Advances in Artificial Intelligence and Machine Learning; Research 3 (2) 1220-1233 Received 22-04-2023; Accepted 23-06-2023; Published 30-06-2023

InkGAN: Generative Adversarial Networks for Ink-And-Wash Style Transfer of Photographs

Keyi Yu† keyiy@andrew.cmu.edu
Google Inc, 1600 Amphitheatre Parkway
Mountain View, CA, USA 94043

Yu Wang† wangyu19@student.deanza.edu
De Anza College, 21250 Stevens Creek Blvd
Cupertino, CA, USA 95014

Sihan Zeng† sihanzen@andrew.cmu.edu
Meta Platforms, 1 Hacker Way
Melon Park, CA, USA 94025

Chen Liang† chenlia2@andrew.cmu.edu
Google Inc, 1600 Amphitheatre Parkway
Mountain View, CA, USA 94043

Xiaoyu Bai† xybai521@gmail.com
Meta Platforms, 1 Hacker Way
Melon Park, CA, USA 94025

Dachi Chen† chendachimail@gmail.com
Meta Platforms, 1 Hacker Way
Melon Park, CA, USA 94025

Wenping Wang wenpingw@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, PA, USA 15213

Corresponding Author: Wenping Wang

Abstract

In this work, we present a novel approach for Chinese Ink-and-Wash style transfer using a GAN structure. The proposed method incorporates a specially designed smooth loss tailored for this style transfer task, and an end-to-end framework that seamlessly integrates various components for efficient and effective image style transferring. To demonstrate the superiority of our approach, comparative results against other popular style transfer methods such as CycleGAN is presented. The experimentation showcased the notable improvements achieved with our proposed method in terms of preserving the intricate details and capturing

† The work is done when author was in CMU, and has no bearing on their current affiliation.

Citation: Keyi Yu, et al. InkGAN: Generative Adversarial Networks for Ink-And-Wash Style Transfer of Photographs. Advances in Artificial Intelligence and Machine Learning. 2023;3(2):71.

the essence of the Chinese Ink-and-Wash style. Furthermore, an ablation study is conducted to evaluate the effectiveness of each loss component in our framework. We conclude in the end and anticipate that our findings will inspire further advancements in this domain and foster new avenues for artistic expression in the digital realm.

Keywords: Generative adversarial networks, Photographs, Photo style transfer

1. INTRODUCTION

Chinese Ink-and-Wash graph is a Chinese traditional art genre. Different from other art works, Ink-and-Wash pictures depict the environment and objects in an abstract way with limit color. A typical Chinese Ink-and-Wash picture is shown in FIGURE 1a. As it is a valuable art format, it is meaningful to find some approaches to rendering such pictures via style transferring. However, the most state-of-the-art style transfer framework based on GAN, such as CycleGAN [1], CycleGan-vc2 [2], cannot produce a satisfaction results. That is due to the dramatically difference in the way Ink-and-Wash pictures and other stroke, color and dyeing style pictures depict the objects. As shown in 1b, the stroke in Chinese Ink-and-Wash picture is smoother than the other two arts format.

Therefore, this paper proposes a simple yet effective loss functions to cope with the substantial style variation, i.e. smooth loss [3, 4]. Besides, cropped images for data augmentation proves that even a small scale of style images can attribute to fairly satisfactory results. The effectiveness of our purposed method is further demonstrated through an ablation study. And finally, the end-to-end framework for image style transfer is built as a pipeline for future productionalization.

(a) Ink-and-Wash graph

(b) Sketch stroke(left), Ink-and-wash stroke(middle), painting stroke(right)

Figure 1: Chinese Typical Ink-and-Wash graph and Stroke Comparison

2. RELATED WORK

2.1 Style Transfer Problem

Rendering the semantic content of an image in different style is an interesting topic in these years. Arguably, before the neural network comes out, the major bottleneck for the rending is to represent the style features and separate picture information from style.

Traditional methods do the style transfer from different perspective. A typical approach is to apply a spatially-invariant transfer functions to an image. However, these function can only perform a simple style transfers, such as work proposed by [5]. Their work used a serials of 1D histogram to transfer the 3D color space. The method does well in some specific cases, but is limited by the ability of function and cannot match complicated style transfer. In a task that transfers to Chinese ink-and-wash style, the objects the environment need to be more abstractly represented, thus cannot be easily done by a simple transfer function.

2.2 Neural Network

To overcome the limitation caused mentioned above, neural networks can be adopted. Convolutional Neural Networks (CNNs) introduced by [6] are powerful tools to solve such dilemma. Unlike the traditional style transfer methods [7], which uses a pair of images to do the style transfer, the CNNs methods can extract the semantic features from a set of aimed style pictures. A variety of CNNs have been proposed to extract the semantic information, such as VGG network [8], ResNet [9], and Clipstyler [10]. As a result, more and more powerful style transfer methods has come out.

Gatys et al [11] is the first one to propose a neural-network based style transfer approach. He uses CNN to extract the semantic features. Pre-trained VGG network is used to extract the content information, and using the global Gram matrix to produce a nice results for transferring the texture. It shows a good performance with limitation that content and style texture information should be somehow similar. In addition, it may fail when multiple object contains in the picture and cannot provide a smooth shading that required by our wash-and-ink style.

2.3 Generative Adversarial Network (GAN)

Besides the CNNs, Generative Adversarial Network [12, 13] have become more popular in recent days due to its excellent performance in plenty of fields, such as text-image translation, visual question answering. The key idea of GAN is to train two networks, i.e. generator network and discriminator network, and uses the adversarial loss provided by the discriminator network to push the generative network to generate the target pictures.

Lots of GANs work have been published and shows a good performance on the picture generating. For human portrait images, StyleGAN [14, 15] focuses on the generation and style transfer of high-resolution human portrait images, shows a very promising performance with the help of a redesigned generator architecture where Gaussian noise is added after each convolution layer. Based on it, DualStyleGAN [16] does better on exemplar-based style transfer which requiring much fewer training data by adding an extra extrinsic style to control the style of the target domain.

Specifically, for wash-and-ink style transfer, ChipGAN[17] proposes a model that incorporate three characteristics of paintings of this style into the loss function that shows a good generation quality. Our approach, however, starts from a different perspective. The original GAN solutions[18, 19] use paired pictures to perform style transfer. CycleGAN [1] breaks this limitation by proposing a novel framework. It trains two set of GAN models at the same time, mapping to each other. However, it also leads to a low running efficiency. Furthermore, this results provide poor results for our wash-and-ink oriented style transfer. Therefore, a special loss dedicated for the style of our pictures is designed in this work.

2.4 Diffusion Models

Recently, diffusion models [20, 21] show impressive result on style transfer with a , but requires labeled dataset and extra text prompt input. InST [22] proposed an inversion-based attention model that relieves the need of complex textual description as input.

However, GAN still retains certain advantages over diffusion models on the requirement of computation resources and training times. GAN generated images are still tend to be more realistic due to its ability to capture intricate details.

3. InkGAN

3.1 Model Architecture

In this section, we will briefly introduce how our model is composed. The GAN model consists of two main CNN networks: Generator Network G and Discriminator Network D. In this work, the Generator Network G takes a real world photo and generate a wash-and-ink style photo and the Discriminator Network D takes the generated photo and judge whether the photo is a real wash-and-ink image. FIGURE 2 shows the architecture used for our Ink-GAN.

As shown in FIGURE 2a, G net uses two down-convolution blocks to spatially compress and extract useful local information. Then the local information goes through the residual blocks to further extract and manifold features. Finally the features are used in transpose-CNN network and reconstruct the ink-and-wash style pictures.

As for the discriminator network D shown in FIGURE 2b, a simple CNN network is employed to judge the image considering the simpleness of the task. This can help reduce the parameters and accelerate the training speed.

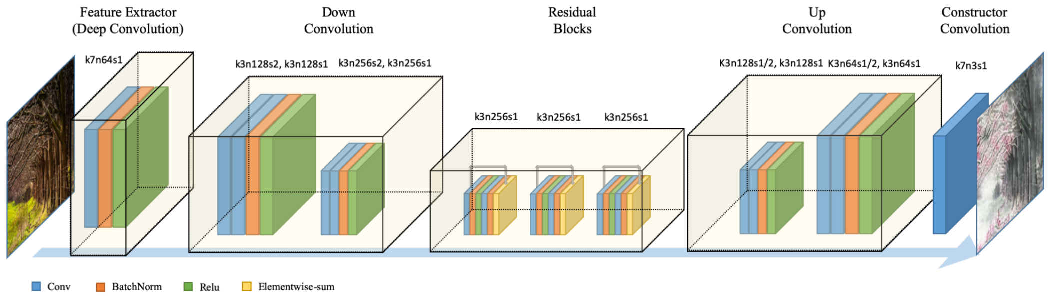
During model architecture tuning, we try to find the optimal network (G^*, D^*), which can meet our max-min loss requirement:

$$(G^*, D^*) = \arg\min_G \max_D L(G, D)$$

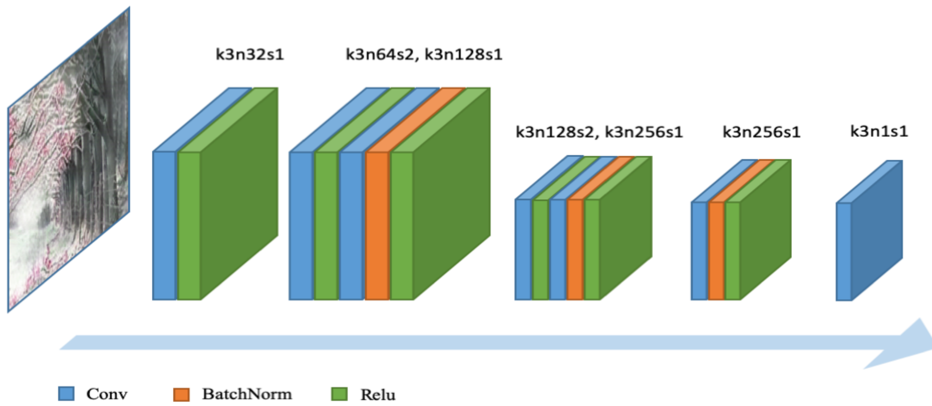
3.2 Loss Function

The total loss function $L(G, D)$ is a combination of two parts. On one hand, the output need to remain the same content for generated images, so a loss term is added to minimize the distance between original input and Generator's output, under the feature space. We mark it as $L_{con}(G, D)$. On the other hand, our Generator should have the ability to do style transfer, so the other part of loss is an adversarial loss $L_{adv}(G, D)$, which drives the generator network to achieve our desired transformation from real world scene to ink-and-wash painting. We use a simple additive form for the loss function:

$$L(G, D) = w * L_{con}(G, D) + L_{adv}(G, D)$$



(a) Generator Network



(b) Discriminator Network

Figure 2: InkGAN Architecture

where w is a parameter to balance two given losses. Larger w leads to higher penalty for content modification, while smaller w encourages more imaginative transformation. Experiment results showed that a dynamically changing fashion for w can accelerate converging as well as bring about a better performance. At the beginning, generator is almost random. A larger w push it to first learn an identity mapping. After several epochs, the generator has already been able to reproduce the content and gradually decrease w to encourage more style transfer. Concretely, w is set to be 10 at the beginning and changed every 5 epochs, finally converge as 2.

3.2.1 Content loss

In addition to transformation between different styles, one more important goal is to ensure that the image generated retain same semantic content of the input photos. VGG16 [8] is a convolutional neural network model which has been demonstrated to have good object preservation ability. In InkGAN, the content distance between original images and generated images in feature space is minimized. The last layer of VGG, whose dimension is 4096, is taken as image's feature. Accordingly, the content loss is designed as:

$$L_{con}(G, D) = E_{r_i \in S_r}[||Conv(r_i) - Conv(G(r_i))||]$$

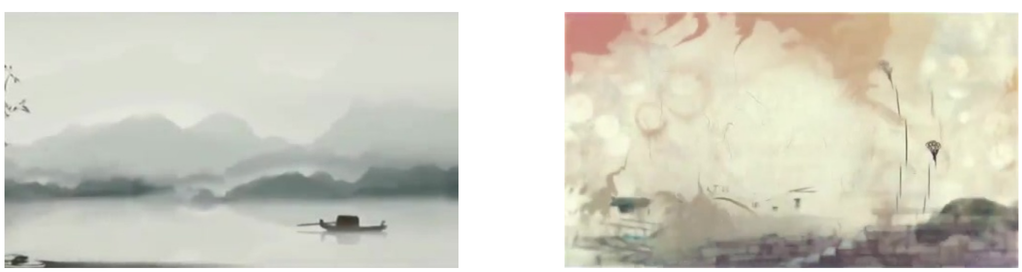
where S_r represent the scope of real world image and r_i is one of the items. The VGG model we used is pre-trained on ImageNet challenge dataset[23]. Regard the distance metrics, $L1$ rather than $L2$ is used for normalization. The main insight is style information is kind of similar to a local filter so it may be captured by original dimensions of feature map, leaving most dimension remain the same. $L1$ norm works well for reducing the value some dimensions to zero, which means force some dimensions of original image and generated image to be exactly same.

3.2.2 Adversarial loss

Similar to the normal Generative Adversarial Network, here the optimization is to solve a min-max problem. The generator take an real world scene as input and output its corresponding ink-and-wash version. The output image as well as a ground-truth ink-and-wash image(their contents are different), will be feed into the discriminator and then their probability or confidence of correct style is acquired. Therefore considering the cross-entropy loss, the target of generator is to generate images that can maximize the CE loss and the target of discriminator is to better identify real images from generated one, that is, to minimize the cross-entropy. So first two items are:

$$E_{s_i \in S_s}[logD(s_i)] + E_{r_i \in S_r}[log(1 - D(G(r_i)))]$$

where S_s is the scope of ground-truth style image and S_r is same as before. However, with this loss function, the results is not good enough. What the generator do is no more than adding some color filters or light filters to input images. Besides we observed that an important character of our target style is blurriness, like FIGURE 3, but the generated images do not hold this feature.



(a) Blurness Example 1 (b) Blurness Example 2

Figure 3: Example of ink-and-wash style images

To circumvent this problem, our idea is to force the discriminator to give prediction using blurness, so that the generator must learn to change the edge form of original image to confuse discriminator. Therefore what we actually do is to add another sharp-contour version of target style image as "negative" example, shown in FIGURE 4.

Those images are generated with sharpness convolution filter. Concretely, three kinds of convolution filter with size 3,5,7 are leveraged and chose uniformly. In this way, every time one positive

(a) Original ink-and-wash image

(b) Sharp-Contour image

Figure 4: By processing original image with sharpness filter, we get images with same content but shaper edge.

input and two negative inputs are feed for the discriminator, and the total adversarial loss is:

$$L_{adv}(G,D) = E_{s_i \in S_s}[logD(s_i)] + E_{r_i \in S_r}[log(1 - D(G(r_i)))] + E_{c_i \in S_c}[log(1 - D(c_i))]$$

where S_c represent the scope of sharp-contour images, we name the additional item as smooth loss.

3.2.3 Decomposition

In this subsection, the total loss function is refined for generator and discriminator(pre-trained VGG network is fixed during training). For generator G, only the second item of $L_{adv}(G,D)$ and $L_{con}(G,D)$ matters and G is supposed to minimize both of them. Thus,

$$L_G = E_{r_i \in S_r}[log(1 - D(G(r_i)))] + w * E_{r_i \in S_r}[||Conv(r_i) - Conv(G(r_i))||]$$

While for discriminator D, all items in adversarial loss matters and it will maximize that. So we add an additional negative sign in the formula:

$$L_D = -E_{s_i \in S_s}[logD(s_i)] - E_{r_i \in S_r}[log(1 - D(G(r_i)))] - E_{c_i \in S_c}[log(1 - D(c_i))]$$

Two loss will be updated alternatively in the training process.

3.3 Model Complexity Analysis

For the spatial complexity, a single convolution layer has about $O(k^2 n_i n_o)$ parameters, where k is the kernel size, n_i and n_o are the input and output channels. As shown in 2, the generator has a combined of around 6 million parameters; while the discriminator has around 1 million parameters, all from the convolution layers.

For the computational complexity, a forward-pass of a single convolution layer generally need $O(d_{ox} d_{oy} k^2 n_i n_o)$ computation operations, where d_{ox} and d_{oy} are the dimensions of the output feature image. The specific time needed depends on the size of the input image.

4. EXPERIMENTS AND RESULTS

In this section, we will introduce how experiments are conducted and show some comparison results. Our InkGAN is implement with Pytorch. Both proposed model and baseline model are trained on 1 Nvidia Tesla P100 GPU for 100 epochs. Training one epoch of InkGAN takes around 380 seconds. In following subsections, we first introduce the pipeline for data collection and then show some example images and compare with CycleGAN[1]. Finally, the training process is further analyzed together with an ablation study.

4.1 Data Collection

Our data set contains 3591 Chinese ink-and-wash images (Style Image Dataset) and 5428 real world scene images (Content Image Dataset). The content images were obtained as a subset of Instagram-styled photographs downloaded from Flickr[1]. To filter the images, we specifically used the keyword ”natural scenery” [2] to identify content that aligns with the characteristics of ink-and-wash images, such as mountains, trees, sea, and more. The primary focus of our research is to train the discriminator to make predictions based on style rather than content. Therefore, we carefully selected the images based on their suitability for ink-and-wash style analysis. All images in the dataset were uniformly reshaped, with the shorter side resized to 224 pixels. A few examples of the original images are presented in FIGURE 5. 500 pictures are randomly selected as test data, left the others as training data.

Figure 5: Reshaped content images.

The Style Images in our dataset were collected meticulously from various sources. Specifically, we crawled over one hundred ink-style videos from platforms such as YouTube, Bilibili, and Youku.

[1] https://www.flickr.com/
[2] https://www.flickr.com/search/?text=natural%20scenery

Screenshots were captured at regular intervals, typically one frame per second, to ensure a comprehensive representation of the ink-and-wash style. After the initial collection, we carefully examined each captured image to ensure its suitability for our dataset. To enhance the diversity of our dataset, we employed data augmentation techniques. In each epoch, every image was randomly cropped to create additional variations while maintaining a scalable image dataset. This augmentation process helps to capture different perspectives and aspects of the ink-and-wash style. Examples have already shown in FIGURE 3.

4.2 Comparison With Other Methods

In our evaluation, we compared generated images of our model with other state-of-the-art methods, especially CycleGan and CNN based style transfer methods [11], which are widely recognized approaches for learning image translation in the absence of paired examples. Identity loss version is used for fairness since the incorporation of this item tends to produce stylized images with better content preservation. Furthermore, all models were trained using the same dataset.

Qualitative results are presented in FIGURE 6, which clearly demonstrate that InkGAN performs better than the two baselines. What CycleGAN do is no more than adding a color or light filter on the original images. By contrast, our InkGAN has the ability to add some special element in ink-and-wash style like mist, peach blossom and painting stroke. Besides, it will convert colorful input images into basically white, pink and light green, which are most common colors for ink-and-wash painting. In addition, since InkGan employs smooth loss into training, images generated become more blur, more like a real painting. On the other hand, compared with CNN based methods, InkGAN demonstrates better preservation of fine details in the input images, such as the intricate details of flowers in the first and last examples, the texture of rocks in the second example, and the depiction of trees in the third image.

There are several other work in the literature we have not reproduced here. While we believe our result are superior than those methods in the specific ink-and-wash style transfer task. DistanceGan [24] further enforce the mapping relation between samples in two domain based on CycleGan. However, this constraint fails to capture the essence of ink-and-wash painting, such as its characteristic colors and mist. Also DistanceGan is only tested on single object images, such as horses or pair of shoes. Ghiasi et al. [25] proposed a method to allow real-time stylization using any content plus style image pair. While as a general solution applicable to various painting styles, it does not leverage target style information in the model training, thus falling short in surpassing our ink-and-wash style-specific approach. He et al. [17] focused on black and wight ink-wash painting style, and incorporated specific constraints like Brush stroke constrain in model's training. In comparison, InkGan did a better job in color rendering and original content preserving. Lastly, Luan et al. [26] proposes an approach for photo realistic style transfer by incorporating a locally affine transformation in color space. However, affine transformation will not capture the characteristic blur effect and atmosphere present in ink painting.

Figure 6: Comparison of InkGAN(ours, second column), CycleGAN (third column) and Gatys's CNN based transfer method (last column). It is clear to see InkGAN can produce more shape transform as well as better style rendering. In contrast, CycleGAN is more like adding a color filter; CNN based method lose content for some dedicated components like flowers and stone's texture.

4.3 Furthur Analysis

In this subsection, we delve deeper to see the actually effect of each component in our loss function. First we visualize how loss value changes as training goes, shown in FIGURE 7. We have three Epoch-Loss curve in blue, orange and green, representing value of content loss $L_{con}(G, D)$, discriminator loss L_D and normalized generator loss $L_G/(w+1)$ respectively.

We observed that content loss gradually increases during training, but finally will converge. That makes sense since style transfer will more or less change content or features of original input.

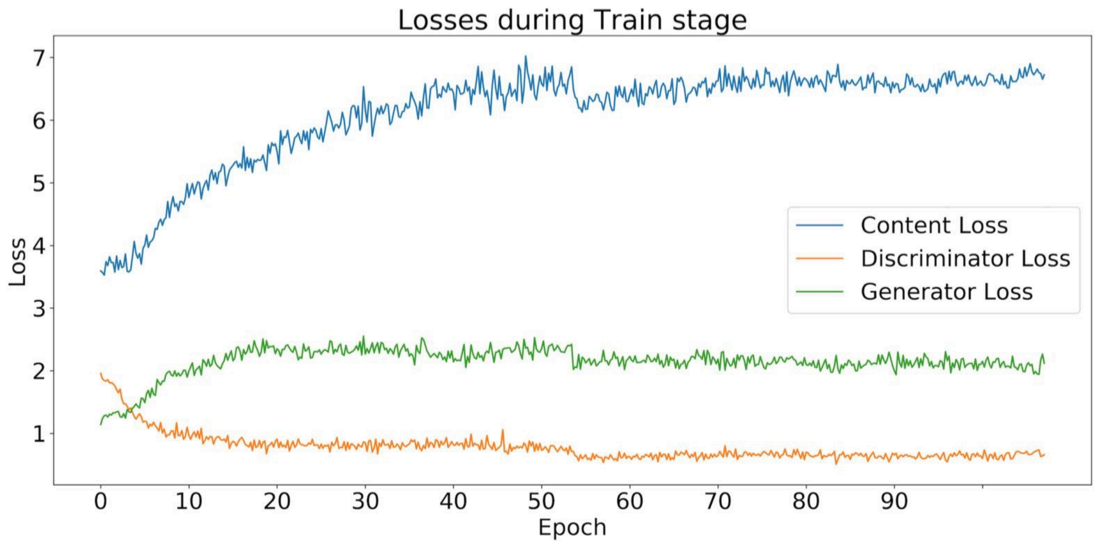


Figure 7: Curve of loss during training stage.

Thanks to adding content loss into training, value of this item does not explode. Discriminator loss continuously decreases to converge as training goes and decreasing speed grows slower and slower, which proves that discriminator is learning things together with generator. Generator loss is the most interesting one. It increases in first 20 epochs and then starts to decrease. One reason is the content loss part continues goes up. Besides, since parameter w is continuously changed, which actually changes the weight of adversarial part and content part in the normalized formula, make it hard to explain. While, we observed images generated at different training stage. Image generated does become better and better, which is the strongest proof that our generator trains well.

Following, we conducted an ablation study to see whether our smooth loss could improve performance. FIGURE 8 shows the comparison. Left column are original input images and the right and middle columns are images generated with and without smooth loss. It is easily to notice that with the help of smooth loss, style images become more blur, more illusory, and thus looks more similar to real ink-and-wash paintings.

5. CONCLUSION

In this paper we proposed InkGAN, a Generative Adversarial Network to transform real-world photos to high-quality ink-and-wash paintings. To realize that, we add two additional loss item on normal GAN's loss, one content loss to help regularize the difference between contents of input and output images, as well as a smooth loss to force images generated to be blur and illusory. The idea is not very complicated but our model actually works well, much better than CycleGAN and very fantastic int-and-wash style images could be generated. Besides, we create a Chinese painting data set with about 3.6K images from scratch. We also do an ablation study to verify the effectiveness of each loss component.

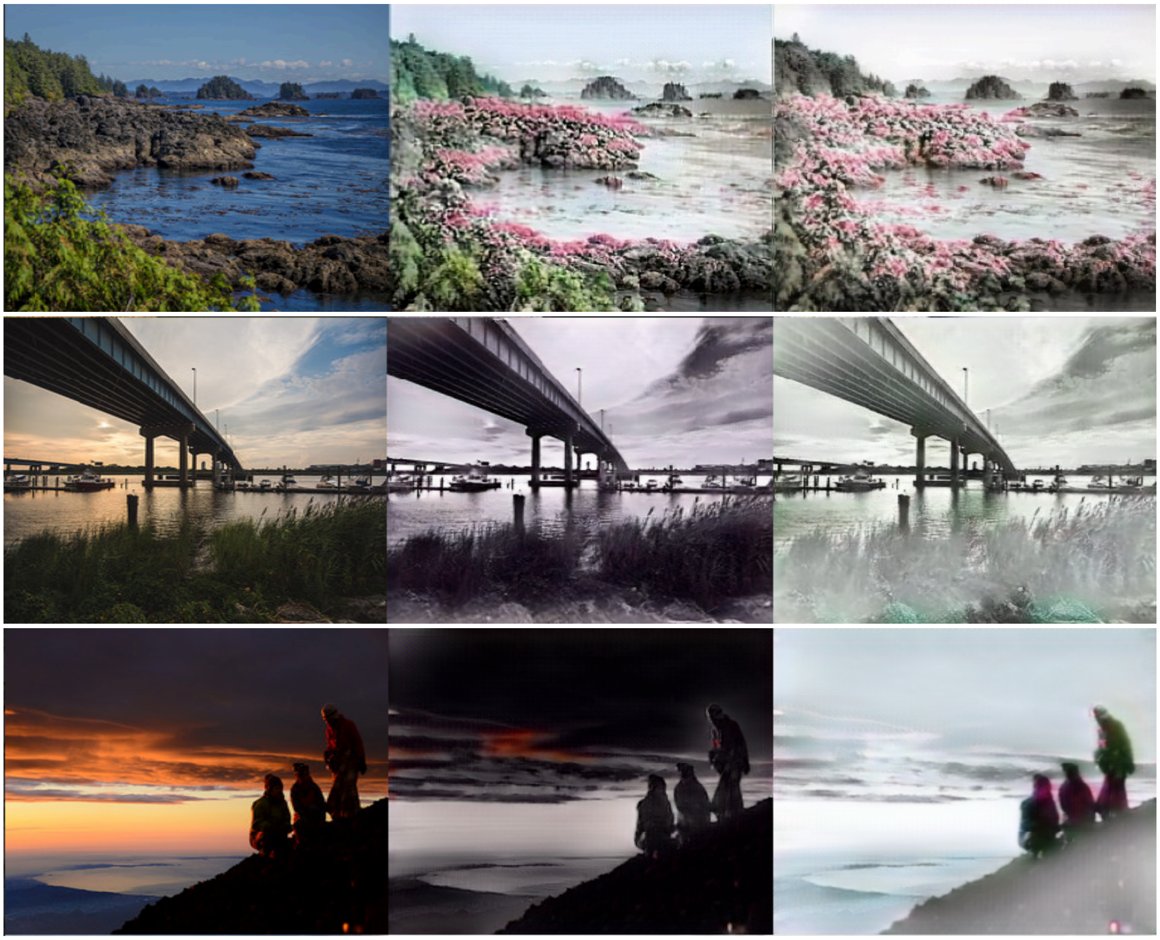


Figure 8: Comparison of images generated with and without smooth loss. Left column are original images, right columns are images generated with our model, and the middle line is generated without smooth loss. It shows that smooth loss help to make images more blur and illusory.

References

[1] Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. In: Proceedings of the IEEE international conference on computer vision. 2017:2242-2251.

[2] Kaneko T, Kameoka H, Tanaka K, Hojo N. Cyclegan-VC2: Improved Cyclegan-Based Nonparallel Voice Conversion. In: ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Vol. 2019. IEEE Publications. 2019:6820-6824.

[3] Akhtar MR, Liu P. Unsupervised generative adversarial network for style transfer using multiple discriminators. In Thirteenth International Conference on Graphics and Image Processing (ICGIP 2021) . SPIE. 2022;12083:178-186.

[4] Liao Y, Huang Y. Deep Learning-Based Application of Image Style Transfer. Math Probl Eng. 2022;2022:1-10.

[5] Li C, Wand M. Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:2479-2486.

[6] Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification With Deep Convolutional Neural Networks. Adv Neural Inf Process Syst. 2012:1097-1105.

[7] Huang SS, Zhang GX, Lai YK, Kopf J, Cohen-Or D, Hu S-M. Parametric Meta-Filter Modeling From a Single Example Pair. Vis Comput. 2014;30:673-684.

[8] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2015. ArXiv preprint: https://arxiv.org/pdf/1409.1556.pdf

[9] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:770-778.

[10] Kwon G, Clipstyler JCY. Image Style Transfer With a Single Text Condition. 2021. ArXiv preprint: https://arxiv.org/pdf/2112.00374.pdf

[11] Gatys LA, Ecker AS, Bethge M. Image Style Transfer Using Convolutional Neural Networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:2414-2423.

[12] Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, et.al. Generative Adversarial Text to Image Synthesis. 2016. ArXiv preprint: https://arxiv.org/pdf/1605.05396.pdf

[13] Kim YuH, Nam SH, Hong SB, Park KR. Gra-Gan: Generative Adversarial Network for Image Style Transfer of Gender, Race, and Age. Expert Syst Appl. 2022;198:116792.

[14] Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, et.al. Analyzing and Improving the Image Quality of Stylegan. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020:8110-8119.

[15] Karras T, Laine S, Aila T. A Style-Based Generator Architecture for Generative Adversarial Networks. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019:4401-4410.

[16] Yang S, Jiang L, Liu Z, Loy CC. Pastiche Master: Exemplar-Based High-Resolution Portrait Style Transfer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022:7683-7692.

[17] He B, Gao F, Ma D, Shi B, Duan LY. Chipgan: A Generative Adversarial Network for Chinese Ink Wash Painting Style Transfer. In: Proceedings of the 26th ACM international conference on multimedia. 2018:1172-1180.

[18] Isola P, Zhu JY, Zhou T, Efros AA. Image-To-Image Translation With Conditional Adversarial Networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:5967-5976.

[19] Karacan L, Akata Z, Erdem A, Erdem E. Learning to Generate Images of Outdoor Scenes From Attributes and Semantic Layouts. 2016. arXiv preprint: https://arxiv.org/pdf/1612.00215.pdf

[20] Dhariwal P, Nichol A. Diffusion Models Beat Gans on Image Synthesis. Adv Neural Inf Process Syst. 2021;34:8780-8794.

[21] Kawar B, Zada S, Lang O, Tov O, Chang H, et al. Imagic: Text-Based Real Image Editing With Diffusion Models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023:6007-6017.

[22] Zhang Y, Huang N, Tang F, Huang H, Ma C, et al. Inversion-Based Style Transfer With Diffusion Models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023:10146-10156.

[23] Deng J, Dong W, Socher R, Li L, Li K, et al . Imagenet: A Large-Scale Hierarchical Image Database. In: IEEE conference on computer vision and pattern recognition. IEEE Publications. 2009:248-255.

[24] Benaim S, Wolf L. One-Sided Unsupervised Domain Mapping. Adv Neural Inf Process Syst. 2017;30.

[25] Ghiasi G, Lee H, Kudlur M, Dumoulin V, Shlens J. Exploring the Structure of a Real-Time, Arbitrary Neural Artistic Stylization Network. 2017. ArXiv preprint: https://arxiv.org/pdf/1705.06830.pdf

[26] Luan F, Paris S, Shechtman E, Bala K. Deep Photo Style Transfer. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:6997-7005.