions. In which j represents the jth dimension of the solution, with constraints of $j = 1, 2, \ldots, D$, $rand \leftrightarrow U(0,1)$. I_{rand} is a random integer of $[1, 2, \ldots, D]$, and I_{rand} guarantees $v_{i,G+1} \neq x_{i,G}$. Therefore, the process can be defined as follows:

$$u_{j,i,G+1} = \begin{cases} v_{j,i,G+1}, \text{ if } rand_{j,i} \leq CR, \text{ or } j = I_{rand} \\ x_{j,i,G}, \text{ if } rand_{j,i} > CR, \text{ and } j \neq I_{rand} \end{cases}. \tag{3}$$

In the selection process, this paper applies the energy Eq. (1) mentioned above to select the optimal solution, eliminates inappropriate solutions, leaves suitable solutions, and performs optimization screening and elimination through iterations. When the loop ends, it can be considered that the optimal solution has been produced and the algorithm ends. In the color distribution discussed in this paper, each selectable color is regarded as a gene, and the selection and sorting of several colors each time is used as a chromosome. The population size is defined as a relatively large amount, which can generate more options [19]. After the iteration stops, the recorded color combination (that is, E reaches the minimum value) will be regarded as the best color scheme, which is used to replace the color of the original natural image. In the actual experiment, we select $K = 1$, $F = 1$, and $CR = 0.2$. When the loop is complete, the algorithm is said to be complete and the optimal solution has been found. All colors that can be selected are regarded as genes in this work, and the selection and sorting of several colors each time is employed as a genome.

3. DIGITAL MEDIA INTELLIGENT ART CREATION BASED ON THE FUSION OF TEXT SEMANTIC FEATURES

3.1 Generative Adversarial Network

In order to improve computing efficiency to complete high-quality digital media intelligent art creation, this paper proposes an end-to-end architecture, as shown in Fig. 1.

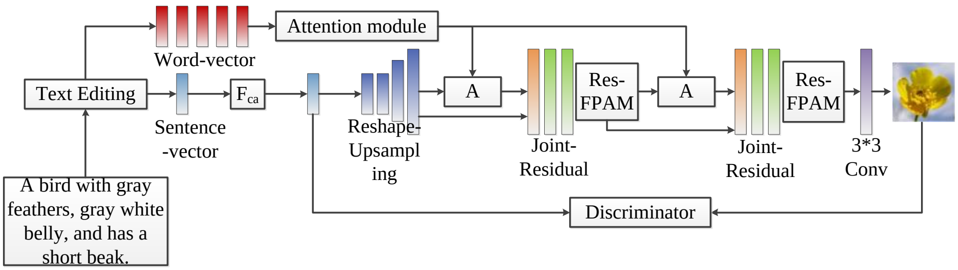


Fig. 1. Digital media intelligent art creation model based on semantic feature fusion.

The bottom-up feature fusion method combines the features in the low-resolution space with the features in the high-resolution space. The proposed feature fusion generation adversarial network contains three important stages, each with multiple inputs and single output, namely:

$$\begin{aligned} & s_i, e_i = \varphi(d_i); I_s^i = F_{ca}(s_i,); I_\omega^{i,j} = A_j^{attn}(e_i, F_j^i) \\ & \hat{t}_i, \left(F_1^i, ..., F_{j+1}^i\right) = G\left(I_s^i, I_\omega^{i,j}, z\right), j \in \{1,2\} \end{aligned} \tag{4}$$

where $\varphi(\ldots)$ is the text encoder designed in [20]; F_{ca} stands for conditional enhancement, which transforms the sentence vector s_i into the global conditional sentence vector I_s^i; $I_\omega^{i,j}$ is the conditional word vectors calculated by the attention model. The conditional word vector has two inputs: word vector e_i and image feature F_j^i.

In our proposed method, an image text pair (t_i, d_i) is a training sample. The generator generates the corresponding composite image $\hat{t}$. The global conditional sentence vector I_s^i is first combined with the noise vector z^i to extract the initial shallow features, and then connected with the conditional word vector $I_\omega^{i,j}$ in the next two stages for higher-level feature synthesis. Finally, we obtain F_1^i, F_2^i, …, F_{j+1}^i, which is a feature map generated by the generative adversarial network at different stages. Each feature map F_j is merged with the next feature map F_{j+1} for global feature fusion. The discriminator D has two training objectives:

(1) Evaluate whether the input image is real or fake;
(2) Determine whether the image text condition pair matches.

3.3 Semantic Feature Fusion

Although the feature map of lower resolution is coarser in space, it is more accurate in semantics. Lower-resolution features reflect the overall structure and color distribution of the image. In this paper, such features are sampled in the horizontal path and merged with the feature map generated by the main path, which helps to collect different features. For the semantic feature fusion, the addition and concatenation operators are two commonly used in researches. In each dense block, each layer is connected with subsequent layers to retain local feature information. The input and output of the current dense block are fused together through identity mapping.

Since such feature maps have the same size, residual learning across different levels is regarded as a weighted global fusion [21]. The output feature map of each stage is added with an up-sampling layer and the convolutional kernel is used to match the size of the feature map in other stages, thereby forming a multi-feature additional fusion layer to support continuous state transitions.

As shown in Fig. 2, our proposed generative adversarial network can be roughly divided into three steps. Step 1 is used to extract the shallow-level feature map F_1 from the concatenation of the sentence vector I_s and the random noise z, namely:

$$F_1 = H_1(z, I_s); F_{d,1} = \left[F_1, I_s, I_\omega^1\right] \tag{5}$$

where $F_{d,1}$ represents the concatenation of the text condition and the F_1 of the feature map, which is applied to deep-level feature fusion, namely:

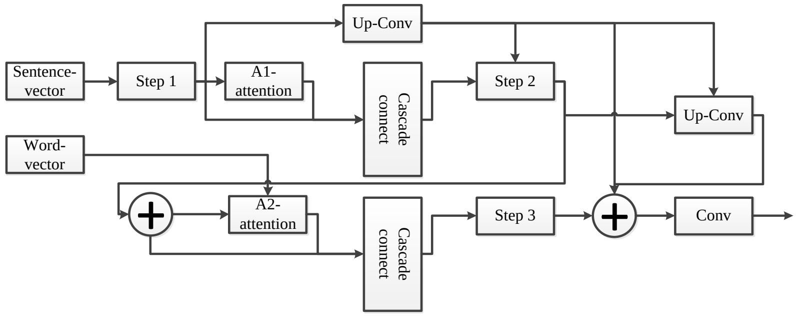


Fig. 2. Schematic diagram of semantic feature fusion.

$$F_2 = H_2(H_{dense}(F_{d,1}) + F_{d,1});\ F_{2,f} = F_2 + F'_1;\ F_{d,2} = [F_{2,f}, I_s, I^2_{\omega}]. \tag{6}$$

Feature fusion from the bottom-up integrates the low-resolution features with the high-resolution features. With numerous inputs and a single output, the proposed feature fusion generating adversarial network has three key steps. According to the research of [22], we use dense layers in the next two steps. The advantage of using dense layer is that it cannot only solve the problem of gradient disappearance to a certain extent, but also make better use of the feature information of different layers. $F_{2,f}$ is one of the global fusion feature maps realized through global residual learning, and $F_{d,2}$ is the concatenation result of the text condition and the fusion feature map $F_{2,f}$. Steps 2 and 3 have the same structure, so we can obtain F_3 in the same way, namely:

$$F_3 = H_3(H_{dense}(F_{d,2}) + F_{d,2}). \tag{7}$$

In order to utilize the features of each step in a global manner, our generator also uses global feature fusion at the end of the final step to adaptively fuse features from different levels, namely:

$$F_{3,f} = F_3 + F'_2 + F''_1. \tag{8}$$

In which, the terms can be defined as:

$$F''_1 = H_{L2}(H_{L1}(F_1));\ F'_2 = H_{L2}(F_2). \tag{9}$$

Through residual learning to directly fuse features in a weighted addition method, this method can reduce the parameters and calculation cost by half compared with the connection method.

4. SIMULATION EXPERIMENT AND RESULT ANALYSIS

4.1 Experiments and Results of Intelligent Digital Media Art Creation Based on Virtual Reality

The implementation environment of the intelligent digital media art creation system described in this paper is based on a 3.6GHz Intel Core i7-9770K processor, a Windows

10 operating system with 16GB of memory, and uses C++ programming. In this experiment, for the DE algorithm, the parameter setting of the color configuration part is specifically iterative 100 generations, and the population of each generation contains 500 possible solutions.

4.1.1 Color matching of different origins and comparison of results

This paper first analyzes the colors of several common art creation works and extracts color templates. This paper uses manual methods to extract common color pre-configurations in their respective origins. In the experiment, two images were stylized with creation works of art from different origins. It can be seen that the image calculated by the stylization algorithm of the proposed art creation work has the characteristics of less color block area, clear boundary contour, lively and warm color, *etc.*, which basically meets the characteristics of the art creation work. As an implementation of non-photorealistic rendering, the proposed algorithm can better express the style of artistic works, and it is easy to operate, eliminate the complicated process of artificially producing artistic works. A closer look at the result map shows that the color matching optimization algorithm of the proposed art creation work will introduce random effects while following the color distribution of the original image. Therefore, the different green areas on the leaves in the parrot map are not consistent.

4.1.2 Discussion on the performance of the color matching optimization of art works

It is worth noting that in the initialization step of the color matching optimization algorithm for art creation works, this paper did not use the usual random initial solution generation scheme. In this paper, a greedy algorithm was used to assign a value to each color block, and the color closest to each color block of the original image was directly selected as the first-generation solution. Based on the greedy algorithm, the optimization was performed and the results are as shown in Fig. 3. It can be seen from Fig. 3 that after using this optimization, it was more reasonable to select the first-generation solution in a completely random manner.

(a) Results from greedy algorithm. (b) Results from randomized generation.

Fig. 3. Comparison of initialization strategies for digital media intelligent art creation based on semantic feature fusion.

4.1.3 Time performance analysis

Table 1 counts the time required for the art creation work stylization system process based on the legends used in the paper. It can be seen from Table 1 that the stylization method proposed in this paper can obtain stylized results in a relatively short period of time. The time required for the algorithm is positively correlated with the number of iterations and the number of populations.

Table 1. Time complexity of intelligent art creation based on the integration of semantic features.

Images	Image Size/Pixels	Image Segmentation	Contour extraction	Color configuration	Detail enhancement
Parrot of (1)	450*315	5.8754	0.4121	13.1021	0.03547
Flower of (2)	550*430	7.8547	0.4565	5.8754	0.04253
Human of (3)	460*425	5.8487	0.5325	13.1258	0.04557
Bird of (4)	420*580	12.2154	0.6432	16.5459	0.04529

4.2 Experimental Results of Intelligent Media Art Creation Based on Semantic Feature Fusion

4.2.1 Time performance analysis

Data set used in the experiment: In order to verify the effectiveness of the model, experiments were carried out on the CUB and Oxford-102 data sets, and the data sets are shown in Table 2. We use the pre-trained text encoder provided by [23] to encode each sentence into a sentence embedding vector and a word embedding vector.

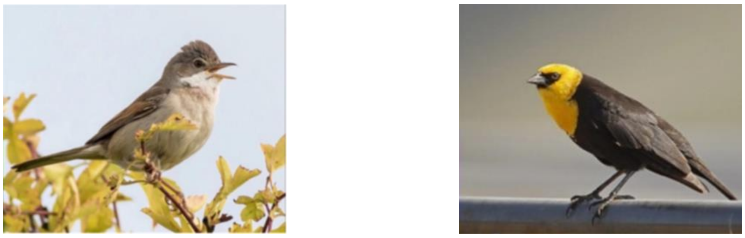
Table 2. Experimental dataset of intelligent media art creation based on semantic special fusion.

	CUB dataset		Oxford-102 dataset	
Number of images	Training	Testing	Training	Testing
	8855	2933	7034	1155
Number for using	20	20	15	15

4.2.2 Experimental results and comparative analysis

Fig. 4 shows the fine image generated by the input text description on the CUB data set, and the visualization result of the fusion feature map in the middle layer. Therefore, our proposed method produces realistic results consistent with the input text as intelligent media art creation. The input texts are given as:

(a) A bird with gray feathers, gray white belly, and has a short beak.
(b) A bird with yellow head, black feathers around body, and has a sharp beak.



(a) Results of sentence (a). (b) Results of sentence (b).

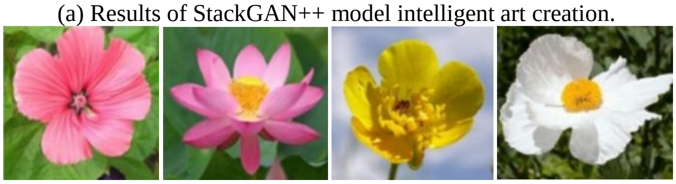
Fig. 4. The result of model synthesis trained on the CUB data set.

Fig. 5 shows the experimental comparison between our method and StackGAN++ [13] on the Oxford-102 flower test set. It can be seen from Fig. 5 that the shape and color of the flowers generated by our proposed method are more real. The flowers generated by StackGAN++ are very fuzzy, there is no clear layering between the petals, and the stamens have problems such as unclear shapes and wrong colors. Therefore, the text cannot be fully displayed for the characteristics of petals and stamens in the description. The input texts are given as:

(i) A flower has large pink petals and a pink stamen.
(ii) A flower has pale pink petals and a cluster of yellow stamen.
(iii) A flower has many yellow stamens surrounded by yellow and large petals.
(iv) A flower has large with petals and a yellow stamen.

(i) (ii) (iii) (iv)

(a) Results of StackGAN++ model intelligent art creation.



(i) (ii) (iii) (iv)

(b) Results of the proposed model intelligent art creation.

Fig. 5. Digital media intelligent art creation results on the Oxford-102 dataset.

An algorithm for extracting image lines using Mean-Shift clustering is used in this paper before proposing an energy equation to measure the color configuration. Finally, effects like paper texture and color blooming enhance the image details, bringing it closer to the original artwork. By using the method in [14], we adopt pre-trained on the ImageNet dataset to evaluate the IS values and FID values. Table 3 lists the comparison of IS values and FID values between different methods. The IS values of our proposed method on such two data sets are 3.744 and 3.324, which are higher than many models proposed in the previous references. The continuous decrease of the FID values has also shown that the samples generated based on our proposed method are closer to the real image.

Table 3. The results of the IS values and FID values of the model on the two test data sets.

Comparative methods	Different datasets for digital media intelligent art creation			
	CUB dataset		Oxford-102 dataset	
	IS index	FID index	IS index	FID index
GAN-INT-CLS [11]	2.56±0.02	–	2.51±0.02	–
StackGAN [14]	3.58±0.04	32.65	3.16±0.05	55.17
StackGAN++ [13]	3.79±0.03	18.27	3.31±0.02	48.26
ResFPA-GAN [12]	4.27±0.04	16.39	3.72±0.04	43.05
Our proposed	4.52±0.05	13.25	3.92±0.05	38.29

5. CONCLUSIONS

In order to apply virtual reality and semantic feature fusion technology to the intelligent art creation of digital media, this paper first uses the Mean-Shift clustering algorithm to extract image lines, then proposes an energy equation to measure the color configuration, and finally enhances the details of the image by adding effects such as paper texture and color blooming to make the picture closer to the real creation work of art. This stylization process can automatically convert a natural image into the style of a work of art, and has a better visual effect. In addition, this paper proposes a generative adversarial network based on feature fusion, and successfully applies it to the task of text synthesizing images. The proposed FF-GAN model introduces multi-scale feature fusion by embedding the residual block feature pyramid structure. Compared with the previous model, our proposed FF-GAN model does a better job in generating consistent and high-quality images. This is due to the fusion of hierarchical feature maps can fully extract and utilize local and global features. The experimental results have shown that the digital media art images generated by our model are not only richer and more delicate in color, but also perform better in some local subtleties.

REFERENCES

1. S. S. C. Chen, H. Cui, P. Tan, X. Sun, Y. Ji, and H. Duh, "Cantonese porcelain image generation using user-guided generative adversarial networks," *IEEE Computer Graphics and Applications*, Vol. 40, 2020, pp. 100-107.
2. Y. Yu, A. Srivastava, and S. Canales, "Conditional lstm-gan for melody generation from lyrics," *ACM Transactions on Multimedia Computing, Communications, and Applications*, Vol. 17, 2021, pp. 1-20.
3. S. N. Esfahani and S. Latifi, "A survey of state-of-the-art gan-based approaches to image synthesis," in *Proceedings of Computer Science and Information Technology*, 2019, pp. 63-76.
4. J. Agnese, J. Herrera, H. Tao, and H. Zhu, "A survey and taxonomy of adversarial neural networks for text-to-image synthesis," *Wiley Interdisciplinary Reviews: Data*

Mining and Knowledge Discovery, Vol. 10, 2020, p. e1345.
5. Y. Men, Y. Mao, Y. Jiang, W. Y. Ma, and Z. Lian, “Controllable person image synthesis with attribute-decomposed gan,” in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5084-5093.
6. M. Qi, Y. Wang, A. Li, and J. Luo, “STC-GAN: Spatio-temporally coupled generative adversarial networks for predictive scene parsing,” *IEEE Transactions on Image Processing*, Vol. 29, 2020, pp. 5420-5430.
7. L. Guarnera, O. Giudice, and S. Battiato, “Fighting deepfake by exposing the convolutional traces on images,” *IEEE Access*, Vol. 8, 2020, pp. 165085-165098.
8. T. Arora and R. Soni, “A review of techniques to detect the GAN-generated fake images,” *Generative Adversarial Networks for Image-to-Image Translation*, 2021, pp. 125-159.
9. A. Notaro, “State-of-the-art: Ai through the (artificial) artist’s eye,” *Electronic Visualisation and the Arts*, 2020, pp. 322-328.
10. P. R. Siavelis, N. Lamprinou, and E. Z. Psarakis, “An improved GAN semantic image inpainting,” in *Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems*, 2020, pp. 443-454.
11. G. Zou, Z. Qin, Q. He, P. Wang, B. Zhang, and Y. Gan, “DeepWSC: clustering web services via integrating service composability into deep semantic features,” *IEEE Transactions on Services Computing*, Vol. 15, 2020, pp. 1940-1953.
12. R. Zhang, C. Chen, Z. Gan, Z. Wen, W. Wang, and L. Carin, “Nested-Wasserstein self-imitation learning for sequence generation,” in *Proceedings of International Conference on Artificial Intelligence and Statistics*, 2020, pp. 422-433.
13. Z. Liu, J. Deng, L. Li, S. Cai, Q. Xu, S. Wang, and Q. Huang, “IR-GAN: Image manipulation with linguistic instruction by increment reasoning,” in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 322-330.
14. H. Tan, X. Liu, B. Yin, and X. Li, “Cross-modal semantic matching generative adversarial networks for text-to-image synthesis,” *IEEE Transactions on Multimedia*, Vol. 24, 2021, pp. 832-845.
15. X. Mao, S. Wang, L. Zheng, and Q. Huang, “Semantic invariant cross-domain image generation with generative adversarial networks,” *Neurocomputing*, Vol. 293, 2018, pp. 55-63.
16. Z. Ma, J. Li, N. Wang, and X. Gao, “Semantic-related image style transfer with dual-consistency loss,” *Neurocomputing*, Vol. 406, 2020, pp. 135-149.
17. P. Andreini, S. Bonechi, M. Bianchini, A. Mecocci, and F. Scarselli, “Image generation by GAN and style transfer for agar plate image segmentation,” *Computer Methods and Programs in Biomedicine*, Vol. 184, 2020, p. 105268.
18. X. Ye, J. Du, and Y. Ye, “MasterplanGAN: Facilitating the smart rendering of urban master plans via generative adversarial networks,” *Environment and Planning B: Urban Analytics and City Science*, 2021, p. 23998083211023516.
19. M. Cheng, Y. Li, S. Nazarian, and P. Bogdan, “From rumor to genetic mutation detection with explanations: a GAN approach,” *Scientific Reports*, Vol. 11, 2021, pp. 1-14.
20. Y. Li, Y. Cheng, Z. Gan, L. Yu, L. Wang, and J. Liu, “Bachgan: High-resolution image synthesis from salient object layout,” in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8365-8374.
21. C. Zhang, Z. Yang, X. He, and L. Deng, “Multimodal intelligence: Representation lea-

rning, information fusion, and applications," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 14, 2020, pp. 478-493.
22. A. Fares, S. Zhong, and J. Jiang, "Brain-media: A dual conditioned and lateralization supported GAN (DCLS-GAN) towards visualization of image-evoked brain activities," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 1764-1772.
23. L. Rossi, M. Paolanti, and R. Pierdicca, "Human trajectory prediction and generation using LSTM models and GANs," *Pattern Recognition*, Vol. 120, 2021, p. 108136.

Yang Liu was born in HeNan, China, in 1986. From 2004 to 2008, she studied in Zhengzhou University of Light Industry and received her bachelor's degree in 2008. From 2008 to 2011, she studied in HeNan University and received her Master's degree in 2011. Currently, she works in Xinyang Agriculture and Forestry University. She has published ten papers. Her research interests include interaction design and graphic design. Email: xiaomeiyaliu@163.com

Cun Shang was born in HeNan, China, in 1982. From 2001 to 2006, he studied in Henan University of Science and Technology and received his bachelor's degree in 2006. From 2006 to 2017, he worked in Xinyang Agriculture and Forestry University, HeNan. From 2007 to 2010, he studied in XinYang Normal University and received his Master's degree in 2010. From 2017 to 2021, he studied in Yeungnam University and received his Master and doctor's degree in 2021. Currently, he works in Xinyang Agriculture and Forestry University. He has published ten papers. His research interests include computer graphic and interaction design. Email: leadnat@163.com

Peng-Xin Liu was born in HeNan, China, in 1992. From 2010 to 2014, he studied in Henan University and received his bachelor's degree in 2014. From 2014 to 2017, he studied in Wuhan Institute of Physical Education and received his Master's degree in 2017. Currently, he works in Zhengzhou Institute of Aeronautical Industry Management. He has published five papers. His research interest includes science of education. Email: liupengxin2021@126.com