Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Special Track on AI, the Arts and Creativity

Style Fader Generative Adversarial Networks for Style Degree Controllable Artistic Style Transfer

Zhiwen Zuo , **Lei Zhao*** , **Shuobin Lian** , **Haibo Chen** , **Zhizhong Wang** , **Ailin Li** , **Wei Xing*** and **Dongming Lu**

College of Computer Science and Technology, Zhejiang University

{zzwcs, cszhl, lshuobin, cshbchen, endywon, liailin, wxing, ldm}@zju.edu.cn

Abstract

Artistic style transfer is the task of synthesizing content images with learned artistic styles. Recent studies have shown the potential of Generative Adversarial Networks (GANs) for producing artistically rich stylizations. Despite the promising results, they usually fail to control the generated images' style degree, which is inflexible and limits their applicability for practical use. To address the issue, in this paper, we propose a novel method that for the first time allows adjusting the style degree for existing GAN-based artistic style transfer frameworks *in real time after training*. Our method introduces two novel modules into existing GAN-based artistic style transfer frameworks: a Style Scaling Injection (SSI) module and a Style Degree Interpretation (SDI) module. The SSI module accepts the value of Style Degree Factor (SDF) as the input and outputs parameters that scale the feature activations in existing models, offering control signals to alter the style degrees of the stylizations. And the SDI module interprets the output probabilities of a multi-scale content-style binary classifier as the style degrees, providing a mechanism to parameterize the style degree of the stylizations. Moreover, we show that after training our method can enable existing GAN-based frameworks to produce over-stylizations. The proposed method can facilitate many existing GAN-based artistic style transfer frameworks with marginal extra training overheads and modifications. Extensive qualitative evaluations on two typical GAN-based style transfer models demonstrate the effectiveness of the proposed method for gaining style degree control for them.

1 Introduction

Since the seminal work of Gatys *et al.* [2016], artistic style transfer has seen a booming development in recent years due to its scientific and artistic values. However, most existing artistic style transfer methods heavily depend on VGG network [Simonyan and Zisserman, 2014], pre-trained on ImageNet [Deng *et al.*, 2009], which requires extensive labeled images and can introduce an extra bias [Geirhos *et al.*, 2018], since the pre-trained network has no access to artistic images during training. In contrast, the methods based on Generative Adversarial Network (GAN) [Goodfellow *et al.*, 2014] directly leverage the collection of artistic images to learn the style representation, such as the works of [Li and Wand, 2016b; Elgammal *et al.*, 2017; Zhu *et al.*, 2017; Sanakoyeu *et al.*, 2018; Huang *et al.*, 2018; Lee *et al.*, 2018; Kotovenko *et al.*, 2019a; Svoboda *et al.*, 2020; Chen *et al.*, 2020; Kotovenko *et al.*, 2019b]. In essence, these methods follow the paradigm of image-to-image translation (I2I) [Isola *et al.*, 2017], where a translated image not only preserves its original content but also is stylized. Thanks to the rapid development of GANs, GAN-based artistic style transfer methods have shown great success in producing visually appealing stylizations.

Despite the promising results, existing GAN-based artistic style transfer methods can not flexibly control the style degree of the generated images. Artistic style transfer, however, is a very subjective task. A thousand people may have a thousand preferences for the stylizations. Therefore, a practical method allowing users to control the style degree of the stylizations for existing GAN-based artistic style transfer frameworks *in real time after training* is of great value. Nonetheless, real-time style degree control for existing GAN-based artistic style transfer models is a non-trivial problem. The challenges mainly lie in two aspects. ***On the one hand***, unlike those methods based on well-defined style losses (*e.g.*, the Gram loss [Gatys *et al.*, 2016] or the first and second order moment matching losses [Huang and Belongie, 2017]), which can explicitly control the strengths of the stylizations by adjusting the weights of the style losses [Gatys *et al.*, 2016; Babaeizadeh and Ghiasi, 2018] or mixing content and style features in the latent space [Huang and Belongie, 2017; Park and Lee, 2019], GAN-based artistic style transfer methods lack a mechanism to parameterize the style degree of the stylizations. Since the stylization is the result of adversarial training in GAN-based models, a possible solution is to adjust the weight of adversarial loss in existing GAN-based frameworks, hoping to produce different stylizations with different style degrees accordingly. But such solution has two fatal defects: **1)** as the adversarial training encourages the generated

*Corresponding authors

images' stylization effects to be indistinguishable from those of the artistic images, adjusting the weight of adversarial loss may not result in style degree changes of the stylizations, which will also be verified by our experiments in Section 4.2; **2)** moreover, even if such solution is feasible, one can only observe the effect of different weights by fully retraining the model with the new set of weights, which is cumbersome and does not fit for real-time applications. ***On the other hand***, from an I2I translation perspective, the training image sets with different style degrees are not available, which hinders direct modeling of the translations between the corresponding domains via conditional GANs (cGANs) [Mirza and Osindero, 2014]. Thus, it is not straightforward to enpower existing data-driven GAN-based artistic style transfer models with abilities to learn stylizations with different style degrees.

Solving the aforementioned problem and learning a style degree-aware representation for existing GAN-based artistic style transfer methods can not only provide more stylized image choices for users to select their favored ones but also are of great significance to promote the understanding and development for style transfer. To this end, we propose a simple yet effective method enabling flexible adjustment of the generated images' style degree *after training* and *in real time* for existing GAN-based artistic style transfer models. Our method mainly introduces two novel modules into existing GAN-based artistic style transfer frameworks: a Style Scaling Injection (SSI) module and a Style Degree Interpretation (SDI) module. The SSI module accepts the value of Style Degree Factor (SDF) as the input and outputs parameters that scale the feature activations in existing models, which offers control signals to alter the style degree of the stylization. And the SDI module interprets the output probabilities of a multi-scale content-style binary classifier as style degrees, providing a mechanism to parameterize the style degree of the stylizations. Moreover, we show that our method can enable existing GAN-based artistic style transfer models to produce over-stylizations by setting the value of SDF beyond 1. We conduct comprehensive experiments on two typical GAN-based artistic style transfer models to verify our method's effectiveness of gaining style degree controllability and learning a style degree-aware representation for them.

In summary, the contributions in this paper are fourfold:

- We propose a novel method that achieves style degree controllability for existing GAN-based artistic style transfer models *in real time after training* by introducing the Style Scaling Injection (SSI) and the Style Degree Interpretation (SDI) module into them.
- The proposed method can be easily integrated into many existing GAN-based artistic style transfer models with marginal extra training overheads and modifications.
- Moreover, we show that our method can enable existing GAN-based artistic style transfer models to produce over-stylizations by setting the value of Style Degree Factor (SDF) beyond 1.
- Comprehensive qualitative evaluations on two typical GAN-based artistic style transfer frameworks show that our method can help them freely alter the style degree of the generated images.

2 Related Work

According to how existing artistic style transfer methods extract and model the style information, we roughly divide them into two categories in this paper: (1) methods based on pre-trained networks (*e.g.*, VGG), and (2) methods based on generative adversarial networks (GANs).

2.1 Methods Based on Pre-Trained Networks

The pioneering work of Gatys *et al.* [2016] first demonstrated the strength of Deep Convolutional Neural Networks (DCNNs) in artistic style transfer, where the content and style can be expressed as multi-level feature statistics extracted from the pre-trained DCNNs. Since then, extensive works have been proposed to improve the performance of artistic style transfer in several aspects, such as efficiency [Johnson *et al.*, 2016; Ulyanov *et al.*, 2016], quality [Li and Wand, 2016a; Ulyanov *et al.*, 2017; Zhang *et al.*, 2019; Chen *et al.*, 2021b; Chen *et al.*, 2021a], generalization [Chen and Schmidt, 2016; Huang and Belongie, 2017; Li *et al.*, 2017b; Sheng *et al.*, 2018; Li *et al.*, 2019; Wang *et al.*, 2021], diversity [Li *et al.*, 2017a; Ulyanov *et al.*, 2017; Wang *et al.*, 2020; Chen *et al.*, 2021c], and controllability [Babaeizadeh and Ghiasi, 2018; Yao *et al.*, 2019].

Controllability. Many feed-forward style transfer methods based on pre-trained network can make trade-offs between the content and the style by adjusting the weights of the corresponding losses, but they have to retrain their models with a new set of weights for each trade-off, which is time-consuming and inefficient. Adjustable style transfer [Babaeizadeh and Ghiasi, 2018] was proposed for adjusting these weights after training and in real-time. Some arbitrary style transfer methods like AdaIN [Huang and Belongie, 2017], WCT [Li *et al.*, 2017b], and SANet [Park and Lee, 2019] also allow content-style trade-off at test time by interpolating between content and style features that are fed to the decoder. Yao *et al.* [2019] developed a method that can control the stroke patterns of the stylizations. However, *all the above methods extract and model the style representations based on the pre-trained network so that they can not be applied to GAN-based style transfer models to gain style degree controllability*.

2.2 Methods Based on GANs

There is another line of research that is free from pre-trained DCNNs, and it relies on GAN [Goodfellow *et al.*, 2014] to mimic the artistic image distribution. We informally regard the methods that only utilize adversarial training in pixel space to learn stylization effects as GAN-based style transfer methods. These methods, in general, learn a translation from a content image domain to a style image domain through a feed-forward network. And we further divide them into two types: methods with and without style image guidance in the translation.

Methods without style image guidance. Some methods utilized a collection of artworks instead of only one style image to learn the stylizations. CycleGAN [Zhu *et al.*, 2017] proposed a cycle-consistency loss for unpaired I2I, which has also been adopted in style transfer to yield visually pleasing

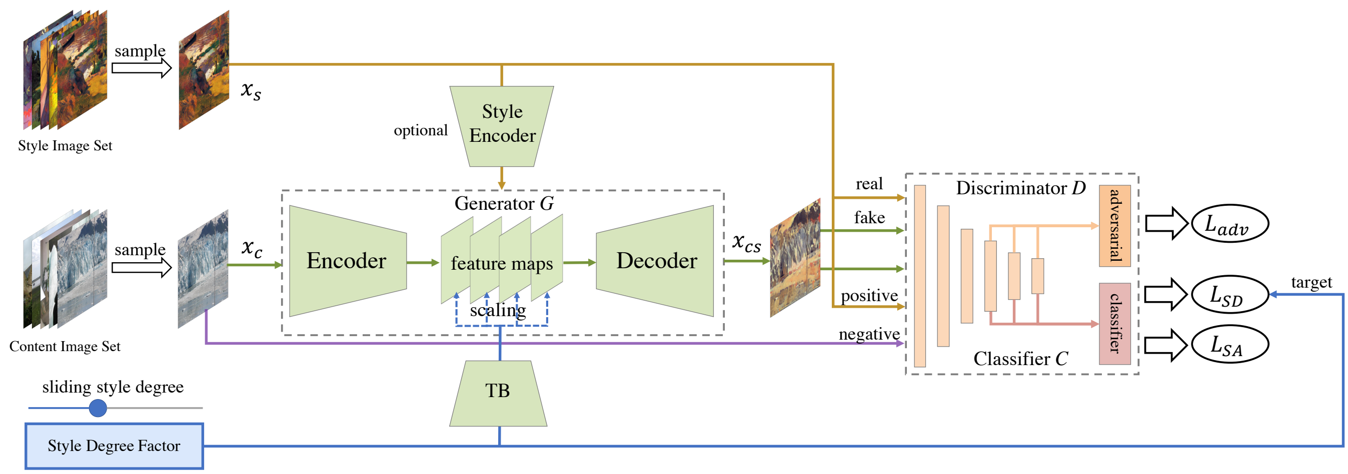

Figure 1: Illustration of the proposed method. We introduce two novel modules: the Style Scaling Injection (SSI) and the Style Degree Interpretation (SDI) module into existing GAN-based artistic style transfer models to control the style degree of the generated images. Refer to text for more details.

stylizations. AST [Sanakoyeu *et al.*, 2018] proposed a style-aware content loss, which can produce stunning stylized images. Kotovenko *et al.* [2019b] introduced a novel content transformation block that achieved content- and style-aware stylization of a content image. Chen *et al.* [2020] extended AST to multiple artists' styles.

Methods with style image guidance. Sharing the same high-level idea, MUNIT [Huang *et al.*, 2018] and DRIT [Lee *et al.*, 2018] proposed to decompose the latent space of the image into a shared content and a domain-specific style space, which can perform reference-guided style transfer. Recently, some methods [Kotovenko *et al.*, 2019a; Svoboda *et al.*, 2020] also proposed to disentangle the content and the style representation of the image by metric learning losses and therefore also allow for reference-guided style transfer.

Although the above methods have achieved superior stylization quality, they suffer from the incapability to freely alter the style degree of the generated images. Aside from these methods, DNI [Wang *et al.*, 2019] proposed to interpolate the model parameters between a reconstruction and an adversarial network for continuous domain translation. But DNI often needs to train two networks for the interpolation, which is inflexible. TM [Yu *et al.*, 2019] proposed a model that interpolates between texture features for smooth texture transition. However, TM was only designed for texture synthesis and used a weighted Gram loss to learn the transition, impeding its generalization to many purely GAN-based style transfer models. To the best of our knowledge, we propose the first method to empower existing GAN-based style transfer models with flexible style degree adjustment ability.

3 Proposed Method

3.1 Preliminaries

Existing GAN-based artistic style transfer frameworks often have an autoencoder-like generator G which accepts a content image x_c or a pair of content and style images (x_c, x_s) as the input (depending on whether they use a reference style image to guide the transfer) and translate the content image into a stylized image $x_{cs} = G(x_c)$ or $G(x_c, x_s)$ accordingly. The generated image x_{cs} needs to maintain the content of x_c and is stylized by the adversarial loss shown below:

$$\max_D \min_G \mathbb{E}_{x_s}[\log D(x_s)] + \mathbb{E}_{x_{cs}}[\log(1 - D(x_{cs}))] \quad (1)$$

where $D(x)$ denotes the probability that an input image x comes from the artistic image distribution rather than the stylized image distribution. The discriminator is trained to distinguish the artistic images from the generated stylized images, while the generator seeks to produce stylized images whose stylization effects can fool the discriminator.

3.2 Style Fader Generative Adversarial Network

We show the main framework of the proposed method in Figure 1, where the proposed method is built upon existing GAN-based artistic style transfer models and many model-specific network designs and losses are omitted for better demonstration. The style encoder is optional as some models do not have style image guidance in the generation, *e.g.*, the works of Zhu *et al.* [2017] and Sanakoyeu *et al.* [2018]. To gain control of the style degree, we introduce two novel modules: the Style Scaling Injection (SSI) and the Style Degree Interpretation (SDI) module into existing models.

Style Scaling Injection Module

We assume that the style degree of the stylizations can be explicitly influenced by the magnitude of the feature activations in existing models; thereby, we design the SSI module that scales the feature activations in existing models. As shown in Figure 1, the SSI module first takes the value of Style Degree Factor (SDF), denoted as a random variable F, as the input. During training, the value of f, which indicates the expected style degree of the generated images, is randomly sampled from a uniform distribution $U[0, 1]$. The SSI module then utilizes a learnable non-linear transformation block (TB) which transforms the value of f to parameters that scale the feature activations in existing GAN-based models. We view each

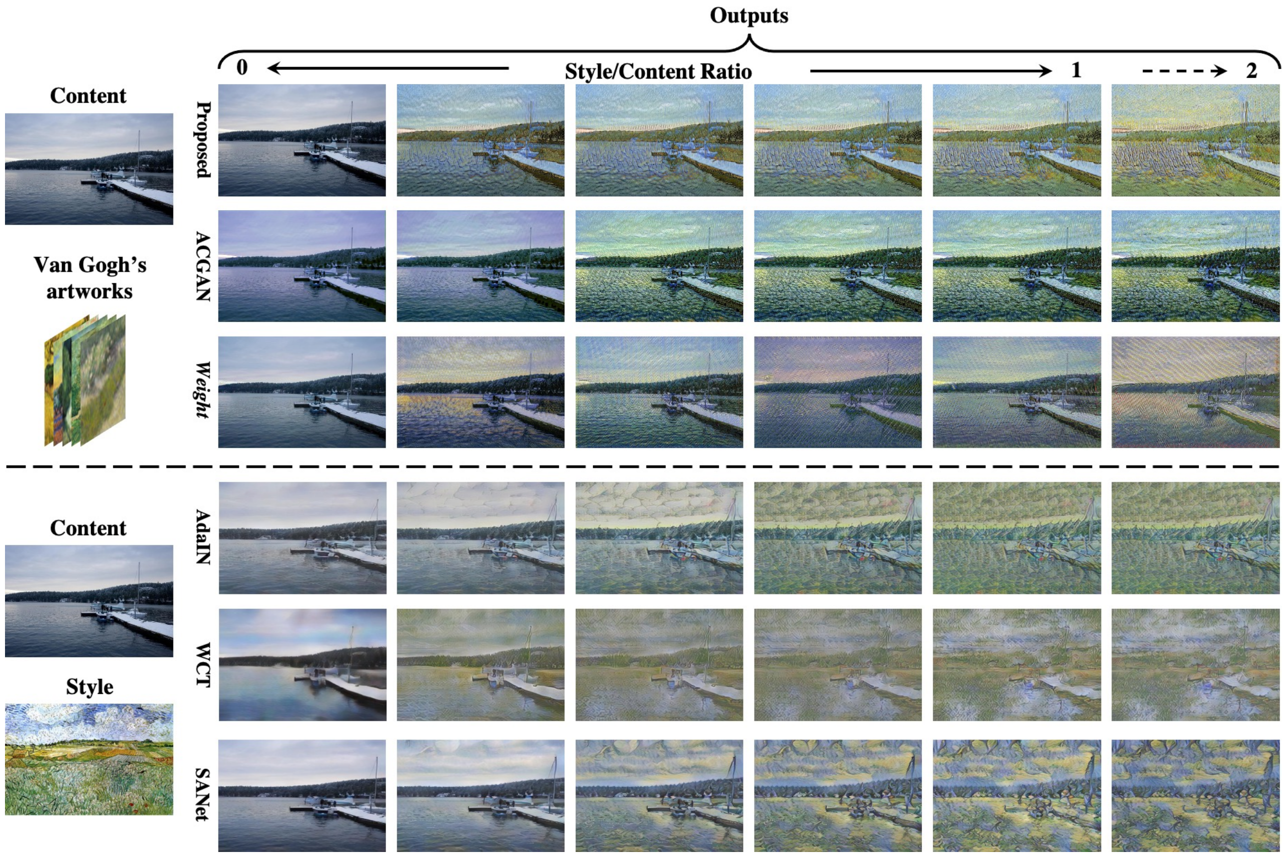

Figure 2: The top part of the figure compares the proposed method built on AST with two methods that may bring style degree control for AST. Refer to text for detailed discussion. And the bottom part of the figure shows the results of some arbitrary style transfer methods which also allow content-style trade-offs. Please zoom in to see more details.

channel of the feature activations as a unit and assign a parameter to scale it. The softplus function $g(x) = \log(1 + e^x)$ is applied to constrain the value range of the output parameters within $(0, +\infty)$. The scaling operation during training can be formulated by the following equations:

$$\gamma = g(TB(f)), f \sim U[0, 1] \tag{2}$$

$$a^i_{scale} = \gamma^i a^i \tag{3}$$

where $g(\cdot)$ and $TB(\cdot)$ denote the softplus and the transformation block's non-linear transformation function respectively, a^i and γ^i denote the i-th feature channel's activations and the scaling parameter for them respectively, and a^i_{scale} denotes i-th feature channel's activations after scaling. At inference, we expect sliding the value of SDF f can smoothly and continuously change the style degree of the generated stylizations.

Style Degree Interpretation Module

The SSI module offers control signals to influence the generation process in existing models. However, without further regularizations, the generated images produced by different values of f can not result in differences in style degree. To make f aware of the smooth and continuous changes in style degree, we additionally introduce a SDI module, which provides a mechanism to parameterize the style degree. The SDI module mainly introduces a multi-scale content-style binary classifier C into existing models, as shown in Figure 1, where the classifier judges whether an input image is an artistic image by its multi-scale features. The classifier C can be built upon the discriminator D in existing models where the classifier and the discriminator share most of the layers. We use a Style Aware loss to train the classifier C to learn the boundaries between the content and the artistic image domain:

$$\mathcal{L}_{SA} = \max_{C} \mathbb{E}_{x_s}[\log C(x_s)] + \mathbb{E}_{x_c}[\log(1 - C(x_c))] \tag{4}$$

where the output probability of the classifier $C(x)$ can be viewed as the likelihood of an image x being an artistic image. As the generated stylized image $G(\cdot, f)$ should reflect the style degree indicated by its input Style Degree Factor f, the output probability of it is then enforced to be consistent with the input f via the following Style Degree loss:

$$\begin{aligned}\mathcal{L}_{SD} = &\min_{G} \mathbb{E}_f[-f \log C(G(\cdot, f)) \\ &- (1 - f) \log(1 - C(G(\cdot, f))], f \sim U[0, 1]\end{aligned} \tag{5}$$

Total Loss

As described above, the proposed method only makes negligible changes to the generator G and the discriminator D in existing GAN-based models. Thus, the proposed method can be

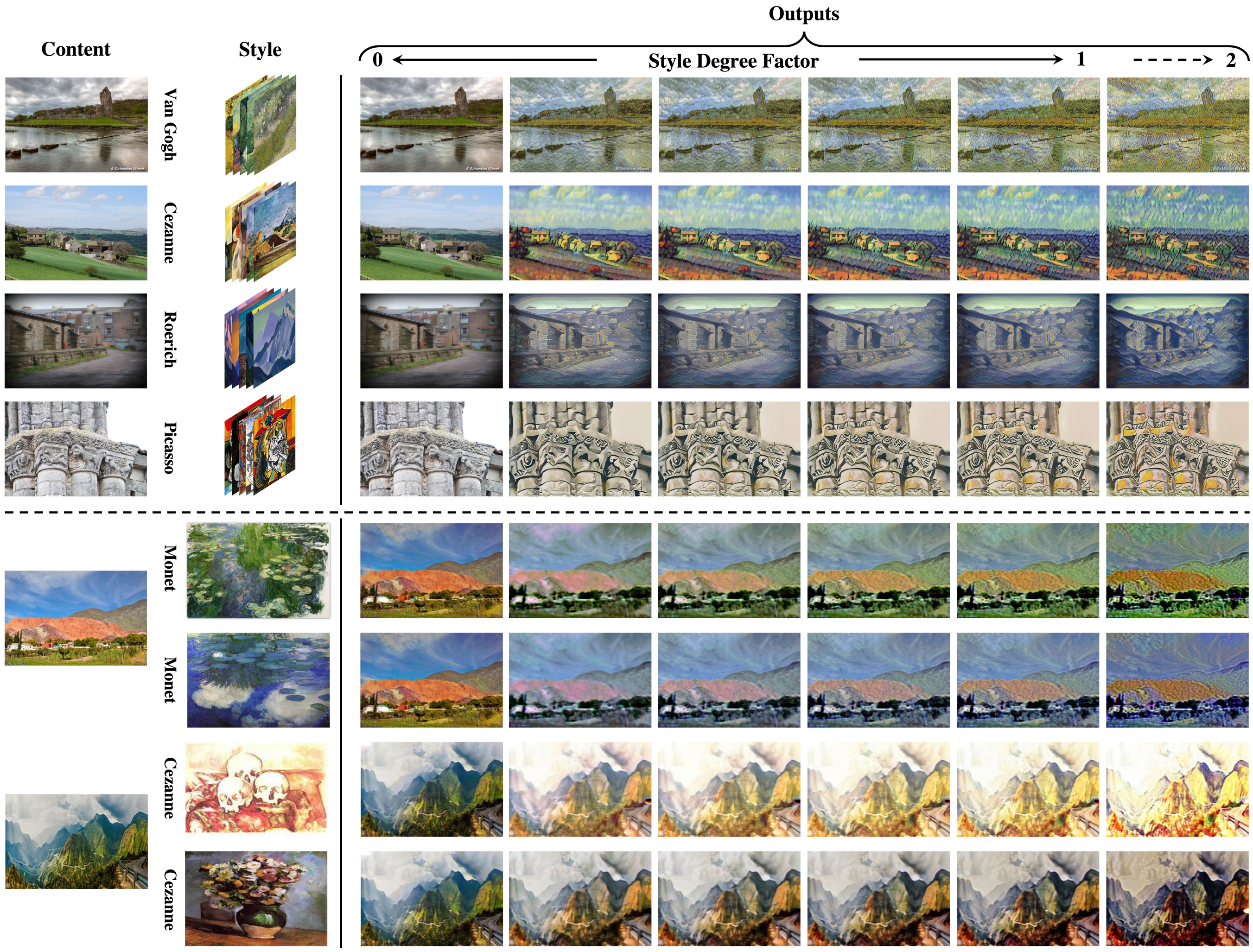

Figure 3: The qualitative results of SFGANs built upon AST and MUNIT are shown in the top and the bottom part of the figure, respectively. Stylized images are generated from uniformly interpolated values of SDF between 0 and 1. Moreover, we also set SDF = 2 for evaluation, which exhibits remarkably extreme stylization effects. Please zoom in to see more details.

easily integrated into them with marginal extra training overheads and modifications. We call the resulting framework as Style Fader Generative Adversarial Network (SFGAN). And the overall training loss of SFGAN is shown below:

$$\mathcal{L}_{\text{SFGAN}} = \alpha_1 \mathcal{L}_{SA} + \alpha_2 \mathcal{L}_{SD} + \mathcal{L}_{ori} \tag{6}$$

where α_1 and α_2 denote the weights of the corresponding losses and $\mathcal{L}_{ori}$ denotes the original losses of the model upon which SFGAN is built.

4 Experiments

To verify the effectiveness of the proposed method, we build SFGAN on two typical GAN-based artistic style transfer models: (1) AST [Sanakoyeu *et al.*, 2018], which learns style representation from an artwork collection and does not have style image guidance in the translation, and (2) MUNIT [Huang *et al.*, 2018], which can perform reference-guided style transfer. The detailed frameworks of SFGANs built on them can be found in the supplementary materials. We refer to their original papers for details about the two models.

4.1 Experimental Settings

We consistently set $\alpha_1 = 1$ and $\alpha_2 = 1$ for SFGANs in the experiments. As for SFGAN built upon AST, the content and style images are image patches of size 768×768 from Places [Zhou *et al.*, 2014] dataset and the artworks of different artists collected from Wikiart [Karayev *et al.*, 2013], respectively. We train separate models on artworks of different artists. As for SFGAN built upon MUNIT, we adopt artwork datasets from [Zhu *et al.*, 2017] and train separate models for each dataset accordingly. We only apply our method to the generator and discriminator that are responsible for translating the content image to the stylized image in MUNIT. SFGAN built on MUNIT is trained with image size 256×256 for both content and style images.

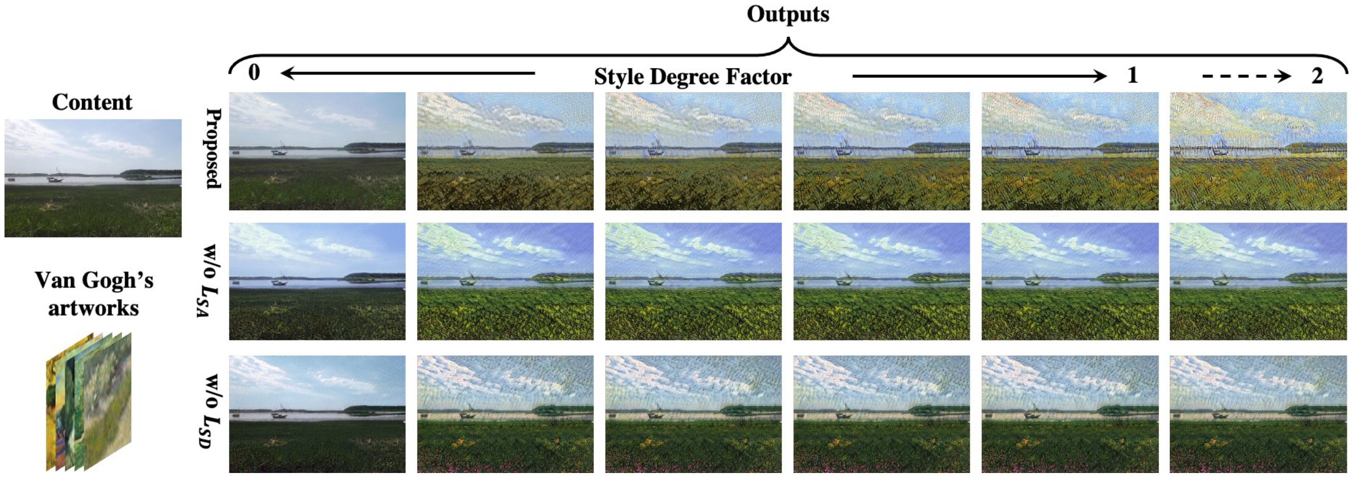

Figure 4: The ablation studies on SFGAN built upon AST model. Ablating either of $\mathcal{L}_{SA}$ or $\mathcal{L}_{SD}$ from our method will lead to a failure.

4.2 Qualitative Comparison

In this section, we first qualitatively compare our method with two methods that may bring style degree control for existing GAN-based artistic style transfer models: **(1)** *Weight*, and **(2)** ACGAN [Odena *et al.*, 2017]. In the comparison, AST [Sanakoyeu *et al.*, 2018] is taken as the baseline model. All the compared methods are trained on Van Gogh's artwork collection [Sanakoyeu *et al.*, 2018]. *Weight* denotes a simple method that adjusts the weight of adversarial loss in AST. ACGAN is taken for comparison as cGANs are well-known for their abilities to yield reasonable interpolations between latent codes or class labels. Moreover, ACGAN incorporates a classifier to guide the class-conditioned image generation, which shares similarities with our method. Specifically, we train ACGAN to reconstruct the input content images when the label is 0 and perform style transfer when the label is 1, where ACGAN adopts the same network architecture as in our method. We would like to examine whether interpolating between label 0 and label 1 in ACGAN can yield images with smooth and continuous style degree changes after training. We use values of (0.00, 0.25, 0.50, 0.75, 1.00, 2.00) for the weight of the adversarial loss, ACGAN's label, and the value of SDF in our method for evaluation.

Except for the above methods, we further show content/style trade-off results of three arbitrary style transfer methods: AdaIN [Huang and Belongie, 2017], WCT [Li *et al.*, 2017b], and SANet [Park and Lee, 2019]. *Note that these arbitrary style transfer methods all rely on the pre-trained DCNN to extract and model the style representation and can not enable style degree control for GAN-based artistic style transfer models.*

As can be observed from the top part of Figure 2, adjusting the weight of adversarial loss in AST brings random changes in stylization effects. Also, the interpolations between ACGAN's labels 0 and 1 result in abrupt changes in style degree (from the 2nd image to the 3rd image in the second row of Figure 2), probably because the images produced by intermediate labels between 0 and 1 in ACGAN are never been trained (*e.g.,* by our proposed Style Degree loss). Only our proposed SFGAN successfully enable AST to yield images with style degree changes in the experiment. In addition, compared with arbitrary style transfer methods like AdaIN, WCT, and SANet, SFGAN built upon AST more accurately learns the artist's holistic style representation. In contrast, the three arbitrary style transfer methods tend to put more style image's textures in the stylizations with the increase in style/content trade-off ratio. Moreover, we show that our method can enable AST model to produce over-stylizations by setting SDF = 2, verifying the effectiveness of our method for learning a style degree-aware representation. Conversely, the arbitrary style transfer methods all fail to yield over-stylizations, as shown in the bottom part of Figure 2 (we multiply the style features that are feeded to the decoder by 2 to produce images in the last column for them).

4.3 Qualitative Results

We show qualitative results of SFGANs built upon AST and MUNIT in the top and bottom part of Figure 3, respectively. The qualitative results clearly demonstrate that the style degree of the generated images continuously and smoothly shifts with the variation in the value of SDF. For example, in the top part of Figure 3, the output images generated by models trained on artwork collections of some impressionist painters like Van Gogh or Cezanne appear more highly organized rhythmic brushstrokes with the increase in value of SDF, which conforms to the painting characteristics of these artists. Also, with the increase in SDF's value, the models trained on Roerich and Picasso's paintings tend to generate images with more geometric deformations. On the other hand, the stylized images demonstrated in the bottom part of Figure 3 show smooth and continuous changes in color, texture, and brightness according to the reference style image by varying the value of SDF. Remarkably, during training, the value of SDF is uniformly sampled from the value range $[0, 1]$, where SFGANs are never trained with values of SDF bigger than 1. However, during inference, we enlarge the value of SDF beyond 1 (*e.g.,* setting it as 2). The stylized images produced with SDF = 2 exhibit consistent style degree increase and extreme stylization effects, verifying the effectiveness of SFGAN for flexibly manipulating the style degree of the generated images and learning a style degree-aware representation. More qualitative results of SFGANs built on AST or MUNIT can be found in the supplementary materials.

4.4 Ablation Studies

We perform ablation studies on our method built upon AST model trained with Van Gogh's artwork collection. Obviously, ablating the SSI module from our framework will cause a lack of control signals to manipulate the style degree of the generated images. As for the SDI module, we ablate $\mathcal{L}_{SA}$ (Equation 4) or $\mathcal{L}_{SD}$ (Equation 5) loss from our framework and show their results in Figure 4. As we see, ablating either of the two losses will lead to a failure of our method.

5 Concluding Remarks

In this paper, we present a novel method that achieves style degree controllability for GAN-based artistic style transfer models by introducing the Style Scaling Injection (SSI) and the Style Degree Interpretation (SDI) module into them. The proposed method can facilitate many existing GAN-based artistic style transfer frameworks with marginal extra training overheads and modifications. Comprehensive qualitative results on two typical GAN-based artistic style transfer frameworks have verified the effectiveness of our method for flexibly adjusting the style degree of the generated images and learning a style degree-aware representation for them.

Ethical Statement

There are no ethical issues.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (business-driven digital cultural heritage trusted sharing and management service system, No. 2020YFC1522704), the National Natural Science Foundation of China (62172365), the Natural Science Foundation of Zhejiang Province (LY21F02005, LY19F020049), the Key Program of the National Social Science Foundation of China (19ZDA197), Zhejiang Cultural Relics Protection Science and Technology Project (2019011), Zhejiang Elite Program (Research and application of Multimedia Fusion digital intelligence service platform based on multimodal data), Key Scientific Research Base for Digital Conservation of Cave Temples of State Administration for Cultural Heritage. MOE Frontier Science Center for Brain Science & Brain-Machine Integration (Zhejiang University).

References

[Babaeizadeh and Ghiasi, 2018] Mohammad Babaeizadeh and Golnaz Ghiasi. Adjustable real-time style transfer. *arXiv preprint arXiv:1811.08560*, 2018.

[Chen and Schmidt, 2016] Tian Qi Chen and Mark Schmidt. Fast patch-based style transfer of arbitrary style. *arXiv preprint arXiv:1612.04337*, 2016.

[Chen *et al.*, 2020] Xuanhong Chen, Xirui Yan, Naiyuan Liu, Ting Qiu, and Bingbing Ni. Anisotropic stroke control for multiple artists style transfer. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 3246–3255, 2020.

[Chen *et al.*, 2021a] Haibo Chen, Zhizhong Wang, Huiming Zhang, Zhiwen Zuo, Ailin Li, Wei Xing, Dongming Lu, et al. Artistic style transfer with internal-external learning and contrastive learning. *Advances in Neural Information Processing Systems*, 34, 2021.

[Chen *et al.*, 2021b] Haibo Chen, Lei Zhao, Zhizhong Wang, Huiming Zhang, Zhiwen Zuo, Ailin Li, Wei Xing, and Dongming Lu. Dualast: Dual style-learning networks for artistic style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 872–881, 2021.

[Chen *et al.*, 2021c] Haibo Chen, Lei Zhao, Huiming Zhang, Zhizhong Wang, Zhiwen Zuo, Ailin Li, Wei Xing, and Dongming Lu. Diverse image style transfer via invertible cross-space mapping. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14860–14869. IEEE Computer Society, 2021.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[Elgammal *et al.*, 2017] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068*, 2017.

[Gatys *et al.*, 2016] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.

[Geirhos *et al.*, 2018] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, 2018.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[Huang and Belongie, 2017] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017.

[Huang *et al.*, 2018] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 172–189, 2018.

[Isola *et al.*, 2017] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[Johnson *et al.*, 2016] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[Karayev *et al.*, 2013] Sergey Karayev, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller. Recognizing image style. *arXiv preprint arXiv:1311.3715*, 2013.

[Kotovenko *et al.*, 2019a] Dmytro Kotovenko, Artsiom Sanakoyeu, Sabine Lang, and Bjorn Ommer. Content and style disentanglement for artistic style transfer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4422–4431, 2019.

[Kotovenko *et al.*, 2019b] Dmytro Kotovenko, Artsiom Sanakoyeu, Pingchuan Ma, Sabine Lang, and Bjorn Ommer. A content transformation block for image style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10032–10041, 2019.

[Lee *et al.*, 2018] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang. Diverse image-to-image translation via disentangled representations. In *Proceedings of the European conference on computer vision (ECCV)*, pages 35–51, 2018.

[Li and Wand, 2016a] Chuan Li and Michael Wand. Combining markov random fields and convolutional neural networks for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2479–2486, 2016.

[Li and Wand, 2016b] Chuan Li and Michael Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *European conference on computer vision*, pages 702–716. Springer, 2016.

[Li *et al.*, 2017a] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Diversified texture synthesis with feed-forward networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3920–3928, 2017.

[Li *et al.*, 2017b] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 385–395, 2017.

[Li *et al.*, 2019] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. Learning linear transformations for fast image and video style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3809–3817, 2019.

[Mirza and Osindero, 2014] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[Odena *et al.*, 2017] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional image synthesis with auxiliary classifier gans. In *International conference on machine learning*, pages 2642–2651. PMLR, 2017.

[Park and Lee, 2019] Dae Young Park and Kwang Hee Lee. Arbitrary style transfer with style-attentional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5880–5888, 2019.

[Sanakoyeu *et al.*, 2018] Artsiom Sanakoyeu, Dmytro Kotovenko, Sabine Lang, and Bjorn Ommer. A style-aware content loss for real-time hd style transfer. In *proceedings of the European conference on computer vision (ECCV)*, pages 698–714, 2018.

[Sheng *et al.*, 2018] Lu Sheng, Ziyi Lin, Jing Shao, and Xiaogang Wang. Avatar-net: Multi-scale zero-shot style transfer by feature decoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8242–8250, 2018.

[Simonyan and Zisserman, 2014] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[Svoboda *et al.*, 2020] Jan Svoboda, Asha Anoosheh, Christian Osendorfer, and Jonathan Masci. Two-stage peer-regularized feature recombination for arbitrary image style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13816–13825, 2020.

[Ulyanov *et al.*, 2016] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, and Victor S Lempitsky. Texture networks: Feed-forward synthesis of textures and stylized images. In *ICML*, volume 1, page 4, 2016.

[Ulyanov *et al.*, 2017] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6924–6932, 2017.

[Wang *et al.*, 2019] Xintao Wang, Ke Yu, Chao Dong, Xiaoou Tang, and Chen Change Loy. Deep network interpolation for continuous imagery effect transition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1692–1701, 2019.

[Wang *et al.*, 2020] Zhizhong Wang, Lei Zhao, Haibo Chen, Lihong Qiu, Qihang Mo, Sihuan Lin, Wei Xing, and Dongming Lu. Diversified arbitrary style transfer via deep feature perturbation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7789–7798, 2020.

[Wang *et al.*, 2021] Zhizhong Wang, Lei Zhao, Haibo Chen, Ailin Li, Zhiwen Zuo, Wei Xing, and Dongming Lu. Texture reformer: Towards fast and universal interactive texture transfer. *arXiv preprint arXiv:2112.02788*, 2021.

[Yao *et al.*, 2019] Yuan Yao, Jianqiang Ren, Xuansong Xie, Weidong Liu, Yong-Jin Liu, and Jun Wang. Attention-aware multi-stroke style transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1467–1475, 2019.

[Yu *et al.*, 2019] Ning Yu, Connelly Barnes, Eli Shechtman, Sohrab Amirghodsi, and Michal Lukac. Texture mixer: A network for controllable synthesis and interpolation of texture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12164–12173, 2019.

[Zhang *et al.*, 2019] Yulun Zhang, Chen Fang, Yilin Wang, Zhaowen Wang, Zhe Lin, Yun Fu, and Jimei Yang. Multimodal style transfer via graph cuts. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5943–5951, 2019.

[Zhou *et al.*, 2014] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. Learning deep features for scene recognition using places database. 2014.

[Zhu *et al.*, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.