Check for updates

IPSM-GAN: A Generative Adversarial Network for Shadow Removal Guided by Mixed Shadow Masks

Chuang Xie[1]

[1] School of software, Hefei University of Technology, 230009, Hefei, China

tsechong@mail.hfut.edu.cn

Abstract. Recently, thanks to the rapid development of deep learning, there are many methods to remove shadows in images by using generative adversarial networks. Most of them can learn the relationship between different domains, like shadow and shadow-free areas, to transform the shadow areas into areas with no shadow. However, due to inaccurate shadow shapes or masks obtained, these methods can not lead to a better performance in the shadow image and even create more artifacts. To solve these problems, the authors propose IPSM-GAN, a new framework that learns to remove shadows in images by formulating cycle-consistency constraints and the guidance of mixed shadow masks. The mixed shadow mask generation method can accurately capture the shape of shadows. Also, the method can be a guide to the learning of the framework, which makes IPSM-GAN achieve better performance in removing shadows. Extensive experimental results verify the effectiveness of the proposed method, which can provide some new insights into the research field of shadow removal.

Keywords: Shadow removal; generative adversarial network; deep learning

1 Introduction

Shadow is a degrading phenomenon caused by poor imaging conditions. In most cases, a shadow is a darker area created by a light source being blocked by an opaque object. The shadow will weaken the optical information in the image, reduce the accuracy and clarity of the image, and make the amount of information reflected by the target missing or disturbed. Shadows are commonly found among digital images.

© The Author(s) 2023

P. Kar et al. (eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Advances in Computer Science Research 108,
https://doi.org/10.2991/978-94-6463-300-9_82

As a result, the image information loss caused by shadow will seriously affect various analyses and applications of images, and adversely affect the subsequent works like feature extraction, image matching, and image processing. Therefore, it is very necessary to detect and process the shadow areas in the image, which can be beneficial to the subsequent image processing.

In earlier times, some edge detection algorithms based on detecting gradient, such as the Canny edge detection algorithm1] and Sobel-Feldman algorithm2], were used to detect the contour of the shadow and select the shadow areas. These methods were suitable for detecting the obvious edges in the image. After obtaining the edges of the shadow areas, additional steps are required to distinguish the shadow areas from other areas, such as foreground areas and background areas. In the past 15 years, with the advent of convolutional neural networks (CNNs), their powerful feature representation capabilities make it possible to capture various information in images accurately. As a result, the shadow detection and processing methods related to convolutional neural network has better performance than the traditional image processing algorithms in some situations. For example, in Khan's works [3], convolutional neural networks were used to detect shadow ares in images. Later, generative adversarial network (GANs) based on unsupervised learning were also adopted to remove shadows. During the training of the generative adversarial network, the generators and discriminators of the networks are opposed to each other, and gradually acquire better performance. For example, in ST-CGAN [4], Wang et al. respectively used 2 generative adversarial networks for detecting and removing shadow areas of images. Further, Cycle-Consistent Adversarial Networks (cycleGAN) [5], a network which could learn mappings between different domains from unpaired data [6], was also applied to remove shadow areas. In Mask-ShadowGAN7], Hu et al. used shadow masks to guide the training of Cycle-Consistent Adversarial Networks, and used the network to remove the shadow areas of images . All the above methods based on generative adversarial network need to detect shadow of input images firstly, generate the shapes or masks of shadow, and then adopt neural networks to remove the shadow. Most of the methods have tried to improve the accuracy of shadow masks as much as possible. However, those methods are still difficult to obtain accurate shadow masks in some images, such as images with complex shadow edges or images with a large number of tiny shadow areas. Inaccurate shadow masks can finally cause the network to struggle to perform as expected when removing the edges of shadows or discrete small shadow areas.

In this thesis, to solve the problem, the author puts forward a shadow removal framework named IPSM-GAN based on cycleGAN and Mask-ShadowGAN. The framework can learn from shadow masks generated by a mix shadow generating method and remove the shadow. The mix shadow mask generating method produces masks in two methods. When generating a shadow mask, a global thresholding method (Otsu's algorithm) [8] is used for obtaining the main body part of the shadow areas to generate the first shadow mask. Then, a adaptive thresholding method [9] is used for capturing more details at the edges of the shadow or some small shadow areas to generate another shadow mask. After that, the two masks are blended to generate a new shadow mask, which will guide the network learn more information in the shadow and obtain better performance in subsequent shadow removing.

2 Related Work

In 2014, Khan et al. proposed a method, which can detect shadow by using convolutional neural networks and adopt a Bayesian model to remove the shadow [10]. Later, Methods have since emerged using unsupervised generative adversarial networks to achieve the same goals. For Instance, as Wang et al. proposed, two generative adversarial network were respectively trained for detecting and removing shadows. In 2017, a network named Cycle-Consistent Adversarial Networks, i.e. CycleGAN emerged. To learn mappings between different domains, CycleGAN used two GANs to formulate cycle-consistency constraint. Based on CycleGAN, Hu et al. used unpaired dataset and shadow masks generated from shadow image to guide the network to learn the mappings between different styles of images, and achieved the shadow removal in images.

All of the above methods required identifying the shadow areas in images or getting the masks of shadow areas at first. Failure to accurately acquire the shadow area in the image will result in poor performance of shadow removal. In most cases, the central part of the shadow can be obtained very accurately by binarizing the image with a global threshold. However, in the areas where the shadow and bright regions intersect, the shadow edges can not be captured by the global thresholding algorithm because the gray values of the edges are close to other areas. Therefore, in the IPSM-GAN, the author proposes that the using of two methods to generate shadow mask by calculating the global threshold and the adaptive thresholds. The results of the two methods are added together to acquire the accurate shadow mask.

3 Method

As shown in Figure 1, the whole framework of IPSM-GAN is composed of three kinds of module, which are generator, discriminator and mask generator. Figure 1 illustrated the learning steps of IPSM-GAN. The framework can also be devided into two main parts, the first part learns from images with shadow. Firstly, the shadow image $\mathrm{P_1}$ is the input of mask generator $\mathrm{G_m}$ and generator $\mathrm{G_a}$. Two generators will generate a shadow mask $\mathrm{M_1}$ and a unshaded image $\mathrm{P_2}$ of the source image, respectively. Then , the mask $\mathrm{M_1}$ and image $\mathrm{P_2}$ were sent to generator $\mathrm{G_b}$ to generate a rebuilt shadow $\mathrm{P_3}$. The mask $\mathrm{M_1}$ will be stored in a first-in-first-out queue and used in the learning from shadow-free images. Another part of the framework learns from shadow-free images. A real unshaded image $\mathrm{P_4}$ and a shadow mask $\mathrm{M_r}$ randomly selected from the mask queue are fed into the generator $\mathrm{G_b}$ and a shadow image $\mathrm{P_5}$ is produced. Then, image $\mathrm{P_5}$ will be recover to a shadow-free image $\mathrm{P_6}$ by generator $\mathrm{G_a}$.

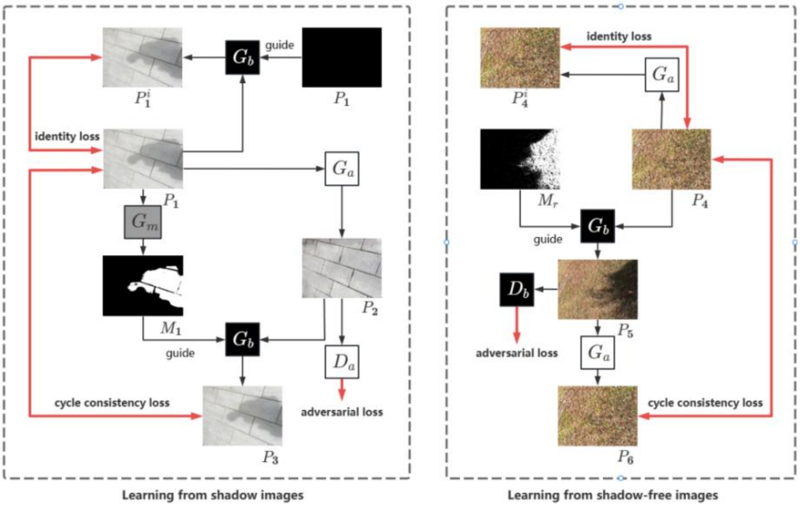

Fig. 1. The framework of IPSM-GAN images (Photo/Picture credit: Original)

3.1 Generators and Discriminators

In the framework, the generator $\mathrm{G_a}$ can convert a shadow image to a unshaded image, and the structure of this module is derived from the generative network designed by Johnson11]. The input of the generator $\mathrm{G_a}$ are 3 channels of the shadow image. To extract features, the image goes through a three-layer convolutional neural network at first. Subsequently, there is a nine-layer residual network in the generator G_a, which transforms the style of the image (from shadow to shadow-free). Then, with 2 layers of deconvolution and one layer of convolution operation, the image is rebuilt to its original size. The structure of the generator G_b is almost the same as the structure of G_a. The difference between the two generators is that the input of Generator G_b needs 4 channels, and the additional channel is a shadow mask randomly selected from the mask queue. Shadow image discriminator D_b and shadow-free image discriminator D_a are both based on PatchGAN12]. They share the same network structure. With a 5-layer convolutional neural network, the discriminators can distinguish if the input image is real or not.

3.2 Mask Generator

As mentioned in the previous sections, no matter in the learning from shadow or shadow-free images, the generator needs to obtain an accurate shadow mask firstly. In this thesis, the network is trained by using unpaired datasets, which make it difficult to get shadow masks directly. In Mask-ShadowGAN, Hu et al. used the OTSU algorithm to acquire the global threshold of the image, which enable the generation of binary shadow mask of the shaded image. This method has some limitations. First of all, for some high-frequency areas like shadow edges, a mask generated by using global threshold can not include these areas, and a lot of details of shadow are lost. Secondly, after several iterations of the model, artifacts will appear on the edge of removed shadows, which means that the removal performance on the shadow edges is not as good as the center areas. Therefore, the author proposes an improved shadow mask generation method.

Figure 2 shows how the mask generator produces a shadow mask. Firstly, the generator G_m calculates the difference between the original shadow image P_1 and the generated unshaded image P_2. Then, Chow-Kaneko adaptive threshold algorithm is adopted to compare the value of each pixel with the average value of the neighborhood. If the value of the pixel exceeds the average value, the pixel is marked as 1, that is, the shadow, otherwise it is marked as 0. G_m generates the mask M_a by this way. At the same time, OTSU algorithm is used for calculating the global threshold t of the image, and t is key to binarize the image and generate mask M_o. The mask M_a and M_o are then added to generate the shadow mask M_1 of the image P_1. The mixed shadow mask can capture more details on the edge of shadows or some smaller shadows, which means the accuracy of the mask-generation is improved.

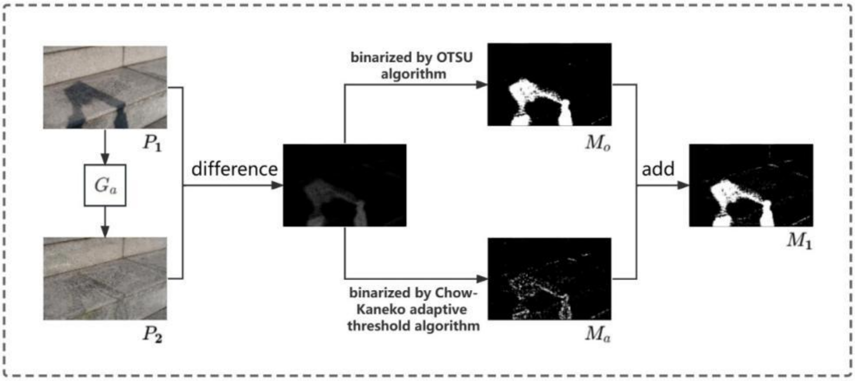

Fig. 2. The generating steps of shadow mask (Photo/Picture credit: Original)

3.3 Loss Function

The loss of IPSM-GAN can be devided into three parts, which are identity loss13], adversarial loss and cycle consistency loss. During training, the framework adopted adversarial loss to optimize the generators and the discriminators of two kinds of image. When the framework learns from shadow images, as shown in Figure 1, the generator G_a uses the shadow image P_1 to generate the unshaded image P_2. Then, the discriminator D_a distinguishes whether the image P_2 is a real unshaded image:

$$D_a(G_a(P_1))) = \text{ Is the shadow-free image real or fake?} \quad (1)$$

The generator G_a and the discriminator D_a are optimized by using the following objective functions:

$$\mathrm{Loss}_{GAN}^{a}(G_a, D_a) = \varepsilon_{P_{4\sim P_{data}}(P_4)}[\log(D_a(P_4))]$$
$$+ \varepsilon_{P_{1\sim P_{data}}(P_1)}[\log(1 - D_a(G_a(P_1)))] \quad (2)$$

In the formula, ε represents the error. $P_{4\sim P_{data}}(P_4)$ and $P_{1\sim P_{data}}(P_1)$ denote that real unshaded image P_4 and the shadow image P_1, following the data distribution P_{data}, are respectively selected from the datasets of unshaded images and images with shadow. Adversarial loss is calculated by the Smooth L1 loss function [14], rather than L2 loss, which can reduces the influence from outliers.

Similarly, when it is learning from images have no shadow, the generator G_b can converts the unshaded image P_4 to a shadow image P_5 by randomly acquiring the mask M_r from the shadow mask queue. Then, the discriminator G_b will determine if the image P_5 is a real shadow image:

$$D_b(G_b(P_4, M_r))) = \text{Is the shadow image real or fake?} \quad (3)$$

To make the generator G_b and discriminator D_b be further optimized, adversarial loss is also applied:

$$Loss^{b}_{GAN}(G_b, D_b) = \varepsilon_{P_{b \sim P_{data}}(P_b)}[\log(D_b(P_b))] + \varepsilon_{P_{a \sim P_{data}}(P_a)}[\log(1 - D_b(G_b(P_a, M_r)))] \quad (4)$$

In the learning of the two kinds of images, the produced images will be rebuilt to their original style. The generator G_b uses the mask M_1 to convert the shaded image P_2 to the shadow image P_3:

$$P_3 = G_b(P_2, M_1) \quad (5)$$

Where the mask M_1, generated by using the mask generator G_m, is a binary mapping of the shadow areas of the image P_1:

$$M_1 = G_m(P_1) \quad (6)$$

The author expect the recovered shadow image P_3 to be as similar as possible to the original shadow image P_1, and cycle-consistency constraint are used to optimize the generators G_a and G_b.

$$Loss^{a}_{cycle}(G_a, G_b) = \varepsilon_{P_{1 \sim P_{data}}(P_1)}[||G_b(G_a(P_1), M_1) - P_1||_1] \quad (7)$$

Similarly, in another part of the framework, the generator G_a converts the image P_5 to the unshaded image P_6. The author also uses cycle consistency loss to optimize two generators:

$$Loss^{b}_{cycle}(G_b, G_a) = \varepsilon_{P_{4 \sim P_{data}}(P_4)}[||G_a(G_b(P_4, M_r)) - P_4||_1] \quad (8)$$

To further optimize the shadow image generator G_b, a shadow image P_1 is transformed by using a null mask M_n:

$$P^{i}_{1} = G_b(P_1, M_n) \quad (9)$$

Moreover, identity loss is applied to regularizes the output. It makes the output image P^{i}_{1} as close as possible to the input image:

$$\mathrm{Loss}^{a}_{identity}(G_b) = \varepsilon_{P_{1\sim P_{data}}(P_1)}[\|G_b(G_1, M_n) - P_1\|_1] \quad (10)$$

Analogously, the author use the real unshaded image P_4 as the input of the generator G_a, which can also be optimized by using identity loss:

$$\mathrm{Loss}^{b}_{identity}(G_a) = \varepsilon_{P_{4\sim P_{data}}(P_4)}[\|G_a(G_4) - P_4\|_1] \quad (11)$$

Totally, the loss of IPSM-GAN is a weighted sum of the three kinds of loss in two learning. The weights of the three loss functions, ω_1, ω_2, and ω_3, are set to 2, 10, and 5 respectively.

$$\mathrm{Loss}_{total} = \omega_1\left(\mathrm{Loss}^{a}_{GAN} + \mathrm{Loss}^{b}_{GAN}\right) + \omega_2\left(\mathrm{Loss}^{a}_{cycle}(G_a, G_b) + \mathrm{Loss}^{b}_{cycle}(G_b, G_a)\right) + \omega_3(\mathrm{Loss}^{a}_{identity}(G_b) + \mathrm{Loss}^{b}_{identity}(G_a)) \quad (12)$$

4 Experiment

4.1 Dataset

In this thesis, the USR dataset used by the author consists of unpaired images. There are 2,445 shadow images with shadow and 1,770 images in the dataset. The images have many kinds of scenes, which contain shadows generated by objects of different kinds, such as plants, people, buildings, traffic signs, fences, umbrellas, etc. During the experiment, shadow images were divided into two groups, with 1956 images randomly selected for training and the remaining 489 images applied to subsequent test. All the images with no shadow were used in training.

4.2 Setting of Parameters

In the IPSM-GAN, the author uses random noise which follows zero-mean normal distribution to initialize the parameters in the generator and discriminator. In order to avoid the discriminators learning too fast, resulting in poor training performance of the generators, the learning rate of discriminators is set lower than the learning rate of generators [15,16]. The initial learning rates of the generators and discriminators are set to 0.0002 and 0.00015, respectively. Their learning rates remain the same value for the first 100 epochs. In the next 100 epochs, the learning rates are linearly decayed to 0, after which the learning will be stopped. During the learning, the framework is optimized by using Adam17]. The first and second momentum values are empirically set as 0.5 and 0.999.

4.3 Results and Comparison

Here, the author compares IPSM-GAN with CycleGAN and Mask-ShadowGAN, the results of them were achieved by running their open source code. The images used for training and testing are from the USR dataset. The visual comparison of the three methods are shown in Figure 3. In some specific situations, for example, when the background texture near the shadow areas is complex (the first row and the fourth row), IPSM-GAN can more accurately remove the shadow in the image. The results illustrate that IPSM-GAN can effectively distinguish the shadow from the background part in images, so as to achieve better learning results in the same dataset of training. Moreover, by using the mixed shadow mask, IPSM-GAN can learn and recognize the edge area of the shadow more accurately. In images transformed by IPSM-GAN, the area between the edges of the shadow areas and the background look more smooth than those shadows removed by using other methods, resulting in better visual effects (the second, third, and fifth row).

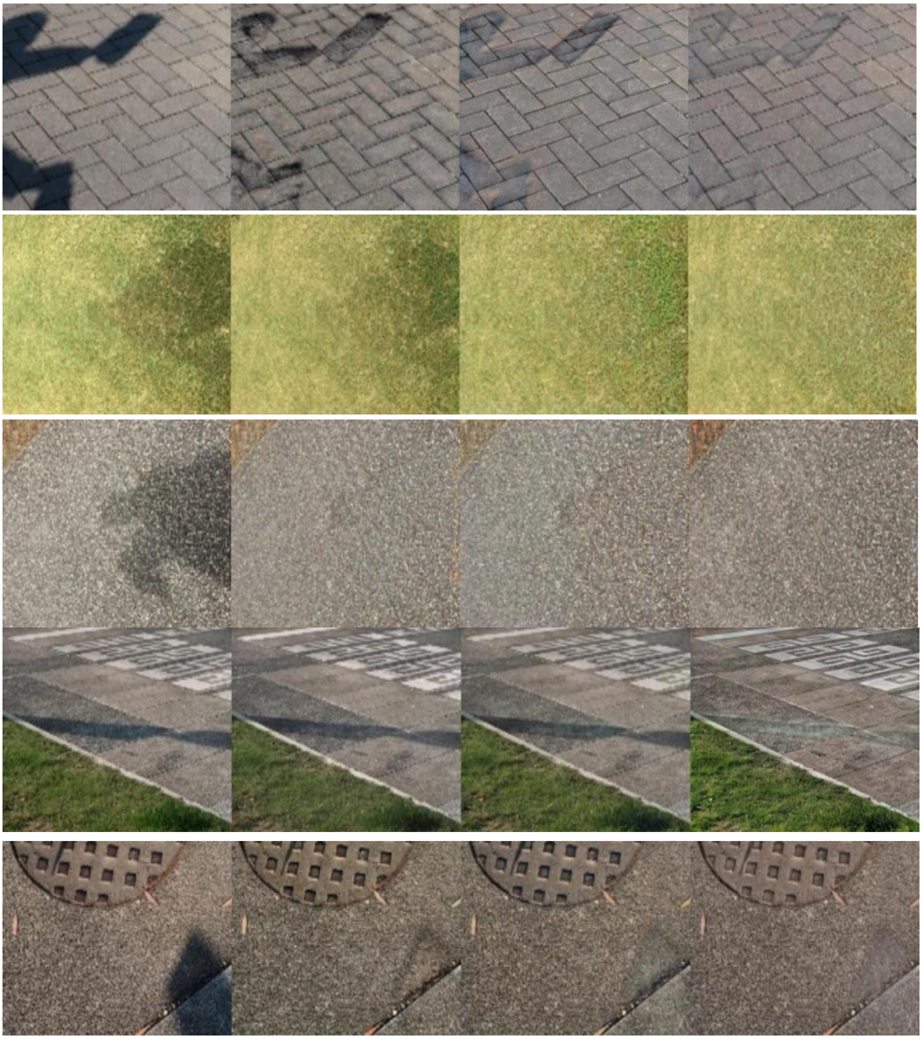

Fig. 3. Original image and results from three different removal methods(Photo/Picture credit: Original)

5 Conclusion

In this thesis, the author proposes a generative adversarial network IPSM-GAN, guided by mixed shadow mask to remove the shadow of images. The key idea is to use adaptive threshold and global threshold to obtain a accurate shadow mask, so that IPSM-GAN can more efficiently learn how to map images with shadow to those images have no shadow, and improve the ability of removing shadows. The experimental results reveal IPSM-GAN's impressive performance in shadow removal, particularly for images with uncomplicated background textures. Nevertheless, challenges arise when dealing with images featuring intricate background textures or challenging shadow patterns. Addressing these challenges demands further exploration and development of techniques to achieve seamless style transformation between shadowed and unshaded images. To address these limitations, the author outlines future directions. One pivotal focus is on dataset expansion, encompassing a wider variety of scenarios and textures. This expanded dataset aims to bolster IPSM-GAN's ability to generalize and adapt to diverse image characteristics. Moreover, the author emphasizes the significance of training the network using high-quality datasets, fostering improved generalization and enhancing IPSM-GAN's overall performance. In summary, the proposed IPSM-GAN framework presents a promising solution for shadow removal, leveraging adaptive thresholding and guided learning. While demonstrating its efficacy on simpler shadowed images, future efforts will center on refining its performance for complex scenarios through dataset augmentation and network enhancement.

References

1. Canny J, A computational approach to edge detection[J], IEEE Transactions on pattern analysis and machine intelligence, 1986 (6): 679-698.
2. I. Sobel and G. Feldman. A 3x3 Isotropic Gradient Operator for Image Processing. Pattern Classification and Scene Analysis, 1973, pp. 271-272.
3. Salman Hameed Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Automatic feature learning for robust shadow detection. In CVPR, pages 1939–1946, 2014.
4. Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In CVPR, pages 1788–1797, 2018.
5. Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In ICCV, pages 2223–2232, 2017.
6. Amjad Almahairi, Sai Rajeshwar, Alessandro Sordoni, Philip Bachman, and Aaron C. Courville. Augmented CycleGAN: Learning many-to-many mappings from unpaired data. In ICML, pages 195–204, 2018.

7. Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, Pheng-Ann Heng. Mask-ShadowGAN: Learning to Remove Shadows From Unpaired Data. In ICCV, pages 2472–2481, 2019.
8. Nobuyuki Otsu. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1):62–66, 1979.
9. C.K. Chow and T. Kaneko. Boundary Detection of Radiographic Images by a Thresholding Method. Frontiers of Pattern Recognition, S. Watanabe, ed., Academic Press, 1972, pp. 61-82.
10. Salman Hameed Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Automatic shadow detection and removal from a single image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(3):431–446, 2016.
11. Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, pages 694–711, 2016.
12. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In CVPR, pages 1125–1134, 2017.
13. Yaniv Taigman, Adam Polyak, and Lior Wolf. Unsupervised cross-domain image generation. arXiv preprint arXiv:1611.02200, 2016.
14. Ross Girshick. Fast R-CNN. In ICCV, pages 1440-1448, 2015.
15. L. Mescheder, A. Geiger, and S. Nowozin. Which training methods for GANs do actually converge? arXiv preprint arXiv:1801.04406, 2018
16. L. Mescheder. On the convergence properties of GAN training. arXiv preprint arXiv:1801.04406, 2018.
17. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.