Image Style Transfer with Generative Adversarial Networks

Ru Li
University of Electronic Science and Technology of China
Chengdu, China
liru@std.uestc.edu.cn

ABSTRACT

Image style transfer is a recently popular research field, which aims to learn the mapping between different domains and involves different computer vision techniques. Recently, Generative Adversarial Networks (GAN) have demonstrated their potentials of translating images from source domain X to target domain Y in the absence of paired examples. However, such a translation cannot guarantee to generate high perceptual quality results. Existing style transfer methods work well with relatively uniform content, they often fail to capture geometric or structural patterns that reflect the quality of generated images. The goal of this doctoral research is to investigate the image style transfer approaches, and design advanced and useful methods to solve existing problems. Though preliminary experiments conducted so far, we demonstrate our insights on the image style translation approaches, and present the directions to be pursued in the future.

CCS CONCEPTS

• **Computing methodologies → Computational photography**.

KEYWORDS

image style transfer, image-to-image translation, generative adversarial networks

ACM Reference Format:
Ru Li. 2021. Image Style Transfer with Generative Adversarial Networks. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21), October 20–24, 2021, Virtual Event, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3474085.3481029

1 INTRODUCTION

Stylizing images in an artistic manner has been widely studied in the domain of non-photorealistic rendering (NPR) [42]. Traditional approaches develop dedicated algorithms for specific styles. However, substantial efforts are required to produce fine-grained styles that mimic individual artists. Recently, learning-based style transfer methods, in which an image can be stylized based on provided examples, have drawn considerable attention [11, 18]. In particular, the power of Generative Adversarial Networks (GANs) [13] is explored to achieve high-quality style transfer, with the distinct feature that the model is trained using two different domain images. The GAN-based image-to-image translation is to capture special characteristics for one image collection and figure out how these characteristics could be translated to target image collection. Many researches have produced attractive transfer results, where image pairs are required [17, 21]. However, the major limitation is that they use paired images as supervision, which is impractical for many tasks due to the challenge of obtaining such corresponding ground truth. Some style transfer works avoid the need for such paired datasets by introducing the unpaired image-to-image translation using generative adversarial networks [4, 50].

A problem when applying style transfer to challenging input images with complex spatial layouts is that the synthesized images tend to distribute style elements evenly across the whole image, making the holistic structure become unrecognizable, therefore, the results are not entirely satisfactory. For inputs with prominent fundamental characteristics or sensitive to structure distortion, the uniformly distributed textures further obscure weak details and destroy the original structure. To solve the problems of detail preservation, Liu *et al.* proposed to improve style transfer results by preserving salient information of content images, which adds a localization network to calculate the region loss. However, it tends to generate discontinuous salient regions and background if it cannot keep the balance between style loss and region loss [33]. Cheng *et al.* combined the depth and edge information to construct a structure representation to solve the disruption of content, which is limited by a single style image and not robust for structure preservation [5].

The goal of this PhD research is to explore the principle of style transfer and propose advanced and useful methods to solve existing problems. We have proposed two GAN-based methods to generate high-quality translation results when the inputs are single image and multi-images, respectively. As for single input, we propose a method to preserve saliency detail information while correctly mapping unpaired images from source domain to target domain, named SDP-GAN. The SDP-GAN introduces an extra saliency network that concurrently predicts the saliency map, which helps the calculation of newly proposed objective functions. The encoder part of the saliency branch is also concatenated into the generator network to yield a smooth overall style transformation, with content details properly preserved in salient regions. As for multi-input images, we refer to the concept of high dynamic range image (HDR) fusion techniques [25–27], to combine multi-input images to generate more informative transfer results. Multi-input images can provide more scene information to the network. However, capturing totally static images at the same time is difficult. A common strategy used in HDR fusion is to obtain a stack of different exposure images and merge them together. We propose a GAN-based method to align the multi-input images and combine the scene information from them simultaneously.

MM '21, October 20–24, 2021, Virtual Event, China

ACM ISBN 978-1-4503-8651-7/21/10…$15.00
https://doi.org/10.1145/3474085.3481029

2 BACKGROUND AND RELATED WORK

2.1 Generative Adversarial Networks

The concept of GAN was originally proposed by Goodfellow *et al.* [13], which has drawn substantial attention in the field of deep learning. The conventional GAN includes two modules, the generator G and the discriminator D, to achieve desirable transformation and generate new samples which are indistinguishable with target images. This framework corresponds to a min-max two-player game, which can be described as:

$$G^*, D^* = \arg\min_G \max_D L(G, D) \tag{1}$$

We adopt a wide diversity of real photos $\{x_i\}_{i=1,\dots,N} \in X$ as our source domain data, and a collection of artistic images $\{y_j\}_{j=1,\dots,M} \in Y$, as the target domain data. The original adversarial loss can be represented as:

$$\begin{aligned} L_{\text{GAN}}(G, D) = {} & E_{y\sim p_{\text{data}}(y)}[\log D(y)] \\ & + E_{x\sim p_{\text{data}}(x)}[\log(1 - D(G(x))] \end{aligned} \tag{2}$$

GANs have achieved impressive results in different image synthesis tasks. The GAN-based image-to-image translation is defined as the problem of translating a possible representation of one scene into another, which can be viewed as a generalization of style transfer, since it is not limited to transfer the styles of images, but can also manipulate attributes of objects. Isola *et al.* proposed Pix2Pix for pixel-to-pixel image translation problems [17]. Although Pix2Pix produces very impressive synthetic images, the major limitation is that it must use paired images as supervision. It requires paired image sets for the training process which is impractical for many tasks due to the challenge of obtaining such corresponding ground truth. Zhu *et al.* introduced CycleGAN to perform image translation with unpaired training data, which trains two sets of GAN models at the same time, mapping from class A to class B and from class B to class A, respectively [50]. Karras *et al.* proposed StyleGAN to embed the latent code to each convolutional block, which can achieve image modification more delicate [22].

The strength of generative adversarial networks makes it ideal for solving various image processing problems, especially for those where the perceptual quality of outputs is the primary evaluation criterion. GAN-based image translation has applied in many vision tasks, including image restoration and enhancement [1], super resolution [23], image denoising [3] and image inpainting [49]. Based on these methods, we proposed two GAN-based networks to achieve high-quality image style translation with single-input image and multi-input images, respectively.

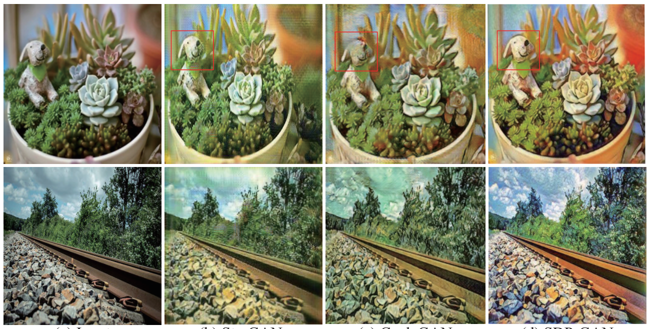
2.2 Image Style Translation

Many non-photorealistic rendering methods have been developed since the mid-1990*s*, and nowadays NPR is a firmly established field in computer graphics [12]. There are some works proposed to mimic specific artistic styles [42]. However, these methods use low-level image features and often fail to capture image structures effectively, such as making object boundaries clear. Inspired by the development of convolution neural networks (CNN), Gatys *et al.* first applied CNN feature activations to recombine the content of a given photo and the style of famous artworks [10]. Their subsequent work [11] used the feature maps of a pre-trained VGG network to optimize the content, and captured texture information of the style image using the Gram matrix [9]. However, the algorithm of Gatys *et al.* does not perform well in preserving the coherence of fine structures and details. Also, it generally fails for photorealistic synthesis, due to the limitations of Gram-based style representation. To solve the problem, some derived Gram-based representation methods are put forward to encode style information [41]. Then, some efficient methods are proposed to optimize a generative model offline and produce the stylized image with a single forward pass [31]. Johnson *et al.* aimed to pre-train a feed-forward style-specific network and produced a stylized result with the forward pass at the testing stage [18]. The design basically follows Gatys *et al.* [11], which suffers from the same aforementioned issues. Recently, some structure-preserving style transfer methods are proposed [5, 32, 33]. Liu *et al.* focused on salient regions in style transfer with the help of region loss, computed by introducing a localization network [33]. Cheng *et al.* used the depth and edge information to construct a structure representation to solve the disruption of content [5]. Liu *et al.* applied the saliency map to fuse the real image and stylized image, generating an output with stylized salient regions and real background regions [32]. All these methods, however, use a single style image and generate results whose style heavily depends on the chosen style image.

For arbitrary style transfer, a few methods [16] adjusted the content features to match the statistics of the style features. Isola *et al.* [17] and Li *et al.* [24] applied generative adversarial learning to achieve the transformation between two domains. Zhu *et al.* further introduced CycleGAN to train unpaired image datasets [50]. Choi *et al.* aimed at obtaining one single model to transfer multiple artistic styles [6] . Park *et al.* concentrated on attention-based style transfer techniques, which lead to some improvement, but still unsatisfactory for generating high-quality results [39].

The inputs of the aforementioned methods are noise or a single image. Some recent image translation works are proposed for combining multi-input images with generative adversarial networks. Perera *et al.* extended unsupervised image-to-image translation to multiple input settings using GAN and validated their method on several tasks, including multi-spectral images to visible images and synthetic to real image translations [40]. Joo *et al.* applied GAN to generate a fusion image combined the shape of image x_1 and the identity of image x_2 with the help of min-patch training [19]. Ma *et al.* used GAN to fuse infrared and visible information, obtaining a fused image with major infrared intensities together with additional visible gradients [34]. Multi-input image synthesis often appears in multi-exposure image fusion research field. Yang *et al.* fused two inputs, the under-exposed image and the over-exposed image, to generate an informative output [48]. Niu *et al.* proposed a reference-based residual merging block for aligning large object motions in the feature domain, and a deep supervision scheme for eliminating artifacts of the reconstructed images [36].

No matter the input is single-input image or multi-input images, the detail preservation in generated images is necessary, which leads to a better perspective experience. To solve such problems that appeared in conventional image style translation methods, we introduce two GAN-based detail preservation architectures to improve the total performance.



(a) Input (b) StarGAN (c) CycleGAN (d) SDP-GAN

Figure 1: The top result of CycleGAN [50] loses some information in the doll's head region, and the results of StarGAN [6] cannot produce desired style transformation, while our method solves these problems well.

3 WORK IN PROGRESS

3.1 Style Transfer with Single Images

Although large improvements have been achieved with learning-based stylization over unpaired datasets, state-of-the-art (SOTA) methods often fail to produce satisfying visual results. These methods aim to transfer the holistic style from the source domain X to a new domain $\hat{Y}$ that has identical distribution to the target domain Y. However, such a translation does not guarantee a high perceptual quality style transfer for the whole image, even if each region in an individual x is correctly transferred into the target style. According to our observation, traditional approaches often produce two types of problems: loss of details in structured regions, and excess of details in smooth regions. One possible explanation for this is that previous models rely heavily on convolution to model the dependencies across different image regions. To achieve better perceptual quality, the transformation should preserve more edge details in structured regions and reduce unwanted artifacts in smooth regions, while the overall style is maintained. For learning-based methods, increasing the size of the convolution kernels can improve the representational capacity of the network but doing so also loses the computational and statistical efficiency.

We propose a GAN-based method to preserve saliency detail information while correctly mapping unpaired images from source domain to target domain, named SDP-GAN [30]. The SDP-GAN introduces an extra saliency network that concurrently predicts the saliency map, which helps the calculation of newly proposed objective functions. The encoder part of the saliency branch is also concatenated into the generator network to yield a smooth overall style transformation, with content details properly preserved in salient regions. Fig. 1 shows the comparisons with two classic GAN-based style transfer methods, CycleGAN [50] and StarGAN [6], in terms of style transformation. Fig. 1 (b) and (c) display their results with a uniform style across the whole image. However, more favorable style transfer results can be obtained when image content is better preserved in certain regions, as shown in the red box in Fig. 1 (d). The top result of CycleGAN apparently loses some information in the doll's head region. This artifact appears not only in CycleGAN, but in other latest style transfer methods [4, 11] as well. Our top result demonstrates that the stylization quality in salient regions is important and should be handled properly. For the comparisons with StarGAN, the proposed method shows a more attractive style transformation than StarGAN.

Table 1: Quantitative comparisons of the proposed SDP-GAN with other representative GAN methods in terms of IS and FID scores.

Styles	IS and FID	StarGAN	CartoonGAN	CycleGAN	SDP-GAN
Miyazaki Hayao	IS	5.48±0.52	6.09±0.83	4.37±0.58	**6.38±0.98**
	FID	169.42	159.69	136.82	**133.76**
Van Gogh	IS	4.28±0.53	4.77±0.48	4.58±0.74	**4.86±0.59**
	FID	162.89	135.93	106.94	**101.59**
Ukiyo-e	IS	5.33±0.52	**6.10±0.72**	5.75±0.64	6.07±0.77
	FID	152.46	123.42	107.25	**101.92**
MEAN	IS	5.03±0.52	5.65±0.68	4.9±0.65	**5.77±0.78**
	FID	161.59	139.68	117	**112.42**

In addition to visual comparisons, we also organize quantitative comparisons against several GAN-based style transfer methods, StarGAN [6], CartoonGAN [4] and CycleGAN [50], to validate the superior performance of SDP-GAN in Table 1. We choose the Inception score (IS) [43] and the Fréchet Inception distance (FID) [14] for quantitative evaluation. IS computes the KL divergence between the conditional class distribution and the marginal class distribution. Higher Inception score indicates better image quality. FID is a more principled and comprehensive metric, and has been shown to be more consistent with human evaluation in assessing the realism and variation of the generated samples. Lower FID values mean closer distances between synthetic and real data distributions. In quantitative comparisons, the proposed SDP-GAN achieves the best IS and FID scores on average, which demonstrates the superiority of the proposed method.

3.2 Style Transfer with Multi-input Images

Some recent image translation approaches are proposed for synthesizing multi-input images, which provide more scene information to generate more informative results. However, capturing totally static images at the same time is difficult. Refer to the multi-exposure image fusion techniques [26], a common strategy is to obtain a stack of different exposure images and merge them together.

The original multi-input image synthesis approaches work only for static scenes because they typically assume constant radiance at each pixel over all exposures. If the scenes exist moving content, these methods produce ghosting artifacts from even small misalignments between inputs. Then, some methods [38] are introduced to synthesize inputs with scene motions or dynamic objects using homography or optical flow. Generally, the middle image is considered as the reference image and the other inputs are aligned to the reference. These algorithms can generate high-quality translation results if inputs are fully aligned, but will produce ghosting and blurring artifacts if there are unaligned errors. Some patch-based methods [15, 45] are then proposed to reconstruct the input images by patch-based synthesis according to one selected reference image, to form a fully registered image stack. However, patch-based reconstruction is not always robust in complicated situations.

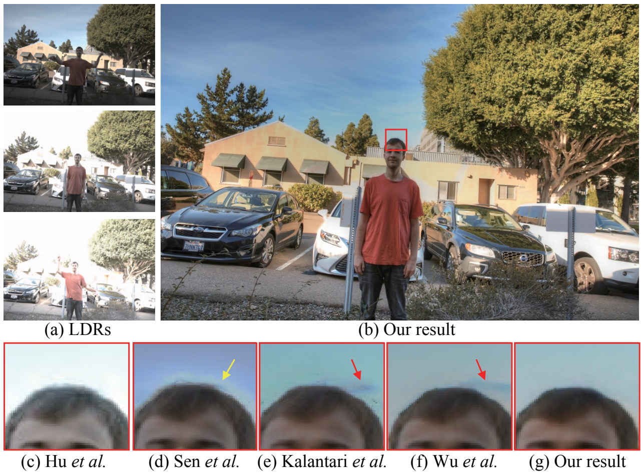


Figure 2: Input images are shown in (a), and our result is shown in (b). (c) Result of Hu *et al.*'s method [15]. (d) Result of Sen *et al.*'s method [45]. (e) Result of Kalantari *et al.*'s method [20]. (f) Result of Wu *et al.*'s method [46]. (g) Zoomed-in areas of our result. The proposed method generates results with fewer artifacts and higher quality.

Recent works [20, 46, 47] used deep neural networks (DNN) to learn the synthesis process. Kalantari *et al.* [20] first proposed a DNN-based generation method, whose inputs are aligned using optical flow before they are sent to the network. However, their method fails to handle the artifacts and distortions caused by the misalignment error of flow-based methods. Wu *et al.* [46] applied a simple homography to align the background first and used the network to achieve main alignment. Yan *et al.* [47] introduced an attention-guided network to detect and align dynamic objects before fusion. However, they still suffer from artifacts when the input images contain large motions or significant misalignment, due to the unreliability of image registration.

We propose a GAN-based multi-input image translation method, which learns the mapping between source and target domains and transforms the inputs to the informative outputs [29]. The proposed method successfully fuses multi-inputs and generates high-quality images by introducing an initialization phase, a modified GAN loss and min-patch training. Fig. 2 shows the comparisons with several typical multi-input synthesis methods, including two patch-based methods [15, 45] and two deep learning-based methods [20, 46]. Our method deals with the multi-inputs properly with the assistance of the modified GAN loss and min-patch training, which pay more attention to the dynamic regions and emphasize the edges.

In addition to visual comparisons, we also organize quantitative comparisons, including PSNR, SSIM and HDR-VDP-2 [35] between the generated images and the real images, to further evaluate the performance of the proposed methods. Table 2 exhibits the comparison results of the proposed method with several multi-input image synthesis methods [15, 20, 45, 46]. Although the comparison methods get similar quantitative scores, the proposed method owns superior performance overall.

Table 2: Quantitative comparison of the proposed method with several SOTA multi-input image synthesis methods.

	PSNR	SSIM	HDR-VDP-2
Sen [45]	40.924	0.9806	56.239
Hu [15]	34.785	0.9725	55.754
Kalantari [20]	42.532	**0.9871**	59.824
Wu [46]	41.667	0.9844	59.908
Ours	**42.857**	0.9869	**60.052**

4 FUTURE WORK

There are numerous self-supervised methods consider solving jigsaw puzzles as pre-text tasks. This learning strategy is a recent variation on the unsupervised learning theme that transfer the pre-trained network parameters on jigsaw puzzle task to other visual recognition tasks [7, 28]. Noroozi *et al.* introduced a context-free network (CFN) to separate the pieces in convolutional process. Their main architecture focused on a subset of possible permutations involving all the image tiles and solved a classification problem [37]. Santa *et al.* proposed to handle the whole set by approximating the permutation matrix and solving a bi-level optimization problem to recover the right ordering [44]. The above methods tackle the problem by dealing with the separate pieces and then finding a way to recombine them. Carlucci *et al.* proposed JiGen to train the jigsaw classifier and object classifier simultaneously. They focused on domain generalization tasks by considering that the jigsaw puzzle solver can improve semantic understanding [2]. Du *et al.* combined the jigsaw puzzle and progressive training to optimize the fine-grained classification by learning which granularities are the most discriminative and how to fuse information cross multi-granularity [8]. These methods assume that a rich universal representation has been captured in pre-trained model, which is useful to be fined-tuned with task-specific data using various strategies. Inspired by their method, we plan to design a jigsaw puzzle solver and transfer the pre-trained parameters to image style tasks, which can concentrate more on the semantic information and further promote the quality of image translation.

5 CONCLUSION

In this paper, we summarize our study on image style transfer with generative adversarial networks. For image style transfer with single-input image, we proposed the SDP-GAN to achieve detail preservation, which introduces an extra saliency network that concurrently predicts the saliency map, which helps the calculation of newly proposed objective functions. For multi-input images, we introduce an initialization phase, a modified GAN loss and min-patch training to align the multi-input images and combine the scene information from them simultaneously. Based on current works, we consider the following directions for future research: designing a jigsaw puzzle solver and transferring the pre-trained parameters to image style tasks to achieve high-quality image translation.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (NSFC) under Grants 61872067, 61720106004 and 62071097.

REFERENCES

[1] Yochai Blau and Tomer Michaeli. 2018. The perception-distortion tradeoff. In *Proc. CVPR.* 6228–6237.
[2] Fabio M Carlucci, Antonio D'Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. 2019. Domain generalization by solving jigsaw puzzles. In *Proc. CVPR.* 2229–2238.
[3] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. 2018. Image blind denoising with generative adversarial network based noise modeling. In *Proc. CVPR.* 3155–3164.
[4] Yang Chen, Yu-Kun Lai, and Yong-Jin Liu. 2018. Cartoongan: Generative adversarial networks for photo cartoonization. In *Proc. CVPR.* 9465–9474.
[5] Mingming Cheng, Xiaochang Liu, Jie Wang, Shaoping Lu, Yukun Lai, and Paul L Rosin. 2020. Structure-Preserving Neural Style Transfer. *IEEE Trans. on Image Processing* 29 (2020), 909–920.
[6] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. 2018. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proc. CVPR.* 8789–8797.
[7] Carl Doersch, Abhinav Gupta, and Alexei A Efros. 2015. Unsupervised visual representation learning by context prediction. In *Proc. ICCV.* 1422–1430.
[8] Ruoyi Du, Dongliang Chang, Ayan Kumar Bhunia, Jiyang Xie, Zhanyu Ma, Yi-Zhe Song, and Jun Guo. 2020. Fine-grained visual classification via progressive multi-granularity training of jigsaw patches. In *Proc. ECCV.* 153–168.
[9] Leon Gatys, Alexander S Ecker, and Matthias Bethge. 2015. Texture synthesis using convolutional neural networks. In *Proc. NIPS.* 262–270.
[10] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015).
[11] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In *Proc. CVPR.* 2414–2423.
[12] Bruce Gooch and Amy Gooch. 2001. *Non-photorealistic rendering.* AK Peters/CRC Press.
[13] Ian Goodfellow, Jean Pougetabadie, Mehdi Mirza, Bing Xu, David Wardefarley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Proc. NIPS.* 2672–2680.
[14] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Proc. NIPS.* 6626–6637.
[15] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. 2013. HDR deghosting: How to deal with saturation?. In *Proc. CVPR.* 1163–1170.
[16] Xun Huang and Serge Belongie. 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proc. ICCV.* 1501–1510.
[17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proc. CVPR.* 1125–1134.
[18] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *Proc. ECCV.* 694–711.
[19] Donggyu Joo, Doyeon Kim, and Junmo Kim. 2018. Generating a Fusion Image: One's Identity and Another's Shape. In *Proc. CVPR.* 1635–1643.
[20] Nima Khademi Kalantari and Ravi Ramamoorthi. 2017. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph* 36, 4 (2017), 144.
[21] Levent Karacan, Zeynep Akata, Aykut Erdem, and Erkut Erdem. 2016. Learning to generate images of outdoor scenes from attributes and semantic layouts. *arXiv preprint arXiv:1612.00215* (2016).
[22] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proc. CVPR.* 4401–4410.
[23] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proc. CVPR.* 4681–4690.
[24] Chuan Li and Michael Wand. 2016. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *Proc. ECCV.* 702–716.
[25] Ru Li, Xiaowu He, Shuaicheng Liu, Guanghui Liu, and Bing Zeng. 2018. Photomontage for robust hdr imaging with hand-held cameras. In *Proc. ICIP.* 1708–1712.
[26] Ru Li, Shuaicheng Liu, Guanghui Liu, Tiecheng Sun, and Jishun Guo. 2020. Multi-exposure photomontage with hand-held cameras. *Computer Vision and Image Understanding* 193 (2020), 102929.
[27] Ru Li, Shuaicheng Liu, Guanghui Liu, and Bing Zeng. 2019. Hybrid Synthesis for Exposure Fusion from Hand-Held Camera Inputs. In *Proc. ICIP.* 4639–4643.
[28] Ru Li, Shuaicheng Liu, Guangfu Wang, Guanghui Liu, and Bing Zeng. 2021. JigsawGAN: Self-supervised Learning for Solving Jigsaw Puzzles with Generative Adversarial Networks. *arXiv preprint arXiv:2101.07555* (2021).
[29] Ru Li, Chuan Wang, Shuaicheng Liu, Jue Wang, Guanghui Liu, and Bing Zeng. 2021. UPHDR-GAN: Generative Adversarial Network for High Dynamic Range Imaging with Unpaired Data. *arXiv preprint arXiv:2102.01850* (2021).
[30] Ru Li, Chi-Hao Wu, Shuaicheng Liu, Jue Wang, Guangfu Wang, Guanghui Liu, and Bing Zeng. 2020. SDP-GAN: Saliency Detail Preservation Generative Adversarial Networks for High Perceptual Quality Style Transfer. *IEEE Trans. on Image Processing* 30 (2020), 374–385.
[31] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. 2017. Diversified texture synthesis with feed-forward networks. In *Proc. CVPR.* 3920–3928.
[32] Xiuwen Liu, Zhi Liu, Xiaofei Zhou, and Minyu Chen. 2019. Saliency-Guided Image Style Transfer. In *International Conference on Multimedia and Expo.* 66–71.
[33] Yijun Liu, Zuoteng Xu, Wujian Ye, Ziwen Zhang, Shaowei Weng, Chinchen Chang, and Huajin Tang. 2019. Image Neural Style Transfer With Preserving the Salient Regions. *IEEE Access* 7 (2019), 40027–40037.
[34] Jiayi Ma, Wei Yu, Pengwei Liang, Chang Li, and Junjun Jiang. 2019. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Information Fusion* 48 (2019), 11–26.
[35] Rafat Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph* 30, 4 (2011), 1–14.
[36] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau. 2020. HDR-GAN: HDR image reconstruction from multi-exposed ldr images with large motions. *arXiv preprint arXiv:2007.01628* (2020).
[37] Mehdi Noroozi and Paolo Favaro. 2016. Unsupervised learning of visual representations by solving jigsaw puzzles. In *Proc. ECCV.* 69–84.
[38] Taehyun Oh, Joonyoung Lee, Yuwing Tai, and In So Kweon. 2015. Robust High Dynamic Range Imaging by Rank Minimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 37, 6 (2015), 1219–1232.
[39] Dae Young Park and Kwang Hee Lee. 2019. Arbitrary Style Transfer With Style-Attentional Networks. In *Proc. CVPR.* 5880–5888.
[40] Pramuditha Perera, Mahdi Abavisani, and Vishal M Patel. 2018. In2I: Unsupervised Multi-Image-to-Image Translation Using Generative Adversarial Networks. In *Proc. ICPR.* 140–146.
[41] Eric Risser, Pierre Wilmot, and Connelly Barnes. 2017. Stable and controllable neural texture synthesis and style transfer using histogram losses. *arXiv preprint arXiv:1701.08893* (2017).
[42] Paul Rosin and John Collomosse. 2012. *Image and video-based artistic stylisation.* Vol. 42. Springer Science & Business Media.
[43] Tim Salimans, Han Zhang, Alec Radford, and Dimitris Metaxas. 2018. Improving GANs using optimal transport. *arXiv preprint arXiv:1803.05573* (2018).
[44] Rodrigo Santa Cruz, Basura Fernando, Anoop Cherian, and Stephen Gould. 2017. Deeppermnet: Visual permutation learning. In *Proc. CVPR.* 3949–3957.
[45] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. 2012. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Trans. Graph* 31, 6 (2012), 1–11.
[46] Shangzhe Wu, Jiarui Xu, Yuwing Tai, and Chikeung Tang. 2018. Deep High Dynamic Range Imaging with Large Foreground Motions. In *Proc. ECCV.* 120–135.
[47] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton Van Den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. 2019. Attention-Guided Network for Ghost-Free High Dynamic Range Imaging. In *Proc. CVPR.* 1751–1760.
[48] Zhiguang Yang, Youping Chen, Zhuliang Le, and Yong Ma. 2020. GANFuse: a novel multi-exposure image fusion method based on generative adversarial networks. *Neural Computing and Applications* (2020), 1–13.
[49] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. 2018. Generative image inpainting with contextual attention. In *Proc. CVPR.* 5505–5514.
[50] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. ICCV.* 2223–2232.