Learning of Image Dehazing Models for Segmentation Tasks

Sébastien de Blois, Ihsen Hedhli, Christian Gagné

Computer Vision and Systems Laboratory, Université Laval, Québec, QC Canada

sebastien.de-blois.1@ulaval.ca, ihsen.hedhli.1@ulaval.ca, christian.gagne@gel.ulaval.ca

Abstract

To evaluate their performance, existing dehazing approaches generally rely on distance measures between the generated image and its corresponding ground truth. Despite its ability to produce visually good images, using pixel-based or even perceptual metrics does not guarantee, in general, that the produced image is fit for being used as input for low-level computer vision tasks such as segmentation. To overcome this weakness, we are proposing a novel end-to-end approach for image dehazing, fit for being used as input to an image segmentation procedure, while maintaining the visual quality of the generated images. Inspired by the success of Generative Adversarial Networks (GAN), we propose to optimize the generator by introducing a discriminator network and a loss function that evaluates segmentation quality of dehazed images. In addition, we make use of a supplementary loss function that verifies that the visual and the perceptual quality of the generated image are preserved in hazy conditions. Results obtained using the proposed technique are appealing, with a favorable comparison to state-of-the-art approaches when considering the performance of segmentation algorithms on the hazy images.

Dehazing, Image segmentation, Deep neural network, Generative models

1 Introduction

Images used for segmentation often suffer from poor weather conditions, including haze, snow, and rain. Nowadays, Deep Learning (DL) techniques are widely used to perform segmentation tasks. They require a large amount of training data covering different situations, including noisy images, in order to properly generalize the segmentation task. However, currently available datasets do not guarantee sufficient representativeness or even the presence of some meteorological conditions (e.g., haze) in their training data. Thus, testing the model with hazy images can hinder the performance for low-level computer vision algorithms such as segmentation [1], even if this algorithm is known to be powerful in various circumstances. Therefore, restoring images from degraded observation under hazy conditions is a useful preprocessing step toward improving better segmentation. Several dehazing techniques have been extensively studied in the literature, most of them being based on physical models of image degradation, the problem being simplified to estimating the transmission map [2]. Such an estimation is made either by using a model (e.g., dark channel prior [3]) or through some learning approach (e.g., dehazenet [4]). Other methods do not assume a physical model and try to build an end-to-end system for haze removal based on generative models (i.e., GAN) and restore original images directly from the hazy ones (e.g., [5, 6]).

Usually, the performance of dehazing is evaluated only with respect to some empirical measures. For instance, we note Structural Similarity Index (SSIM) and Peak signal to noise ratio (PSNR) [7] as the most widely used measures found in the literature to evaluate the effectiveness of denoising methods. SSIM measures the similarity between two images, taking into account the similarity of the edges. PSNR is an indicator of the quality of the transmission of information. However, despite the performance of DL algorithms in dehazing in terms of SSIM and PSNR, there is no guarantee that they will produce images fit for being used as input to segmentation methods [8].

This issue has been tackled by only a few papers found in the literature, which aim at reducing haze in order to improve segmentation quality. In Li et al. [9], a dehazing algorithm for single hazy images was performed then fine-tuned for detection using a Fast-RCNN, with improvement reported on detection accuracy on the generated images. Sakaridis et al. [10] proposed an end-to-end system for segmentation of a foggy scene, with training only on foggy dataset. Results show that the model was able to perform segmentation well on hazy images. The proposed model differs from Li et al. [9] by adding a segmentation loss to the training procedure and not only a fine-tuning on segmentation and differ from Sakaridis
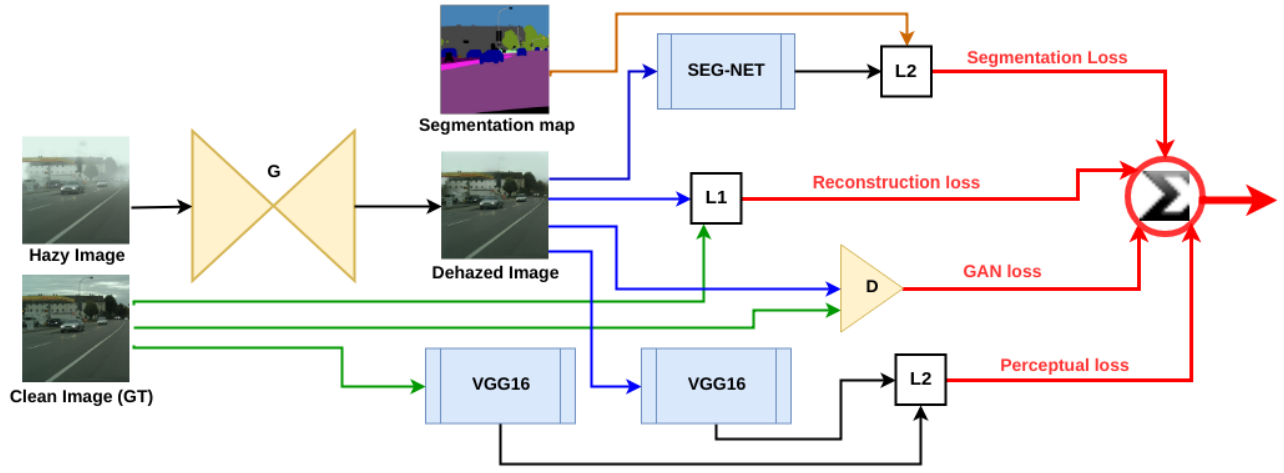
Figure 1: Schema representing the DFS model

et al. [10] by being a preprocessing step toward segmentation of clean images and not an end-to-end segmentation technique for hazy images. More recently, Liu et al. [11] proposed a deep neural network solution that matches image denoising to computer vision tasks, and using the joint loss for updating only the denoising network via back-propagation. However, it assumes that only an independent and identically distributed Gaussian noise with zero mean is added to the original image as the noisy input image during training, which is not reflecting the usual noise we found in real-world problems. In the current paper, we are rather aiming at learning how to handle image degradation related to natural artifacts (i.e., haze).

The main contribution of the paper is a new DL-based dehazing system designed to take into account segmentation performance during training, for improved performance. We continue by detailing the methodology proposed (Sec. 2), before reporting experiments and results (Sec. 3) validating the capacity of the proposed technique for an effective dehazing of images, usable for image segmentation.

2   Methodology

2.1   Single Image Dehazing

Currently, several dehazing algorithms attempt to generate clean images using a generative model such as GAN [12], in which a generator attempted to fool a discriminator, providing most realistic fake sample possible. Specifically, Li et al. [5] used a conditional GAN (CGAN) to perform single image dehazing. We are proposing a model derived from the Pix2Pix architecture [13] for dehazing images, with the generator part of the CGAN composed of a downsampling, a residual, and an upsampling block structure inspired from Johnson et al. [14]. The loss function of the generator for single image dehazing is:

$$L_{generator} = L_{GAN} + \lambda_1 \, L_{pixel} + \lambda_2 \, L_{percep}, \qquad (1)$$

where L_{GAN}, L_{pixel}, and L_{percep} are themselves loss functions on specific elements of the task and λ_i are weighting relative influence in a linear combination. L_{GAN} is the loss function from Isola et al. [13] used to generate fake images. L_{pixel} is the reconstruction loss between the ground truth for dehazing (a.k.a. the real image) and the fake dehazed image, based on their individual pixel values, allowing the network to produce crisper images. L_{percep} is the perceptual loss used for preserving important semantic elements of the image in the output of the generator. Indeed, it was shown [5, 6, 14] that the use of perceptual loss improves the quality of the output by relying on the high-level representation features of a neural network (here a frozen VGG-16 [15] is used), which are compared between the real image and the fake dehazed one. It is worth mentioning here that the same GAN discriminator loss function proposed in the original formulation of Pix2Pix is kept.

| (1) Hazy | (2) Dehaze | (3) DFS | (4) GT | (5) Hazy | (6) Dehaze | (7) DFS | (8) GT |

| (9) Hazy | (10) Dehaze | (11) DFS | (12) GT | (13) Hazy | (14) Dehaze | (15) DFS | (16) GT |

| (17) Hazy | (18) Dehaze | (19) DFS | (20) GT | (21) Hazy | (22) Dehaze | (23) DFS | (24) GT |

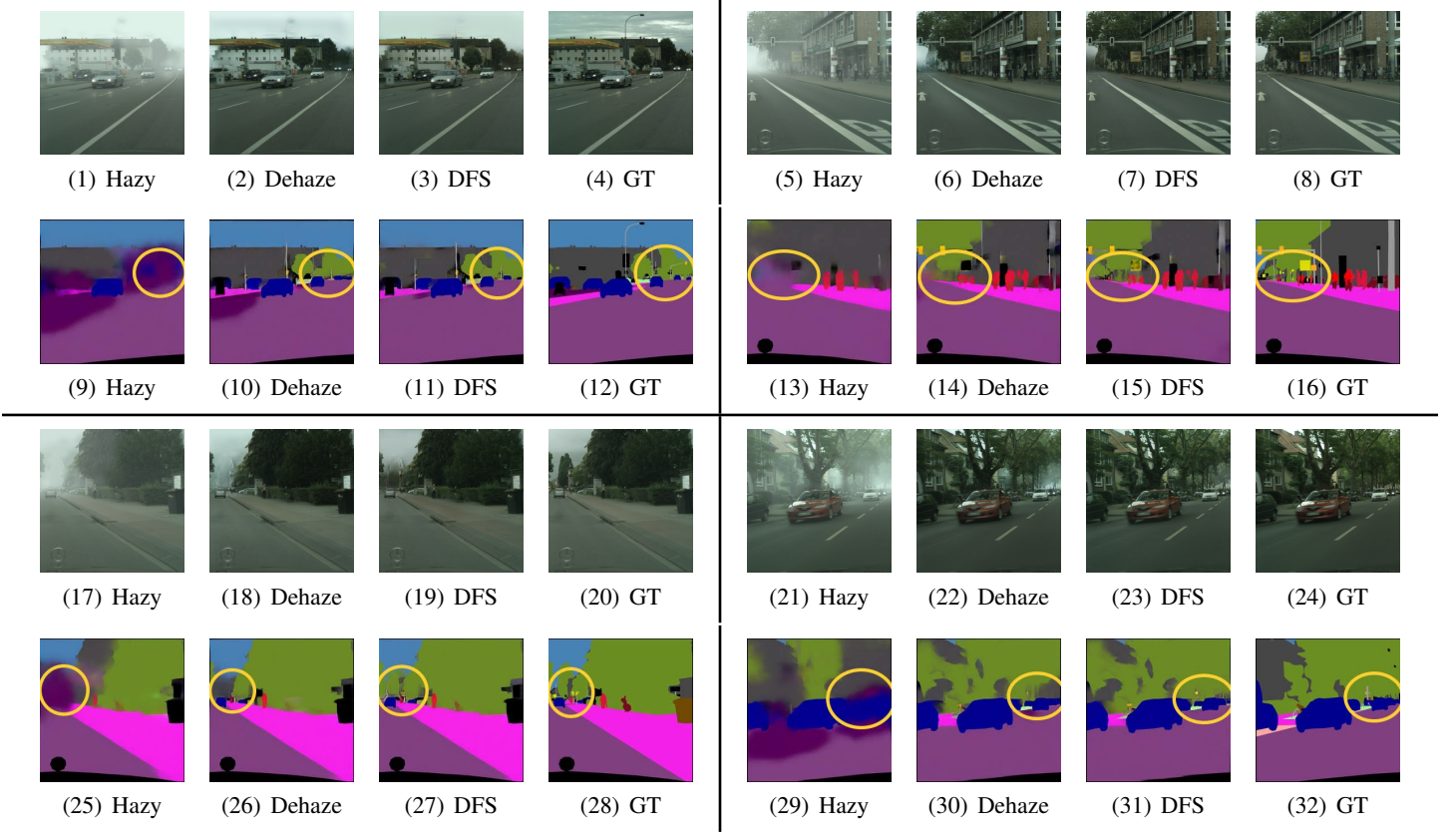| (25) Hazy | (26) Dehaze | (27) DFS | (28) GT | (29) Hazy | (30) Dehaze | (31) DFS | (32) GT |

Figure 2: Four examples showing the performance of the proposed technique: Images 9-12, 25-28 and 29-32 present the problem of far away segmentation and images 13 to 16 show difficulties of far away human segmentation, they are the resulting segmentation map of respectively the images 1-4 for 9-13, 5-8 for 12-16, 17-20 for 25-28 and 21-24 for 29-32, segmentation is done using the SEG-NET, except for the GT segmentation map which is the ground truth, not the output of the SEG-NET.

2.2 Dehazing for Segmentation (DFS)

The quality of segmentation relies on the input quality, which does not depend only on the acquisition devices but also on the environmental conditions such as haze or other weather conditions. Thus, removing these artifacts from the input images (e.g., by dehazing) can have an impact on segmentation quality. A dehazing method for segmentation purposes (referred to as DFS in this paper) is performed by using the following loss function to train the CGAN generator:

$$L_{generator} = L_{GAN} + \lambda_1 L_{pixel} + \lambda_2 L_{percep} + \lambda_3 L_{seg}. \tag{2}$$

Compared to Eq. 1, it adds a new loss component, L_{seg}, which evaluates the impact of dehazing on the segmentation performance.

The architecture of the DFS model used is shown in Fig. 1. The generator network receives an image with haze as an input and gives a candidate of a dehazed image as the output. Similar to the single image dehazing model, the generator loss (Eq. 2) is computed through L_{GAN}, L_{pixel}, and L_{percep}. The segmentation loss L_{seg}, is computed by placing the output of the generator (i.e., the dehazed image) into the segmentation network. The obtained segmentation map is then compared to the ground truth segmentation map, using the L_2 loss. Basically, the model tries at the same time to remove haze as much as possible while preserving, or even improving segmentation performance.

Table 1: Results of the dehazing model on the NYU depth foggy dataset in comparison to the literature. DCP stands for Dark Channel Prior [3], CAP for Color Attenuation Prior [18], NCP for Non-Local Color Prior [19], MSCNN for Multi-Scale CNN [20], DN for DehazeNet [4], CG for CycleGan [21], DDN for Disentangled Dehazing Network [6], while Hazy is the original hazy image. Results marked with a star (*) have been produced in our experiments, while the other non-marked results are reported from Yang et al. [6].

Metrics	DCP [3]	CAP [18]	NCP [19]	MSCNN [20]	DN [4]	CG [21]	DDN [6]	Ours*	DCP*	CG*	Hazy*
PSNR	10.98	12.78	13.05	12.27	12.84	13.39	15.55	***17.89***	*9.12*	*9.75*	*3.67*
SSIM	0.65	0.71	0.67	0.70	0.72	0.52	**0.77**	*0.744*	*0.52*	*0.45*	*0.20*

3 Experiments and Results

3.1 Single Image Dehazing

3.1.1 Dehazing Dataset

The D-Hazy dataset is used for our experiments [16]. D-Hazy contains 1449 pairs of synthetic hazy images with ground truth based on the NYU Depth dataset [17], which is composed of indoor images and their corresponding depth maps. The haze is computed from the depth map using physical models, providing realistic haze. The dataset is split 80%/20% between training and test partitions. Hyperparameters tuning is done through optimization over a random 80%/20% split of the training set. All the training set is used to infer the model with the selected hyperparameters.

3.1.2 Results for single image dehazing

The proposed dehazing model was trained for 200 epochs using a batch size of 16 and Adam as the optimizer. Every image is resized to a resolution of 256x256. For the generator loss function (Eq. 1), λ_1 and λ_2 are set as 10, following results obtained with a grid search over the validation set, which is made by testing 10 values of λ between 1 and 50. Results obtained with the proposed technique achieve a PSNR of 17.89 dB versus 15.55 dB for the state-of-the-art [6] and a SSIM of 0.744 versus 0.77, a light degradation of 3.4%. Table 1 presents the results of the proposed model versus multiple dehazing models, including the actual state-of-the-art, Disentangled Dehazing Network [6]. Because the test set used in our paper is different from the one used by Yang et al. [6], we need to test algorithms appearing in the results of [6] with our test set. According to the results in Table 1, results obtained with our test set (marked with a *) show that it is of similar difficulty, maybe even harder, than the one used by Yang et al. [6], with DCP and CycleGan performing less well on our test set. Our results also demonstrate empirically the capacity of the proposed model to remove haze from single images.

3.2 DFS

3.2.1 Dataset for segmentation of hazy images

The second set of experiments is conducted on the Cityscape dataset [22], which includes a depth map, to simulate a realistic haze level while providing relatively complex scenes (i.e., a high traffic level in the images). For that purpose, we follow the methodology of Sakaridis et al. [10] to generate realistic hazy image datasets. The final dataset is composed of ≈550 foggy images with the ground truth segmentation map in fine annotation. Around 50 images were randomly selected from the training set for the validation set while the 51 images proposed as a validation set for the foggy Cityscape [10] were used as a testing set.

3.2.2 Training methodology and parameters

In addition to its dehazing capability the proposed DFS model contains a segmentation loss which is computed using a segmentation model (referred to as SEG-NET). Except for the segmentation loss, this model is based on the same architecture as the model for single image dehazing. To train SEG-NET we use a subset of the Cityscape dataset not used to train the dehazing algorithm. The training was for 40 epochs using a batch size of 16 and Adam as the optimizer. The DFS model was trained for 100 epochs using a batch size of 8 and Adam as the optimizer. Here all of the images (size 2048x1024) are cropped into two squares (1024x1024) and resized to 256x256, cropping is necessary to keep the maximum amount of

Table 2: Results on the test set with SEG-NET segmentation (foggy Cityscape). For dehaze (single image dehazing) and dehaze with segmentation loss (DFS), results were done by averaging the best model obtained in validation on each on five different runs. Results in bold are the bests and results in italic are from the ground truth (GT).

Metrics	Hazy	Dehaze	DFS	GT
PSNR	11.95	14.77	**15.36**	*15.45*
SSIM	0.624	0.727	**0.747**	*0.748*

Table 3: Results on the test set with DeepLabv3 segmentation (foggy Cityscape). For dehaze (single image dehazing) and dehaze with segmentation loss (DFS), results were obtained by averaging the best model obtained in validation on each on five different runs. Results in bold are the bests and results in italic are from the ground truth (GT).

Metrics	Hazy	Dehaze	DFS	GT
IoU-cl	0.556	0.553	0.557	*0.570*
iIoU-cl	0.294	0.309	**0.316**	*0.340*
IoU-ca	0.745	0.774	0.775	*0.794*
iIoU-ca	0.518	0.554	**0.569**	*0.627*

information and structure of the image before resizing it to a size and shape suited for this network. The parameters λ_1, λ_2 and λ_3 of Eq.2 have been set to 10, 10 and 5, respectively, by the application of a grid search using the same logic as for the single image dehazing, this time only on λ_3, both λ_1 and λ_2 are the same as for the single image dehazing. During the training of the DFS model, SEG-NET is frozen (i.e., not updated by gradient descent).

3.2.3   Testing methodology

During testing, segmentation results are reported on the same SEG-NET used for training the DFS model, and on another state-of-the-art segmentation model trained on Cityscape, DeepLabv3 [23]. The DeepLabv3 model was not used for training nor in validation, only for the final test reported here. Comparisons have been done between segmentation performances, after dehazing using models trained with and without segmentation loss, using foggy images and comparing with the ground truth non-hazy images. The DFS model and the single image dehazing model are trained using exactly the same training parameters.

Table 2 presents the results of the PSNR, SSIM and MSE (mean squared error) with the hazy images, dehazed images, dehazed images with segmentation loss and the ground truth images, without haze, all using SEG-NET segmentation. Table 3 presents the results with DeepLabv3, IoU is the intersection over union metric, ca is for categories, cl is for classes and iIoU refers to instance-level intersection-over-union metric. According to Cordts et al. [22], the IoU measure is biased toward object instances that cover a large image area. In street scenes, where there can be a strong scale variation, this can be problematic, especially for traffic participants. To address this problem it has been suggested to use the iIoU metric, where the contribution of each pixel is weighted by the ratio of the class average instance size to the size of the respective ground truth instance, only classes with instance annotations are included for this measure (classes: person, rider, car, truck, bus, train, motorcycle and bicycle, categories: human and vehicle).

3.2.4   Discussion of results

In the light of the results obtained with both SEG-NET and DeepLabv3 segmentation network, it appears that adding a loss for segmentation in the dehazing network significantly increases the accuracy of subsequent segmentation.

With SEG-NET and DeepLabv3, results are always better when the output of DFS is used compared to dehazing models trained without segmentation loss. On average, the boost in segmentation with SEG-NET (PSNR/SSIM) from hazy to dehazed is around 20%, while from dehazed to DFS, a 3.5% gain is achieved. With DeepLabv3, the IoU metric differences are not significant for classes even with and without haze (0.556 versus 0.570), but there is a nice improvement for IoU with categories from hazy to dehaze (boost of 3.9%), while a little gain (0.13%), is observed from dehaze to DFS. Yet a significant improvement is observed with DeepLabv3 using segmentation loss with iIoU metric, the boost in segmentation for classes

using iIoU from hazy to dehazed is of 5.1%, while from dehazed to DFS the gain is 2.3%. The boost in segmentation for categories using iIoU from hazy to dehazed is 7% and for dehazed to DFS it reaches 2.7%.

Looking at Fig. 2, subfigures 10, 11, 14, 15, 26, 27, 30, and 31, a significant improvement is perceptible for targets of interest located far away in the scene, especially with segmentation of cars and pedestrians, the largest differences being circled. The difference in segmentation performance between hazy (e.g., Fig. 2.9) and both dehazing techniques (e.g., Fig. 2.10-11) are big, adding the segmentation loss is giving a little boost making the results more similar to the ground truth (e.g., Fig. 2.12). Comparison between normal dehazing and DFS (e.g., Fig. 2.22-23) shows more similar results, with DFS appearing to dehaze a little better for targets far away in the scene. The segmentation network also appears to be sensitive to elements that we usually don't notice when looking at the images.

4   Conclusion

In brief, this paper demonstrates the usefulness of including segmentation loss in an end-to-end training of deep learning models for dehazing. The learning-based dehazing model is generated not just for denoising metrics, but also with an optimization criterion aimed at achieving something useful for a specific task, and with performance improvements that can be significant in comparison to results obtained with an unguided approach. Moreover we can consider to boost even more the performance of DFS using directly an approximation of the IoU/iIoU measures for gradient descent [24], which are better optimization measure than mean square error and similar.

Acknowledgments

References

[1] Robby T Tan. Visibility in bad weather from a single image. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008.

[2] Raanan Fattal. Single image dehazing. *ACM transactions on graphics (TOG)*, 27(3):72, 2008.

[3] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2011.

[4] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[5] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8202–8211, 2018.

[6] Xitong Yang, Zheng Xu, and Jiebo Luo. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *Association for the Advancement of Artificial Intelligence (AAAI)*, 2018.

[7] Cosmin Ancuti, Codruta Orniana Ancuti, Radu Timofte, Luc Van Gool, Lei Zhang, Ming-Hsuan Yang, Vishal M Patel, He Zhang, Vishwanath A Sindagi, Ruhao Zhao, et al. Ntire 2018 challenge on image dehazing: Methods and results. 2018.

[8] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019.

[9] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *IEEE International Conference on Computer Vision (ICCV)*, pages 4770–4778, 2017.

[10] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, pages 1–20, 2018.

[11] Ding Liu, Bihan Wen, Xianming Liu, Zhangyang Wang, and Thomas Huang. When image denoising meets high-level vision tasks: A deep learning approach. In *International Joint Conference on Artificial Intelligence, IJCAI*, pages 842–848. International Joint Conferences on Artificial Intelligence Organization, 2018.

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[13] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1125–1134, 2017.

[14] 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.

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

[16] Cosmin Ancuti, Codruta O Ancuti, and Christophe De Vleeschouwer. D-hazy: a dataset to evaluate quantitatively dehazing algorithms. In *IEEE International Conference on Image Processing (ICIP)*, pages 2226–2230. IEEE, 2016.

[17] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *European Conference on Computer Vision (ECCV)*, pages 746–760. Springer, 2012.

[18] Qingsong Zhu, Jiaming Mai, Ling Shao, et al. A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Processing*, 24(11):3522–3533, 2015.

[19] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1674–1682, 2016.

[20] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*, pages 154–169. Springer, 2016.

[21] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017.

[22] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3213–3223, 2016.

[23] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[24] Maxim Berman, Amal Rannen Triki, and Matthew B Blaschko. The lovász-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4413–4421, 2018.