IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 April 2023, accepted 2 May 2023, date of publication 9 May 2023, date of current version 21 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3274666

RESEARCH ARTICLE

TwinGAN: Twin Generative Adversarial Network for Chinese Landscape Painting Style Transfer

DER-LOR WAY[1], CHANG-HAO LO[2], YU-HSIEN WEI[3], AND ZEN-CHUNG SHIH[2]

[1]Department of New Media Art, Taipei National University of the Arts, Taipei 112, Taiwan
[2]Institute of Multimedia Engineering, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan
[3]Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan

Corresponding author: Der-Lor Way (adler@newmedia.tnua.edu.tw)

This work was supported by the Ministry of Science and Technology, Taiwan, for financially supporting this research under Contract No. MOST 110-2221-E-119-002.

ABSTRACT Recently, style transfers have received considerable attention. However, most of these studies were suitable for Western paintings. In this paper, a deep learning method is proposed to imitate multiple styles of Chinese landscape paintings. Twin generative adversarial network style transfer was proposed based on the characteristics of Chinese landscape ink paintings. SketchGAN and renderGAN were performed using generative models based on generative adversarial networks. The SketchGAN involves determining the structure and simplifying the content of an input image. RenderGAN involves transferring the results of sketchGAN into the final stylized image. Moreover, a loss function was designed to maintain the shape of the input content image. Finally, the proposed TwinGAN was successfully used to imitate five styles of Chinese landscape ink paintings. This study also provided ablation studies and comparisons with previous works. The experimental results show that our algorithm synthesizes Chinese landscape stylized paintings that are higher in quality than those produced by previous algorithms.

INDEX TERMS Deep neural networks, style transfer, generative adversarial network (GAN), loss function, Chinese landscape painting.

I. INTRODUCTION

Deep learning is widely used in computer vision, computer graphics, and other fields. Gatys et al. [1] were the first to develop a convolutional neural network (CNN) scheme for style transfer. They found that correlations between features could be used as a basis for accurate style transfer and texture synthesis. The image-style transfer method proposed by the aforementioned authors mainly focuses on the transformation of the global style. This method is suitable for Western paintings, because the styles of all abstract paintings are often similar. However, existing style transfer algorithms perform poorly in Chinese landscape paintings. Rock textures (''Tsun'') represent the orientation of mountains in Chinese landscape paintings, and different landscape painting skills are required depending on the type of rock.

In addition, the Chinese concept of perspective, which differs from the scientific perspective of the West, is an idealistic approach that allows one to depict more than what can be seen with the naked eye. Based on the drawing process for Chinese landscape ink paintings, a twin generative adversarial network (GAN) style transfer method was designed in this study. TwinGAN involves sketching and rendering based on a generative adversarial network (GAN), as shown in Figure 2. First, SketchGAN was used to convert an input photo into a brushstroke-like composition sketch. Second, the sketch image produced in SketchGAN is converted into a Chinese landscape painting using RenderGAN. A reconstruction loss function is used to ensure that the stylized image maintains the shape of the content image. This function calculates the errors of the image features extracted by the extended differenceB-ofB-Gaussian (XDoG) filters. In this study, the proposed method was used to transfer five major styles of Chinese landscape paintings, as shown in Figure 1. The main contributions of this study are as follows: First, TwinGAN

The associate editor coordinating the review of this manuscript and approving it for publication was Mingbo Zhao.

FIGURE 1. Five styles of chinese landscape painting experimental results by the proposed TwinGAN.

based on a GAN was developed to transfer Chinese landscape ink painting styles. Second, XDoG reconstruction loss was designed to ensure that the stylized image retained the shape of the input content image.

II. RELATIVE WORKS

Goodfellow et al. [2] proposed a GAN consisting of a generator and discriminator. The generator produces data to fool the a discriminator, which distinguishes the real data from fake data. The generator can synthesize more realistic data through competition between the generator and discriminator. GAN are widely used for image synthesis. Radford et al. [3] proposed a deep convolutional GAN that could synthesize an image from a noise vector. Isola et al. [4] proposed a pix2pix model that used a label map to synthesize an image. Wang et al. [5] used a coarseB-toB-fine generator and multi-scale discriminator to enhance the resolution of an image for the pix2pix model. The disadvantage of the pix2pix model is that it must be trained using paired data that are difficult to gather in the real world. Therefore, numerous methods [6], [7] based on unpaired data have been proposed.

Neural style transfer refers to changing the style of an image using neural network algorithms. CNNs can be used to extract high-level semantic features from an image. The synthesized image can be modified by manipulating the features of the neural network. Gatys et al. [1] were the first to use a neural network to synthesize artistic paintings. They combined the content of the photo and the style of the artwork by minimizing the feature loss in a pre-trained CNN; however, their iterative optimization-based approach was too slow. To solve this problem, numerous methods based on an encoder–decoder architecture have been proposed. This architecture was used to extract the input features. The encoder reduces the dimensionality of the input and transforms it into latent space. Subsequently, the decoder converts the input back to the original space. The output can be modified by doing something in the latent space. The encoder–decoder architecture is typically used as a generative model for image synthesis. Johnson et al. [8] and Uluanov et al. [9] used a pre-trained VGGB-16 (Visual Geometry Group) network as a discriminator to train a generative model based on the encoder–decoder architecture. However, the performances of the aforementioned method and the Gatys method are the same because the aforementioned method involves the use of the object function presented by Gatys et al. [1].

Most studies used a single model to transfer multiple styles. Dumoulin et al. [10] suggested a conditional instance normalization method, in which affine parameters are learned in the instance normalization layer for every style. Chen et al. [11] used multiple convolution filters (called a style bank) to learn different styles individually. However, their method simply generated the style used in the training stage, and the model size increased with the number of styles. Therefore, achieving the transfer of an arbitrary style has become at focus of research. Chen and Schmidt [12] added a swap layer to a neural network that can substitute features similar to content features. Huang et al. [13] recommended an adaptive instance normalization method to achieve style transfer by adjusting the variance and mean of features.

Li et al. [14] performed whitening and color transformations to stylize content images. However, the performance of their method for arbitrary style transfers was relatively poor. Way et al. [15] performed a semantic-based segmentation of video-style transfers. Given a video and two styled images as inputs, their segmentation involves foreground and background objects, followed by object style transfer. This methodology overcomes the problems of incorrect motion boundaries, choppy segmentation, and low-level noise in the stylized images. In addition, Ulyanov et al. [16] found that instance normalization can improve stylization quality. Liu et al. [17] proposed a loss function that can retain depth information in stylized images. Jing et al. [18] controlled the stroke size by controlling the receptive fields. Furthermore, Luan et al. [19] proposed a photoB-realistic style transfer method that uses a photorealism regularization term to prevent the distortion of the content structure.

Chen et al. [20] proposed a dual-style-learning artistic style transfer to simultaneously learn the overall artistic style and specific art style. Yang et al. [21] presented a recurrent convolutional neural subnetwork to control stroke size. Their method not only delivers more flexibility for simplifying to unseen grander stroke sizes, but also produces multiple stroke sizes with only one residual unit. In addition, they proposed two runtime control approaches to concurrently control the style and stroke size in a feed-forward style.

Some studies have provided Chinese-painting-style transfers using generative adversarial networks. Xue [22] proposed a sketch-and-paint GAN that created Chinese landscape paintings end-to-end without conditional input. Lv and Zhang [23] studied the impact of different loss functions and generator models on training time and results under this framework. They also applied CycleGAN to reduce the training time. Wang et al. [24] provided a high-quality ChinaStyle dataset containing 1913 images with six categories. Mask-aware generative adversarial networks have been used to transfer different styles of Chinese painting from realistic portraitures. Yang et al. [25] applied a visual geometry group (VGG) network to perform style transfer, which was trained on a pair of content and style images to output an image that rendered the target content with the desired style.

III. PROPOSED APPROACH

The proposed TwinGAN scheme was inspired by the painting process. When artists paint, they first roughly sketch the shapes of objects based on their intuition and then complete the final work according to a rough sketch. In accordance with this process, a two-stage TwinGAN style transfer method comprising two stages was developed, as shown in Figure 2. SketchGAN and RenderGAN were used to conduct the cross-domain transfers. The goal of SketchGAN is to perform an image-composition sketch, and the goal of RenderGAN is to transform the abstracted image obtained in the sketching into the final stylized image.

A. STRUCTURE EXTRACTION

Chinese landscape paintings emphasize the implicit meanings of objects. Therefore, brushstrokes and ink tones are often used to depict the shape of the landscape. Therefore, the structure of an image's content is the most important feature of Chinese landscape paintings. Mountains often appear in Chinese landscape paintings; therefore, this study focused on identifying mountain structures.

Several algorithms have been proposed for edge detection in image processing. Figure 3 illustrates the results obtained when an input image is processed using XDoG [26] and holistically nested edge detection (HED) algorithms [27]. The HED algorithm [27] can accurately extract the outline of the image content, and the XDoG algorithm can extract clearer, cleaner, and more detailed edges than other methods. The XDoG algorithm accurately extracts the structure of a mountain; therefore, it was used for the structure extraction in this study. The aforementioned algorithm is formulated as follows.

$$D_{\sigma,k,\tau}(x) = G_{\sigma}(x) - \tau \cdot G_{k\sigma}(x) \tag{1}$$

where x is the input image, G is the Gaussian filter, σ is the standard deviation of a Gaussian distribution, D is the difference between two Gaussian filters, namely G_{σ} and $G_{k\sigma}$, and τ denotes the tradeB-off parameter between these two Gaussian filters.

$$\mathrm{T}_{\varepsilon,\varphi}(D) = \begin{cases} 1, & D \geq \varepsilon \\ 1 + \tanh(\varphi \cdot (D - \varepsilon)), & otherwise \end{cases} \tag{2}$$

where $T_{\varepsilon,\varphi}$ is a thresholding function with a continuous ramp, ε and φ are the parameters that control the threshold.

B. MODEL ARCHITECTURE

The problem of this study is a domain transfer problem and can be described as follows: Given a dataset X as the source domain and another dataset Y as the target domain, the goal is to find a function G that can map domain X to domain Y. To achieve this goal, the GAN was used in this study. The proposed system consists of two models: the generative network G and the discrimination network D, as shown in Figure 4.

Because the problem in this study is image-to-image translation, the goal of the adopted generative network is to alter the appearance of an image. We refer to the study by Johnson et al. [8] to define the architecture of the generative network used in this study. This network is based on an encoder–decoder architecture and contains several residual blocks as transformers. As shown in Figure 4, encoder **f** contains three convolutional layers. It first transforms an input image sampled from the target domain from the pixel space into the latent space. Subsequently, the latent code is mapped from source domain X to target domain Y through nine residual blocks. Finally, decoder **g**, which consists of three deconvolutional layers, transforms the latent code from the latent space back into the pixel space. The style of the

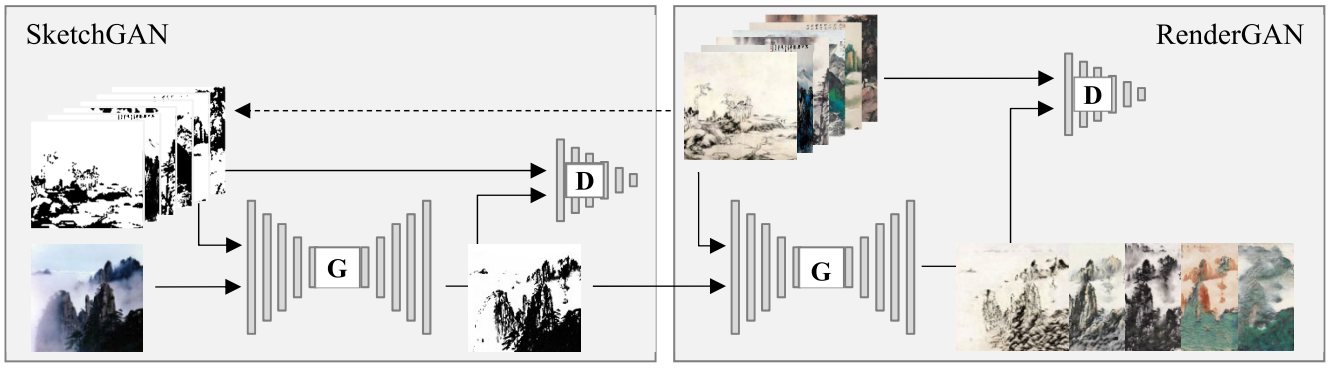


FIGURE 2. Twin generative adversarial network for chinese landscape painting style transfer.

(a) Input image (b) XDoG (c) HED

FIGURE 3. (a) Input image, (b) image obtained after using the XDoG algorithm, and (c) image obtained after using the HED algorithm.

input image was similar to that of the image sampled from the target domain.

The discrimination network used was PatchGAN [4]. In contrast to a normal GAN discriminator, which maps an image to a scalar, PatchGAN maps an image to an $N \times N$ array. The value of this array depends on the local patch of the image and is used to calculate the GAN loss. The quality of the patch can be improved by identifying real or fake data in a local patch. The patch size was equivalent to the receptive field of the CNN. Therefore, the patch size can be varied by stacking convolutional layers. The discriminator network consisted of five convolutional layers, and the patch size used for discrimination was 70 × 70 pixels. Jing et al. [18] found that the scale of a style image can affect the stroke size. Therefore, a multiscale discriminator network [5] was used in this study to train the generative network. The multiscale discriminator contains three discriminators that can discriminate between images of different sizes (256 × 256, 128 × 128, and 64 × 64). A generative network can improve the quality of a stylized image by calculating the adversarial loss of the multiscale discriminator.

C. LOSS FUNCTION

The loss function used to train the developed model is based on the loss function proposed by Taigman et al. [28], who used three loss terms to achieve cross-domain transfer. Taigman's loss function is suitable for aligned data, including facial photos and emojis. However, an unaligned dataset was used in the present study; thus, preserving the content of the input image in the stylized image is difficult. Consequently, we added a new loss term to the loss function of Taigman et al. to solve the aforementioned problem. The loss function is expressed as follows.

$$\mathcal{L}_{total} = \lambda_{adv} \sum_{k=1}^{3} \mathcal{L}_{adv}^{k} + \lambda_{const} \cdot \mathcal{L}_{const} + \lambda_{identity} \cdot \mathcal{L}_{identity} + \lambda_{rec} \cdot \mathcal{L}_{XDoG\_rec} \tag{3}$$

where λ_{adv}, λ_{const}, $\lambda_{identity}$, and λ_{rec} are tradeB-off parameters between the loss functions; L_{adv}^{k} is the adversarial loss, which is calculated by the kth discrimination network; L_{const} and $L_{identity}$ are the encoderB-constancy loss and identity loss proposed by Taigman et al. [28], and L_{XDOG_rec} is the proposed XDoG reconstruction loss. As shown in Equation (3), four loss functions were used to train the network. The function L_{adv}^{k} makes the synthesized image approximate the stylized image; L_{const} and $L_{identity}$ enable the generative model to conduct mapping more easily; and L_{XDOG_rec} allows the stylized image to preserve the content of the input image.

1) ADVERSARIAL LOSS

Consider a dataset Y containings stylized images as the target domain. The task of the generative model G is to transfer an image sampled from another dataset X to the target domain. To make the generative model G synthesize images similar to the images sampled from dataset Y, adversarial discrimination network D was used to train G. Network D can be formulated as follows.

$$\mathcal{L}_{adv} = \min_{G} \max_{D} \mathbb{E}_{y \backsim P_y} log D\,(y) + \mathbb{E}_{x \sim P_x} \log(1 - D\,(G(x))) \tag{4}$$

where y is sampled from the data distribution P_y, x is sampled from the data distribution P_x, $G(x)$ denotes the data generated by the generative network G, and D denotes the discrimination network.

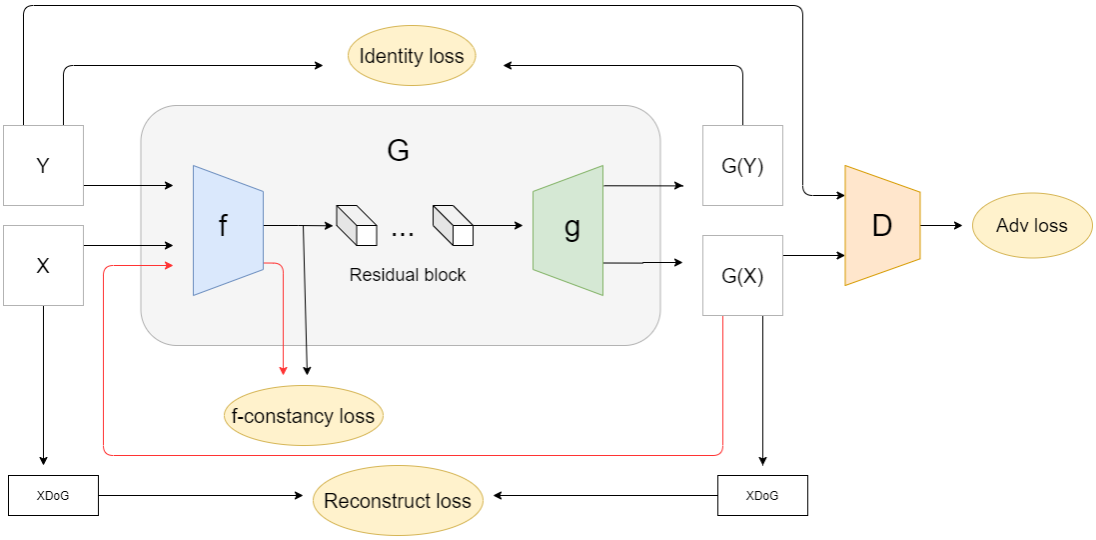
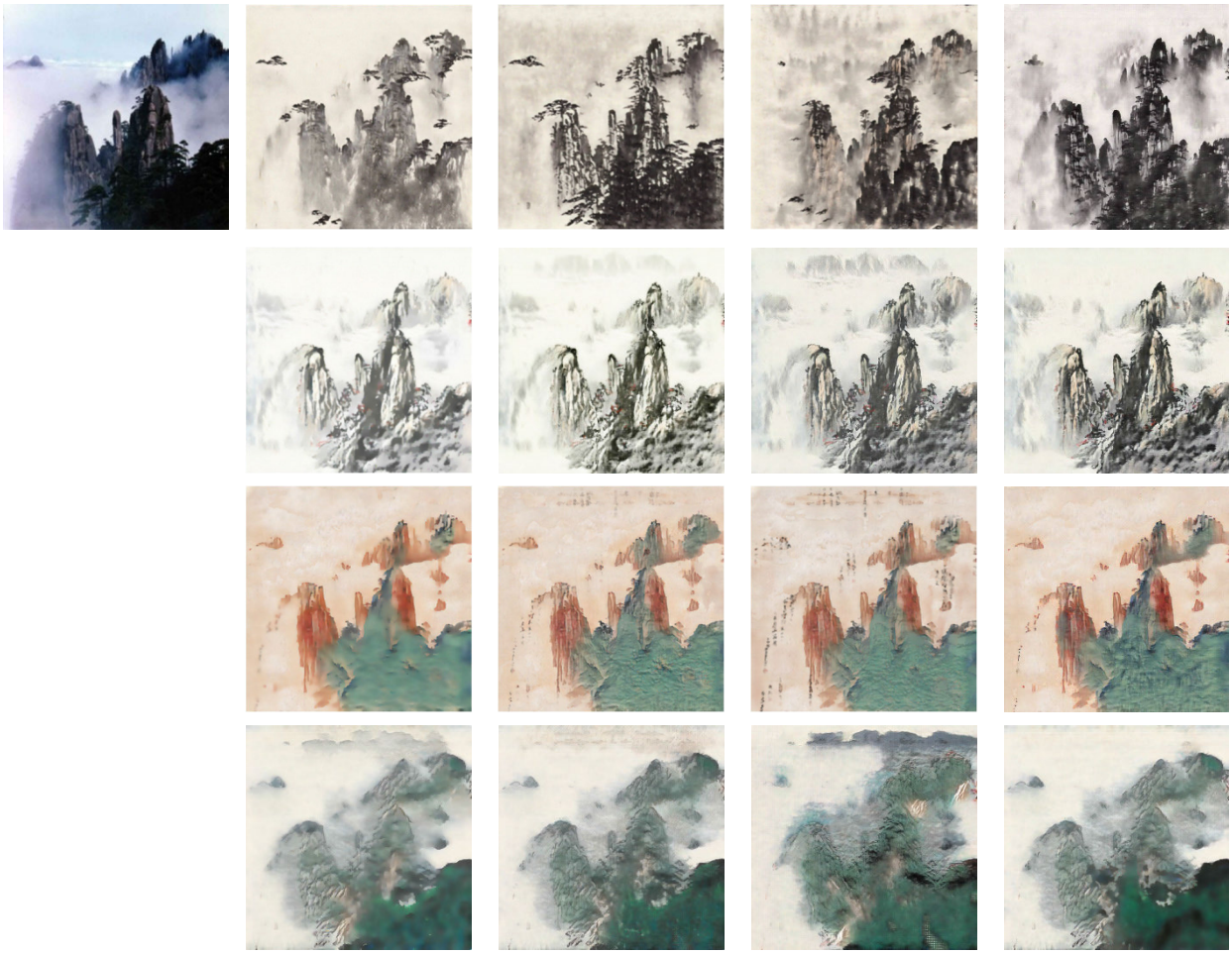
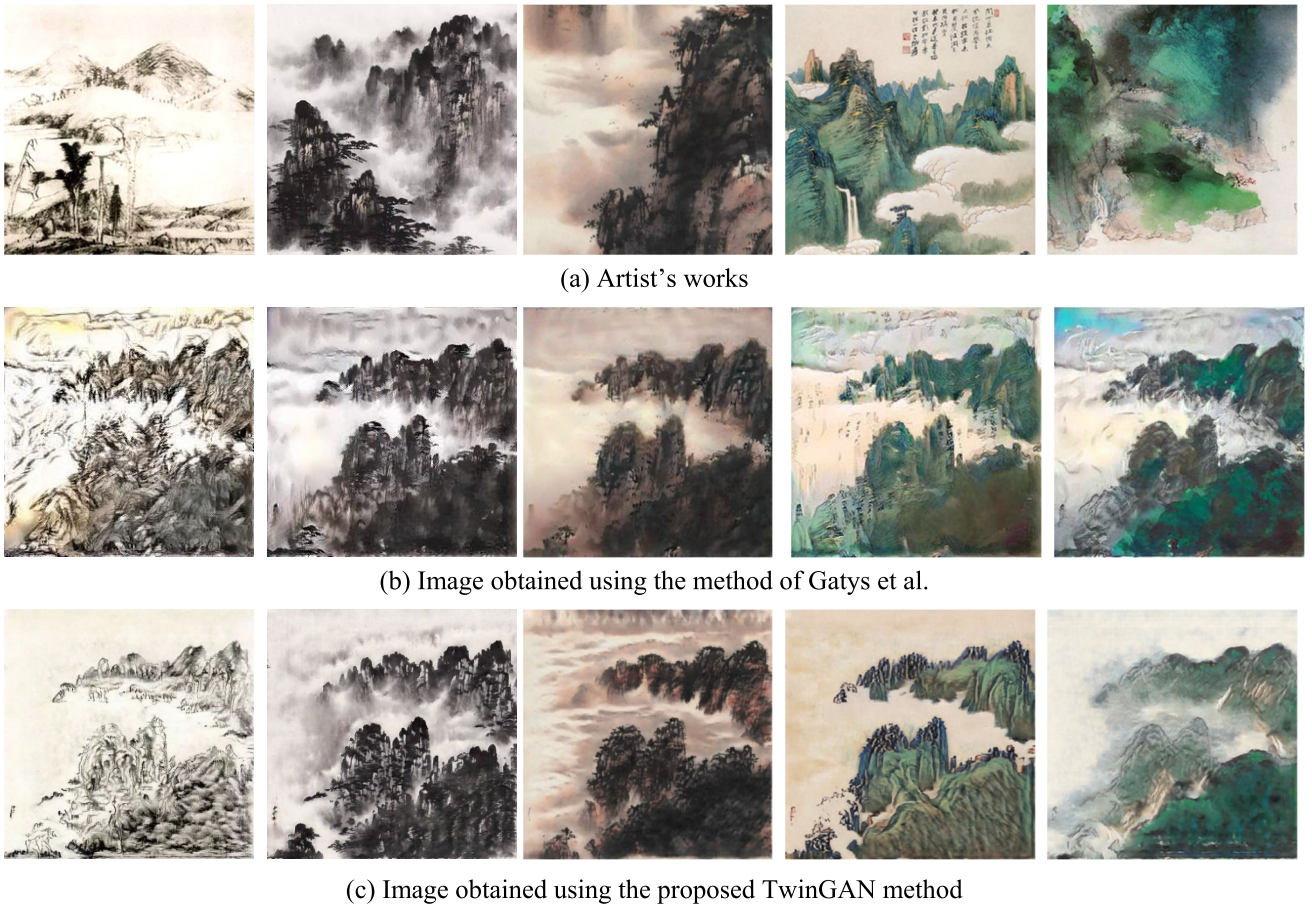


FIGURE 4. Architecture of the developed TwinGAN model.



(a) without reconstruction loss (b) without multiscale discriminator (c) only SketchGAN (d) with SketchGAN and RenderGAN

FIGURE 5. Ablation results: (a) image obtained without using the reconstruction loss. (b) Image obtained without using the multiscale discriminator. (c) Image obtained only using SketchGAN. (d) Image obtained with SketchGAN and RenderGAN.



(a) Artist's works

(b) Image obtained using the method of Gatys et al.

(c) Image obtained using the proposed TwinGAN method

FIGURE 6. Comparison of images obtained using the method of Gatys et al and those obtained using the proposed algorithm.

2) EncoderB-CONSTANCY LOSS

The encoder is used to extract image features and transform an image from pixel space into latent space. The latent code of the input image x is altered to the latent code of the stylized image $G(x)$ by residual blocks. By calculating the feature loss of the encoder between the input image x and stylized image $G(x)$, the encoder can be trained to extract features according to style characteristics. The encoder constancy loss is expressed as follows:

$$\mathcal{L}_{const} = \mathbb{E}_{x\sim P_x}\left[\|f(x) - f(G(x))\|_2^2\right] \tag{5}$$

where $f(x)$ is a feature of an input image and $f(G(x))$ is a feature of a stylized image.

3) IDENTITY LOSS

Our goal is to enable a generative network to approximate a mapping function that can map one domain to another. To strengthen the generative network, the least absolute deviations were used to enable the generative network to perform identity mapping. Therefore, the output of the generative model must approximate the input image when it is sampled from the target domain. The identity loss can be expressed as follows.

$$\mathcal{L}_{identity} = E_{y\backsim P_y}\left[\|y - G(y)\|_1\right] \tag{6}$$

where y is sampled from the target domain P_y and $G(y)$ represents the data generated by the generative model.

4) RECONSTRUCTION LOSS

To ensure that the content of the stylized image $G(x)$ is similar to that of the input image x, a constraint is applied between the input and stylized images. The simplest constraint that can be used is reconstruction loss, which is expressed as follows.

$$\mathcal{L}_{rec} = \mathbb{E}_{x\backsim P_x}\left[\|x - G(x)\|_1\right] \tag{7}$$

The reconstruction loss has been successfully used in autoencoders, which generates an output that approximates the input. However, direct application of the reconstruction loss function is unsuitable for the present study. This loss function is used to calculate the pixel space. However, the color and shape of the stylized image are usually different from those of the input image. Therefore, the calculation of the reconstruction loss in the pixel space results in unstable training. To solve this problem, some modifications were made to the reconstruction loss function. Considering the

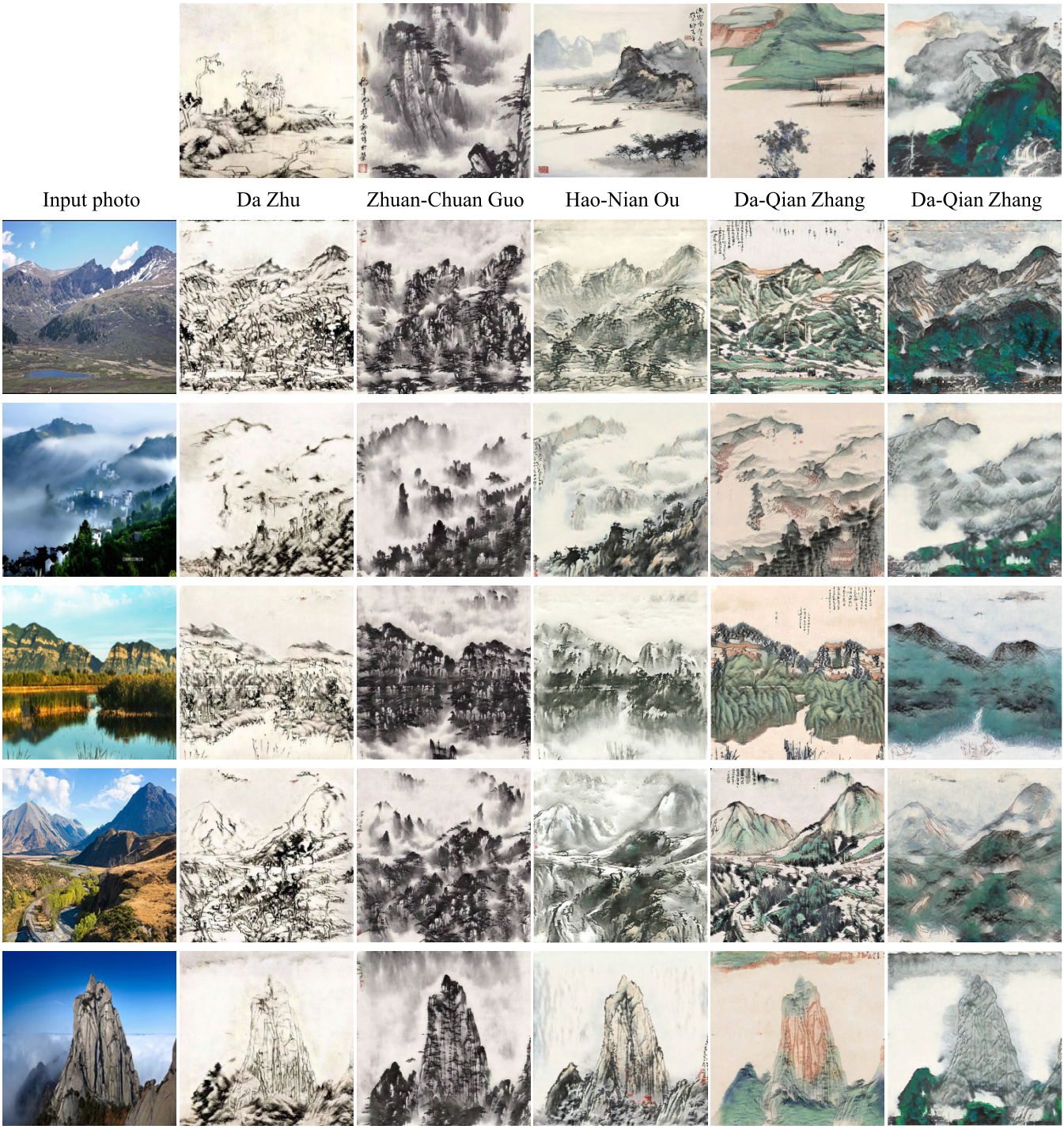


FIGURE 7. Experimental results: Five styles of Chinese landscape paintings.

characteristics of Chinese landscape paintings, the structure of the image content is an important image feature. In this study, the XDoG operation is used to extract the structures of the input and stylized images. Subsequently, the least absolute deviations of the two images were used to make their structures similar. The XDoG reconstruction loss is expressed as follows:

$$\mathcal{L}_{XDoG\_rec} = \mathbb{E}_{x \backsim P_x}\left[\|XDoG(x) - XDoG(G(x))\|_1\right] \quad (8)$$

IV. IMPLEMENTATION AND EXPERIMENTAL RESULTS

We used Pytorch and CUDA 9.0 to implement the proposed algorithm on an Intel Core i7B-4770k CPU@3.50 GHz with an Nvidia GeForce GTX 1080 graphics card. The tradeB-off parameters in Equation (3) are different at each stage of the proposed TwinGAN method. For SketchGAN, the tradeB-off parameters were set as $\lambda_{adv} = 1$, $\lambda_{const} = 5$, $\lambda_{identity} = 5$, and $\lambda_{rec} = 5$. For RenderGAN, the tradeB-off parameters were set as $\lambda_{adv} = 1$, $\lambda_{const} = 100$, $\lambda_{identity} = 1$, and $\lambda_{rec} = 1$. The

Adam optimizer [29] was used to train the neural network, and the learning rate was set to 0.0002. In the SketchGAN, training was conducted for 15 epochs. In RenderGAN, the training was conducted for 200 epochs. We maintained the same learning rate during the first 100 epochs and linearly decreased it during the last 100 epochs. The developed model required approximately 0.21 and 0.66 s to generate a 256 × 256 pixel image and 512 × 512 pixel image, respectively. The total time required to convert the 256 × 256-pixel image and 512 × 512-pixel image into the final stylized image was approximately 0.42 and 1.32 second, respectively.

Because we did not find open-source datasets of Chinese landscape ink paintings, we manually collected relevant images from different websites. The collected Chinese landscape ink paintings depicted natural objects such as mountains, forests, and rivers. For the dataset of stylized images, we collected the artworks of four Chinese landscape ink painting artists, Hao-Nian Ou, Zhuan-Chuan Guo, Da Zhu, and Da-Qian Zhang. We created five style datasets from these artworks: line drawing, ink painting, color ink, blue-green painting and splash ink. To enable easier model training, the collected images were cropped and resized to 256 × 256 pixels. After cropping and resizing, the content image dataset comprised 3502 photos, and the stylized image dataset comprised 686 paintings. The number of images belonging to the five style categories was 200, 132, 57, 178, and 119, respectively.

A. EXPERIMENT RESULTS

Experiments were performed on five styles of Chinese landscape ink paintings produced by four famous Chinese painters. The five styles of paintings considered in this study were as follows: the line drawing of the Ming dynasty painter Da Zhu, the Lingnan school style of Hao-Nian Ou, the modern style of Zhuan-Chuan Guo, and the blue-green landscape and splash ink styles of Da-Qian Zhang. The resolution of all collected paintings was 512 × 512 pixels. Figure 1 and Figure 7 display the five styles of experimental results for the Chinese landscape paintings.

B. ABLATION STUDIES AND COMPARISON

The ablation of single units in the loss function affects network accuracy in different ways. In general, ablation caused a decrease in the overall accuracy but exerted different effects on the accuracies for different classes. Figure 5 depicts the effect of the proposed loss term. The shape of the stylized image could not be retained without the application of XDoG reconstruction loss, as shown in Figure 5(a). When the XDoG reconstruction loss is used, the stylized image contains more detailed strokes, which can result in a clear depiction of details such as rock textures. Figure 5(b) shows the results obtained when a multiscale discriminator was not used. When a multiscale discriminator was used to train the developed model, the obtained results contained richer brush strokes. Figure 5(c) shows the results obtained after SketchGAN, in which the input photo is directly converted into a stylized image. Figure 5(d) shows the images obtained after SketchGAN and RenderGAN. This image shows the rock texture in Figure 5(d). Furthermore, artifacts can be reduced through SketchGAN.

We compared our method with that proposed by Gatys et al. [1]. The method proposed by Gatys et al. involves arbitrary style transfer; that is, a style image is required in their method for achieving style transfer. Therefore, we selected a stylized image from each style dataset for comparison, as shown in Figure 6(a). Figure 6 shows the results of the comparison. The images in Figure 6(b) obtained using the method of Gatys et al. contained some artifacts. These images contained unnatural chaotic lines and an uneven color distribution. The images in Figure 6(c) obtained using the proposed algorithm contain more detailed brushstrokes than those obtained using the method described by Gatys et al.

V. CONCLUSION

This paper presents TwinGAN for the synthesis of Chinese landscape paintings. Based on the characteristics of Chinese landscape ink paintings, the proposed algorithm comprises two GAN to simplify style transfer. In sketchGAN, the main structure in the input photo and the locations of uneven color distribution were identified. An uneven color distribution results in artifacts that are reduced by masking the noise in the sketching stage. In renderGAN, XDoG reconstruction loss is used to ensure that the shapes of the stylized and content images are consistent. In this study, the proposed algorithm successfully simulated five styles of Chinese landscape ink paintings. In the future, the proposed algorithm could be applied to other types of Chinese painting styles, such as meticulous paintings. The proposed algorithm causes the content image to approximate the stylized image and does not consider the semantic meaning of these images. Future studies can perform style transfers based on semantic meanings.

REFERENCES

[1] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[3] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.

[4] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.

[5] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807.

[6] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1501–1510.

[7] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 1857–1865.

[8] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 694–711.
[9] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images," in *Proc. ICML*, vol. 1, 2016, p. 4.
[10] V. Dumoulin, J. Shlens, and M. Kudlur, "A learned representation for artistic style," 2016, *arXiv:1610.07629*.
[11] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua, "StyleBank: An explicit representation for neural image style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1897–1906.
[12] T. Qi Chen and M. Schmidt, "Fast patch-based style transfer of arbitrary style," 2016, *arXiv:1612.04337*.
[13] X. Huang, M. Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 172–189.
[14] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M. H. Yang, "Universal style transfers via feature transforms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 386–396.
[15] D.-L. Way, R.-J. Chang, C.-C. Chang, and Z.-C. Shih, "A video painterly stylization using semantic segmentation," *J. Chin. Inst. Eng.*, vol. 45, no. 4, pp. 357–367, May 2022.
[16] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6924–6932.
[17] X. C. Liu, M. M. Cheng, Y. K. Lai, and P. L. Rosin, "Depth-aware neural style transfer," in *Proc. Symp. Non-Photorealistic Animation Rendering*, 2017, pp. 1–10.
[18] Y. Jing, Y. Liu, Y. Yang, Z. Feng, Y. Yu, D. Tao, and M. Song, "Stroke controllable fast style transfer with adaptive receptive fields," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 238–254.
[19] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4990–4998.
[20] H. Chen, L. Zhao, Z. Wang, H. Zhang, Z. Zuo, A. Li, W. Xing, and D. Lu, "DualAST: Dual style-learning networks for artistic style transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 872–881.
[21] L. Yang, L. Yang, M. Zhao, and Y. Zheng, "Controlling stroke size in fast style transfer with recurrent convolutional neural network," *Comput. Graph. Forum*, vol. 37, no. 7, pp. 97–107, Oct. 2018.
[22] A. Xue, "End-to-end Chinese landscape painting creation using generative adversarial networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3862–3870.
[23] X. Lv and X. Zhang, "Generating Chinese classical landscape paintings based on cycle-consistent adversarial networks," in *Proc. 6th Int. Conf. Syst. Informat. (ICSAI)*, Nov. 2019, pp. 1265–1269.
[24] Y. Wang, W. Zhang, and P. Chen, "Chinastyle: A mask-aware generative adversarial network for Chinese traditional image translation," in *Proc. SIGGRAPH Asia Tech. Briefs*, 2019, pp. 5–8.
[25] X. Yang and J. Hu, "Deep neural networks for Chinese traditional landscape painting creation," in *Proc. 2nd Int. Conf. Artif. Intell., Automat., High-Perform. Comput. (AIAHPC)*, vol. 12348, 2022, p. 9, doi: 10.1117/12.2641585.
[26] H. Winnemöller, J. E. Kyprianidis, and S. C. Olsen, "XDoG: An extended difference-of-Gaussians compendium including advanced image stylization," *Comput. Graph.*, vol. 36, no. 6, pp. 740–753, 2012.
[27] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1395–1403.
[28] Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image generation," 2016, *arXiv:1611.02200*.
[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

DER-LOR WAY received the Ph.D. degree in computer science from National Chiao Tung University, in 2005. From 2016 to 2022, he was the Dean of the School of Film and New Media. He was also with the Industrial Technology Research Institute (ITRI) to research multimedia and virtual reality for 11 years. He is currently a Professor with the Department of New Media Art, Taipei National University of Arts, Taiwan. His research interests include non-photorealistic rendering, digital art, and virtual reality.

CHANG-HAO LO was born in Hsinchu, Taiwan, in 2000. He received the B.S. degree from the Department of Computer Science and Information Engineering, National Chung Cheng University, the M.S. degree from the Institute of Multimedia Engineering, National Yang Ming Chiao Tung University, Hsinchu, in 2020, and the B.S. degree from the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, in 2023. He is currently an Engineer with MediaTek Inc. His current research interests include computer graphics, computer vision, image processing, and deep learning.

YU-HSIEN WEI was born in Hsinchu, Taiwan, in 2000. He received the B.S. degree from the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, in 2023. His research interests include computer vision, image processing, and deep learning.

ZEN-CHUNG SHIH received the B.S. degree in computer science from Chung-Yuan Christian University, in 1980, and the M.S. and Ph.D. degrees in computer science from National Tsing Hua University, in 1982 and 1985, respectively. He is currently a Professor with the Department of Computer and Information Science, National Yang Ming Chiao Tung University, Hsinchu. His current research interests include procedural texture synthesis, non-photorealistic rendering, global illumination, and virtual reality.