

Domain Adaptation for Synthesis of Hazy Images

Haoying Sun, Yutong Zheng, Qing Lang

School of Information Engineering, Minzu University of China, Beijing, China

Email: zhengyutong68@aliyun.com

How to cite this paper: Sun, H.Y., Zheng, Y.T. and Lang, Q. (2021) Domain Adaptation for Synthesis of Hazy Images. *Journal of Computer and Communications*, 9, 142-151.

<https://doi.org/10.4236/jcc.2021.910010>

Received: October 16, 2021

Accepted: October 26, 2021

Published: October 29, 2021

Abstract

Most existing image dehazing methods based learning are less able to perform well to real hazy images. An important reason is that they are trained on synthetic hazy images whose distribution is different from real hazy images. To relieve this issue, this paper proposes a new hazy scene generation model based on domain adaptation, which uses a variational autoencoder to encode the synthetic hazy image pairs and the real hazy images into the latent space to adapt. The synthetic hazy image pairs guide the model to learn the mapping of clear images to hazy images, the real hazy images are used to adapt the synthetic hazy images' latent space to real hazy images through generative adversarial loss, so as to make the generative hazy images' distribution as close to the real hazy images' distribution as possible. By comparing the results of the model with traditional physical scattering models and Adobe Lightroom CC software, the hazy images generated in this paper is more realistic. Our end-to-end domain adaptation model is also very convenient to synthesize hazy images without depth map. Using traditional method to dehaze the synthetic hazy images generated by this paper, both SSIM and PSNR have been improved, proved that the effectiveness of our method. The non-reference haze density evaluation algorithm and other quantitative evaluation also illustrate the advantages of our method in synthetic hazy images.

Keywords

Haze Scene, Domain Adaptation, Generative Adversarial Net, Variational Autoencoder, Deep Learning

1. Introduction

Single image dehazing is an important and difficult task in computer vision, so people usually only pay attention to how to dehaze, and seldom pay attention to

how to synthesize hazy images and how to synthesize more realistic hazy images. This will cause some problems today with the rapid development of deep learning. Having a large amount of high-quality data can make the deep learning model perform better, however, most of the datasets in single image dehazing based learning are artificially synthesized. These synthetic datasets cannot fully simulate the hazy images in the real world. Therefore, it is difficult for the model based learning using such data sets to perform well on the real hazy images.

Currently, there are three main methods for hazy image synthesis. The first is synthesis based on the physical scattering models [1] [2]. It need to measure or estimate the depth map corresponding to the image, and then synthesize the hazy images [3] [4] with different parameters of the physical scattering model. Because the depth map inevitably has errors, and the physical scattering model is difficult to fully simulate the real environment, the synthesized hazy images have certain defects. The second is synthesis based learning, which uses paired data to train to generate hazy images. Since this method also requires large amounts of paired data which is mostly synthetic, there may be deviation between the generated hazy images and the real. The third is some others, such as synthetic software, rendering and so on. However, the quality of the hazy images generated by these methods [5] is low, or these hazy images synthesized from virtual environment [6], and there may be still a deviation from the real hazy images.

In order to alleviate the above problems, this paper proposes a domain adaptive hazy image synthesis model. The model is based on variational autoencoder (VAE) and generative adversarial network (GAN), which adapts the synthetic hazy images to the real hazy images by generative adversarial loss, thereby generating a more realistic hazy images.

The main contributions of this paper are as follows:

- 1) Propose a new domain adaptive model to generate a more realistic hazy images which have various densities.
- 2) Propose to use generative adversarial loss to adapt to the real hazy images' domain.

2. Related Works

2.1. Atmospheric Scattering Model

Haze is produced by the scattering of fine substances suspended in the atmosphere, which will reduce the quality of the image and affect subsequent analysis. In computer vision and computer graphics, the following equation is widely used to describe the generation of a hazy image [1] [2].

$$I(x) = J(x)t(x) + A(1-t(x)) \quad (1)$$

where I is the hazy image, J is the scene radiance representing the haze-free image, A is the atmospheric light, t is the medium transmission, β is the scattering coefficient of the atmosphere and d is the depth of scene, t can be calculated with the following equation.

$$t(x) = e^{-\beta d(x)} \quad (2)$$

Image dehazing can restore the lost details of the image and facilitate in-depth analysis. Therefore, this task has always been paid attention in computer vision and progressing rapidly. Traditional methods in single image dehazing based on physical priors [7] [8] do not require paired data to train, but often have certain requirements on the environment to perform well. Methods based learning [9] [10] require a lot of data for training, because it is difficult to obtain the paired images, the hazy images in the training set are usually artificially synthesized by the atmospheric model. There is a certain deviation between the synthesized hazy images and the real hazy images, which causes some models to perform poorly on the real hazy images.

2.2. Hazy Image Synthesis

2.2.1. Traditional Methods

The traditional method of hazy image synthesis mainly adopts Equation (1) and Equation (2), and the synthesis process is shown in **Figure 1**. The available datasets for indoor hazy image synthesis include NYUv2 dataset [11], Middlebury Stereo [12], etc. The source data of NYUv2 is obtained through Microsoft Kinect, and there are 1499 pictures after the filling of the depth missing value and the post-processing of marking. In general, these data can be used for tasks such as depth estimation and semantic segmentation. The available data for outdoor hazy image synthesis include HazeRD dataset [4], Cityscapes [13], KITTI [14], *et al.*

Adobe Lightroom CC can synthesize hazy images, it contains a dehazing function that can make the picture clear. When the value is adjusted to a negative value (-70 in this paper), a hazy image will be generated. The dehazing function of this software is used to synthesize hazy images for control experiments. Hazy images can also be generated in some 3D games or rendering software.

2.2.2. Deep Learning Method

The generative adversarial network (GAN) [15] has been receiving widespread attention and developing rapidly [16] [17] [18], it can be used to synthesize hazy

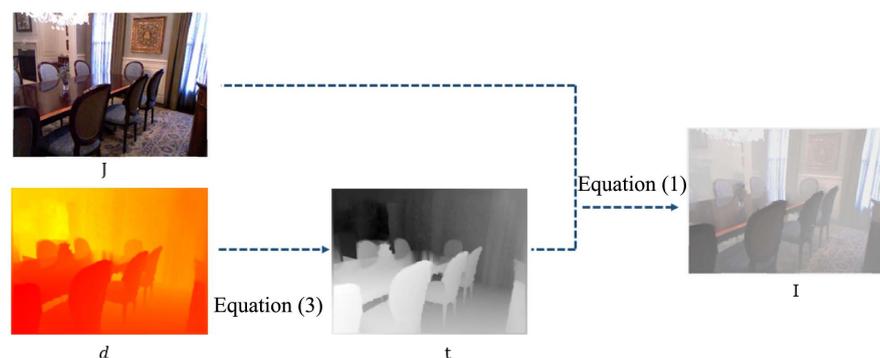


Figure 1. Haze image synthesis based on atmospheric scattering model.

images. The image translation model pix2pix [16] is composed of a multi-resolution generator and a multi-scale discriminator. The discriminator which uses the Patch GAN structure can perform multi-scale discrimination on the generated result. The basic structure of the generator is U-Net [19], it can generate high-resolution and clear images. But this method is supervised learning, if used for synthesizing hazy images, the training data that can be used is artificially synthesized hazy images, there may be a certain difference between the generated result and the real hazy image.

The well-known unsupervised image conversion method Cycle-GAN [17] has been proposed to provide an effective loss function named cycle consistency loss for unpaired training data, and the real hazy images can be used for synthesizing hazy images. But this method without absolute mapping will cause the result to appear undesired conversion style.

3. Method

3.1. Domain Adaptation and Image to Image Translation

Domain adaptation aims to reduce the differences between different domains, and it is mainly divided into feature level or pixel level. The feature-level adaptive method aims to align the feature distribution between the source domain and the target domain by minimizing the maximum mean difference (MMD) or applying adversarial learning strategies on the feature space [20]. Another study focused on pixel-level adaptation [21] [22]. These methods deal with the domain transfer problem by applying image-to-image translation.

Image-to-image translation [16] [23] refers to the translation of images from one domain to the other through learning. Image translation of supervised learning requires paired data sets, and the corresponding image translation algorithms include pix2pix [16] model, pix2pixHD [23] model, etc. Unsupervised learning image translation [17] [24] does not require paired datasets.

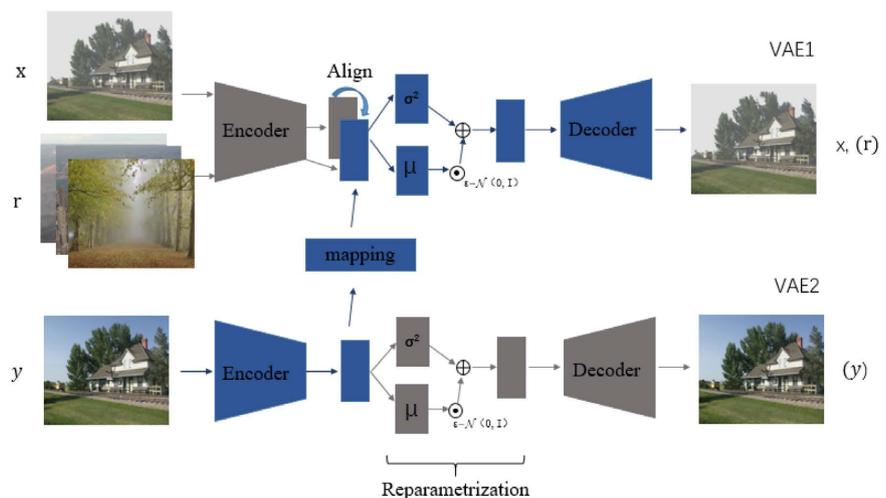


Figure 2. Overview of the proposed domain adaptation for synthesis of hazy images network.

3.2. Synthetic Hazy Images

The overview model structure is shown in **Figure 2**. It is trained on paired data and real hazy images, paired data contains synthetic hazy images and clear images.

The synthetic hazy images domain and the real hazy image domain are respectively recorded as D_X, D_R . The clear image domain corresponding to the synthesized fog image domain is denoted as D_Y . Synthetic hazy images, real hazy images, and clear images are recorded as $x \in D_X, r \in D_R, y \in D_Y$. The latent space of x, y, r is recorded as Z_x, Z_y, Z_r . VAE₁ which is trained on x, r includes encoder E_{D_X}, E_{D_R} and generator G_{D_X}, G_{D_R} . VAE₂ includes encoder E_{D_Y} and decoder G_{D_Y} . Inspired by [25], our model first encodes x and r into the latent space, and then adapts their domains; there is a big difference between the latent space Z_x and Z_y , we use mapping net (denote it as T_Z) to connect them.

The VAE1 losses for the real hazy images (r) are as follows:

$$L_{VAE_1}(r) = KL(E_{D_R, D_X}(Z_r \parallel \mathcal{N}(0, I))) + \alpha \mathbb{E}_{Z_r \sim E_{D_R, D_X}}(Z_r | r) [\|G_{D_R, D_X}(r_{D_R \rightarrow D_R} | Z_r) - r\|_1] + L_{VAE_1, GAN}(r) \quad (4)$$

$z_r \in Z_{D_R}$ is the Hidden code of real hazy images (r), $r_{D_R} \rightarrow D_R$ is generated by G_{D_R} , KL stands for KL divergence. The second term is L1 loss, in order to reconstruct the image. The third term is the least-square loss [26], to make the reconstructed image realistic.

To adapt the domain from synthetic hazy images to realistic hazy images, we use adversarial loss as follows.

$$L_{VAE_1, GAN}^{latent}(r, x) = \mathbb{E}_{x \sim D_X} [1 - D_{D_R, D_X}(E_{D_R, D_X}(x))^2] + \mathbb{E}_{r \sim D_R} [D_{D_R, D_X}(E_{D_R, D_X}(r))^2] \quad (5)$$

So the total loss of VAE1 is as follows.

$$\min_{E_{D_R, D_X}, G_{D_R, D_X}} \max_{D_{D_R, D_X}} L_{VAE_1}(r) + L_{VAE_1}(x) + L_{VAE_1, GAN}^{latent}(r, x) \quad (6)$$

The mapping net learns the connection between the VAE1 and VAE2, it's loss is as follows.

$$L_{T_Z}(x, y) = \delta_1 L_{T_Z} + L_{T_Z, GAN} + \delta_2 L_{perceptual} \quad (7)$$

where L_{T_Z} is stand for the L1 loss between $T_Z(Z_x)$ mapped by T_Z and Z_y ,

$$L_{T_Z} = \mathbb{E} \|T_Z(Z_x) - Z_y\|_1 \quad (8)$$

$L_{T_Z, GAN}$ is the least-square loss(LSGAN), $L_{perceptual}$ is perceptual loss, as shown in Equation (9).

$$L_{perceptual} = \mathbb{E} \left[\sum_i \frac{1}{n_{D_T}^i} \|\phi_{D_T}^i(x_{D_X \rightarrow D_Y}) - \phi_{D_T}^i(y_{D_Y \rightarrow D_Y})\|_1 \right] + \sum_i \frac{1}{n_{VGG}^i} \|\phi_{VGG}^i(x_{D_X \rightarrow D_Y}) - \phi_{VGG}^i(y_{D_Y \rightarrow D_Y})\|_1 \quad (9)$$

where $\phi_{VGG}^i(x)$ is the i^{th} layer feature map of the discriminator (VGG network), and n_{VGG}^i indicates the number of activations in that layer. D_r is the same as VGG.

4. Experiment

4.1. Implementation

The experiment environment uses the deep learning framework Pytorch1.6.0. The graphics card is NVIDIA RTX 3090Ti, the graphics memory is 32 GB.

The 8970 pairs of images selected from the RESIDE-OTS [3] and 2864 real hazy images selected from RESIDE-beta [29] are used to train our model. We choose some images from HazeRD [4], SOTS (indoor and outdoor) to test the model's ability on synthesizing hazy images.

4.2. Comparisons

We compare the model proposed qualitatively and quantitatively in this part. AuthESI [27], PSNR and SSIM is used to evaluate the quality of images. Density [28] of images is also measured.

As shown in **Figure 3**, some photos selected from the SOTS are used to synthesize hazy images. It can be seen from the second row of the figure that in some cases, the overall color shift will occur on the hazy images synthesized by Adobe Lightroom CC. And our model doesn't have this kind of problem and generates hazy images successfully.

Table 1 is a quantitative test of the corresponding image of **Figure 3**. It can be

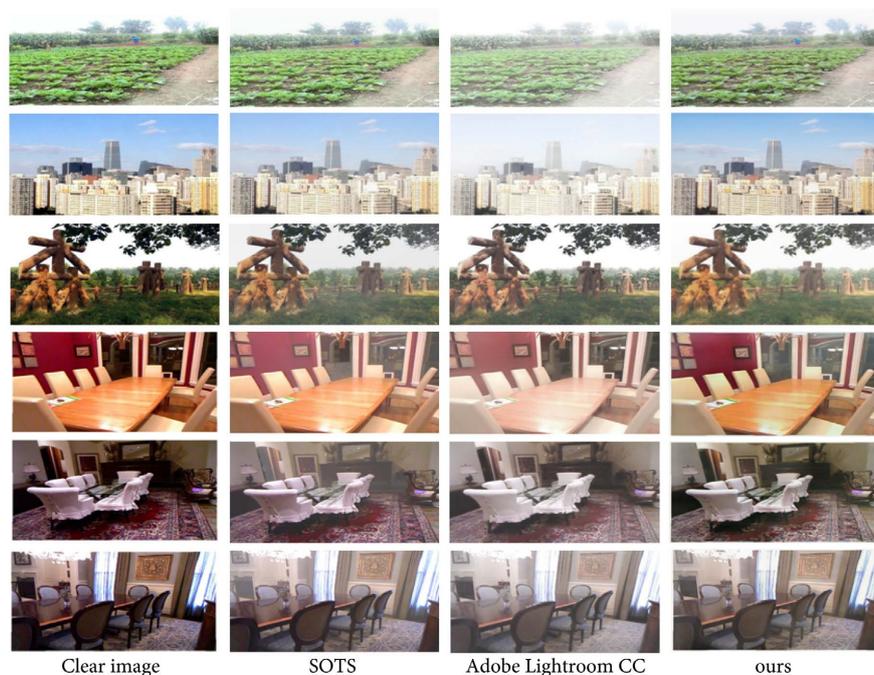


Figure 3. Qualitative comparison of Adobe Lightroom CC and our method on the SOTS outdoor testset.

seen from the table that the quality of the hazy images synthesized by our model is quite high, and this is in the case of similar hazy images' density.

As shown in **Figure 4**, these images are from HazeRD dataset, and some are synthesized by Lightroom CC and our method. It can be seen from the second column of the figure that there are some abnormalities around the plant, because the nearby depth has been wrongly measured, and our model performs well.

The quantitative results of **Figure 3** are shown in **Table 2**. Similarly, our method performs well, indicating that our method also has a certain degree of generalization on various types of images.

To further improve the effective of our model, we use the DCP [7] to dehaze the iamges and measure the PSNR and SSIM. Since the DCP [7] is invalid for the sky, we choose the indoor pictures in **Figure 3** for testing. First, test the PSNR and SSIM of the hazy image synthesized by our model, and then use DCP [7] to dehaze and then test PSNR and SSIM again. The results are as shown in **Table 3**. It can be seen that after DCP [7] processing our synthetic hazy images' SSIM and PSNR are improved, which proves the effectiveness of our model.

Running our model repeatedly can obtain hazy images with different densities, as shown in **Figure 5**.

Table 1. Compare our method with Adobe Lightroom CC software on the SOTS outdoor dataset.

clear image		SOTS outdoor		Lightroom CC		ours	
AuthESI	density	AuthESI	density	AuthESI	density	AuthESI	density
3.422	0.578	1.585	0.739	0.837	1.225	0.868	0.911
5.314	0.649	2.593	1.189	0.303	2.279	1.435	0.982
2.348	0.292	1.289	0.650	1.075	0.637	1.047	0.700
4.303	0.333	2.362	0.658	1.886	1.221	2.367	0.609
6.716	0.241	2.236	0.666	1.982	0.894	4.376	0.479
6.985	0.306	3.525	0.722	2.008	1.095	6.884	0.716

Table 2. Compare our method with Adobe Lightroom CC software on the HazeRD dataset.

clear image		HazeRD		Lightroom CC		ours	
AuthESI	density	AuthESI	density	AuthESI	density	AuthESI	density
4.113	0.194	1.712	0.754	1.693	0.589	1.685	0.725
6.378	0.445	2.084	0.585	1.360	1.485	4.211	0.579
4.879	0.481	1.209	1.138	0.154	1.738	1.260	0.695
2.764	0.611	1.415	0.812	0.685	1.719	0.515	0.939

Table 3. Changes of SSIM and PSNR before and after using DCP to remove haze.

Synthesizes Hazy images by ours model		DCP	
PSNR	SSIM	PSNR	SSIM
16.336213	0.772629	19.863129	0.850797
17.996079	0.752121	24.859871	0.864840
15.900667	0.751683	21.645540	0.893571



Figure 4. Qualitative comparison of Adobe Lightroom and our method on the HazRD dataset.

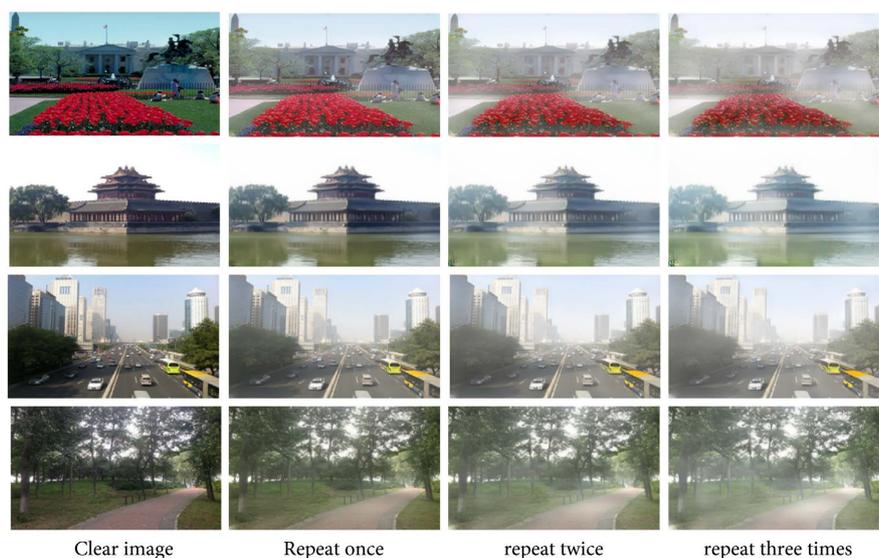


Figure 5. Hazy images of different densities generated by our model.

5. Conclusion

In order to alleviate the problem that the dehazing dataset deviates from reality and deep learning image hazing methods usually perform poorly on real hazy images, we propose a model for synthesizing hazy image based on domain adaptation, which can generate hazy images without depth map. Both qualitative and quantitative experiments show the effectiveness of our method.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Narasimhan, S.G. and Nayar, S.K. (2000) Chromatic Framework for Vision in Bad Weather. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Piscataway, 598-605. <https://doi.org/10.1109/CVPR.2000.855874>
- [2] Narasimhan, S.G. and Nayar, S.K. (2002) Vision and the Atmosphere. *International Journal of Computer Vision*, **48**, 233-254. <https://doi.org/10.1023/A:1016328200723>
- [3] Li, B., Ren, W., Fu, D., Dan, F., Zeng, W. and Wang, Z. (2017) Reside: A Benchmark for Single Image Dehazing. arXiv preprint arXiv:1712.04143. <https://arxiv.org/pdf/1712.04143v1.pdf>
- [4] Zhang, Y., Li, D. and Sharma, G. (2017) Hazerd: An Outdoor Scene Dataset and Benchmark for Single Image Dehazing. *IEEE International Conference on Image Processing (ICIP)*, Piscataway, 3205-3209. <https://doi.org/10.1109/ICIP.2017.8296874>
- [5] Xiao, J., Shen, M., Lei, J., Xiong, W. and Jiao, C. (2020) Image Conversion Algorithm for Haze Scene Based on Generative Adversarial Networks. *Chinese Journal of Computers*, **43**, 165-176.
- [6] Tarel, J.P., Hautiere, N., Caraffa, L., Cord, A., Hakmaoui, H. and Gruyer, D. (2012) Vision Enhancement in Homogeneous and Heterogeneous Fog. *IEEE Intelligent Transportation Systems Magazine*, **4**, 6-20. <https://doi.org/10.1109/MITS.2012.2189969>
- [7] He, K.M., Sun, J. and Tang, X. (2010) Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**, 2341-2353. <https://doi.org/10.1109/TPAMI.2010.168>
- [8] Fattal, R. (2018) Single Image Dehazing. *ACM Transactions on Graphics*, **27**, 1-9. <https://doi.org/10.1145/1360612.1360671>
- [9] Xu, Q., Wang, Z., Bai, Y., Xie, X. and Jia, H. (2020) FFA-Net: Feature Fusion Attention Network for Single Image Dehazing. *Proceedings of the AAAI Conference on Artificial Intelligence*, Menlo Park, 11908-11915. <https://doi.org/10.1609/aaai.v34i07.6865>
- [10] Liu, X., Ma, Y., Shi, Z. and Chen, J. (2019) GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, 7313-7322. <https://doi.org/10.1109/ICCV.2019.00741>
- [11] Silberman, N., Hoiem, D., Kohli, P. and Fergus, R. (2012) Indoor Segmentation and Support Inference from RgbD Images. *European Conference on Computer Vision*, Springer, Berlin, 746-760. https://doi.org/10.1007/978-3-642-33715-4_54
- [12] Scharstein, D. and Szeliski, R. (2002) A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, **47**, 7-42. <https://doi.org/10.1023/A:1014573219977>
- [13] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B. (2016) The Cityscapes Dataset for Semantic Urban Scene Understanding. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR.2016.350>
- [14] Geiger, A., Lenz, P., Stiller, C. and Urtsun, R. (2013) Vision Meets Robotics: The Kitti Dataset. *The International Journal of Robotics Research*, **32**, 1231-1237. <https://doi.org/10.1177/0278364913491297>
- [15] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2020) Generative Adversarial Networks. *Communications of the ACM*, **63**, 139-144. <https://doi.org/10.1145/3422622>

- [16] Isola, P., Zhu, J., Zhou, T. and Efros, A. (2017) Image-to-Image Translation with Conditional Adversarial Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Piscataway, 1125-1134. <https://doi.org/10.1109/CVPR.2017.632>
- [17] Zhu, J., Park, T., Isola, P. and Efros, A. (2017) Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision*, Piscataway, 2223-2232. <https://doi.org/10.1109/ICCV.2017.244>
- [18] Mirza, M. and Osindero, S. (2014) Conditional Generative Adversarial Nets. arXiv preprint arXiv:1411.1784. <https://arxiv.org/pdf/1411.1784.pdf>
- [19] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Berlin, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [20] Tzeng, E., Hoffman, J., Saenko, K. and Darrell, T. (2017) Adversarial Discriminative Domain Adaptation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7167-7176. <https://doi.org/10.1109/CVPR.2017.316>
- [21] Bousmalis, K., Silberman, N., Dohan, D., Erhan, D. and Krishnan, D. (2017) Unsupervised Pixel-Level Domain Adaptation with Generative Adversarial Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3722-3731. <https://doi.org/10.1109/CVPR.2017.18>
- [22] Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W. and Webb, R. (2017) Learning from Simulated and Unsupervised Images through Adversarial Training. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2107-2116. <https://doi.org/10.1109/CVPR.2017.241>
- [23] Wang, T., Liu, M., Zhu, J., Tao, A., Kautz, J. and Catanzaro, B. (2018) High-Resolution Image Synthesis and Semantic Manipulation with Conditional Gans. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Piscataway, 8798-8807. <https://doi.org/10.1109/CVPR.2018.00917>
- [24] Liu, M., Breuel, T. and Kautz, J. (2017) Unsupervised Image-to-Image Translation Networks. *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, 700-708.
- [25] Wan, Z., Zhang, B., Chen, D., Zhang, P., Chen, D., Liao, J. and Wen, F. (2020) Bringing Old Photos Back to Life. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Piscataway, 2747-2757. <https://doi.org/10.1109/CVPR42600.2020.00282>
- [26] Mao, X., Li, Q., Xie, H., Y.K. Lau, R., Wang, Z. and Paul Smolley, S. (2017) Least Squares Generative Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2794-2802. <https://doi.org/10.1109/ICCV.2017.304>
- [27] Zhang, N., Zhang, L. and Cheng, Z. (2017) Towards Simulating Foggy and Hazy Images and Evaluating Their Authenticity. *International Conference on Neural Information Processing*, Springer, Cham, 405-415. https://doi.org/10.1007/978-3-319-70090-8_42
- [28] Choi, L.K., You, J. and Bovik, A.C. (2015) Referenceless Prediction of Perceptual Fog Density and Perceptual Image Defogging. *IEEE Transactions on Image Processing*, **24**, 3888-3901. <https://doi.org/10.1109/TIP.2015.2456502>
- [29] Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W. and Wang, Z. (2018) Benchmarking Single-Image Dehazing and beyond. *IEEE Transactions on Image Processing*, **28**, 492-505. <https://doi.org/10.1109/TIP.2018.2867951>