Mutual Learning for Domain Adaptation: Self-distillation Image Dehazing Network with Sample-cycle

Erkang Chen, Sixiang Chen, Tian Ye, Yunchen Zhang, Liang Chen, Xingjie Lai, Yun Liu

***Abstract*—Deep learning-based methods have made significant achievements for image dehazing. However, most of existing dehazing networks focus on training models using synthetic hazy images, resulting in generalization performance degradation when applying on real-world hazy images because of the domain shift problem. In this letter, we propose a mutual learning dehazing framework for domain adaption. Specifically, we first devise two siamese networks: a teacher network in the synthetic domain and a student network in the real domain, and then optimize them in a mutual learning manner by leveraging the exponential moving average (EMA) and joint loss. Moreover, we design a sample-cycle strategy based on haze density augmentation (HDA) module to introduce pseudo real-world image pairs provided by the student network into training for further improving the generalization performance. Extensive experiments on both synthetic and real-world dataset demonstrate that the proposed framework outperforms state-of-the-art dehazing techniques in terms of subjective and objective evaluation.**

***Index Terms*—Single image dehazing, domain adaptation, self-distillation, semi-supervised and self-training.**

SINGLE image dehazing is a significant image processing problem in computer vision community, which aims to recover the clean image from a hazy input. Early dehazing methods attempt to remove the haze relying on hand-crafted priors [2], [8], [10], [23]. However, these hand-designed priors from human observation may not always hold in case of complex real-world hazy scenes. To circumvent hand-crafted priors dependency, numerous deep learning-based networks [3], [11], [14], [16], [19]–[22] have been proposed for image dehazing, which achieves better performance than prior-based approaches owing to the powerful feature representation ability of deep convolutional neural networks. As learning-based techniques typically require a large number of synthetic paired images and the real-world image paired data under haze and haze-free conditions is difficult to collect, most of them focus on training dehazing models with synthetic hazy dataset. Unfortunately, they fail to generalize well to real-world hazy scenes due to the domain gap between synthetic and real data.

(a) Hazy image (b) DA (c) PSD (d) Ours

Fig. 1: Dehazed results generated by DA [17], PSD [5] and our method on a real-world hazy image.

To solve the domain shift problem, some works [5], [13], [15], [17] made use of domain translation strategy to reduce the discrepancy between different domains. Li *et al*. [13] designed a semi-supervised dehazing framework through sharing weights. Shao *et al*. [17] developed a domain adaptation (DA) paradigm to bridge the gap between the synthetic and real domain. A principle synthetic-to-real dehazing (PSD) framework [5] is presented to attempt to adapt synthetic data based models to the real domain. Liu *et al*. [15] develop a disentangle-consistency mean-teacher network (DMT-Net) collaborating with unlabeled real-world hazy images. These above approaches leverage weights-sharing or domain translation to build up the relationship between the synthetic and real domains. However, in Fig. 1, their dehazing ability on the real domain is insufficient to provide the clean image because they fail to achieve the domain adaptation in essence, resulting in a performance gap between the synthetic and real domains.

To achieve the domain adaptation actually, we propose a novel mutual learning paradigm for single image dehazing. The main contributions are summarized:

- A mutual learning paradigm for domain adaptation is proposed to address the domain gap through knowledge mutual transfer, in which the implicit knowledge is transferred from teacher to student using EMA and meanwhile the rich external knowledge is constructed by the pixel-wise adversarial loss and the dark channel prior loss and explicitly transferred from student to teacher.
- A simple but effective self-distillation framework for single image dehazing is presented, which consists of two siamese networks: a teacher network in the synthetic domain and a student network in the real domain. These two networks can learn useful knowledge from each other through the mutual learning paradigm. The final student network is capable to achieving superior dehazing performance on both domains, simultaneously.

This work is supported partially by the Natural Science Foundation of Fujian Province of China under Grant (2021J01867), Education Department of Fujian Province under Grant (JAT190301), Foundation of Jimei University under Grant (ZP2020034), the National Nature Science Foundation of China under Grant (61901117), Natural Science Foundation of Chongqing, China under Grant (No. cstc2020jcyj-msxmX0324).

E. Chen and S. Chen have the same contributions to this work.

E. Chen, S. Chen, T. Ye and X. Lai are with School of Ocean Information Engineering, Jimei University, Xiamen 361021, China (e-mail:ekchen@jimei.edu.cn)

Y. Zhang is with China Design Group Co., Ltd. Nanjing, China(cydiachen@cydiachen.tech).

L. Chen is with Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou, China (cl_0827@126.com)

Y. Liu is with College of Artificial Intelligence, Southwest University, Chongqing 400715, China (e-mail: yunliu@swu.edu.cn).

Corresponding author: Yun Liu

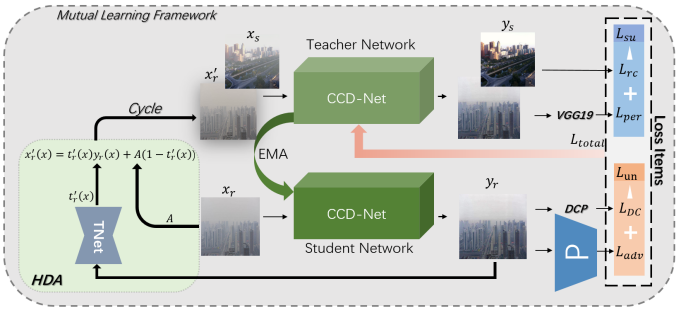

Fig. 2: Overview of the proposed self-distillation dehazing framework. Our model mainly consists of three parts: a teacher network, a student network and a HDA module. The T-Net of our HDA module estimates the transmission map from the output of the student network. The global atmospheric light A is predicted by the DCP. The trapezoid with "P" denotes the discriminator of PatchGAN [9].

- We develop a data augmentation strategy dubbed as haze density augmentation (HDA) in a sample-cycle fashion. The haze-free images acquired by the student network are fed into the HDA module to form the pseudo real-world image pairs with diverse haze density, which is introduced into the teacher network for data augmentation.

I. Proposed Method

A. Overall framework

Given a synthetic dataset $X_{syn} = \{x_i, y_i\}_{s=1}^{N_s}$ and a real-world hazy dataset $X_{real} = \{x_i\}_{r=1}^{N_r}$, where N_s and N_r respectively represent the number of the synthetic and real hazy images. We attempt to acquire a simple and accurate dehazing model which can recover accurately clear images at inferencing phases. We found that most previous dehazing models trained only on the synthetic dataset, and generalized well on the synthetic domain. However, due to the domain gap between the synthetic domain and real domain, these methods fail to work well to the image of both domains simultaneously.

To deal with this problem, we present a novel self-distillation image dehazing framework for both domain adaptation, which consists of three main parts: a teacher network $G_{s\to c}$, a student network $G_{r\to c}$ and a HDA module $G_{Density}$, depicted in Fig. 2. With the help of supervised learning, the teacher network learns how to build a mapping from synthetic hazy images to clean images on paired synthetic dataset X_{syn}. Under the useful knowledge transferred from $G_{s\to c}$, the student network learns how to generate satisfactory dehazing results from real-world hazy images. Implicit knowledge transfer considers to transfer the knowledge from teacher to student by weights sharing and joint training, however, which only provides limited internal knowledge.

To fully explore the actual presentation learning ability of the student network, we design a mutual learning mechanism to evaluate the "real" dehazing performance of the student network online. This end-to-end approach avoids the complicated multi-domain translation training scheme. Moreover, it also enables a "flywheel effect" that the mutual learning of $G_{s\to c}$ and $G_{t\to c}$ can mutually reinforce each other, so that both get better and better as the training goes on.

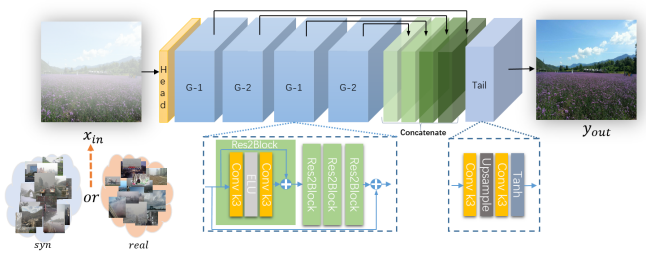

Fig. 3: Architecture of compact cascaded dehazing network (CCD-Net).

B. Mutual Learning of Teacher and Student

Most previous distillation frameworks only transfer implicit knowledge unidirectionally, thus the student network only perform well in a single domain (e.g, real or synthetic). Why not let the teacher network understand the actual real performance of the student network on the other domain online and know how to teach the student better? Based on the above consideration, the proposed mutual learning mechanism boosts the well performance of student on both domain with the help of mutual learning of teacher and student, which indicates the final inferencing network (student) generalizes well on both real and synthetic domain, different from the previous domain-splitting adaptation method for image dehazing.

Furthermore, the proposed mutual learning paradigm allows unsupervised loss function to online evaluate the actual performance of the student in the real domain. Specifically, unlabeled real-world hazy images x_{in}^r as a batch are randomly sampled from X_{real}. The student model is applied on these images by forwarding propagation, thus capturing the adversarial and dark channel prior loss as the "online evaluated errors":

$$\mathcal{L}_{un} = \mathcal{L}_{adv}(G_{r\to c}(x_{in}^r)) + \mathcal{L}_{DCP}(G_{r\to c}(x_{in}^r)). \quad (1)$$

Afterwards, $\mathcal{L}_{un}$ will be feedback to the teacher network $G_{s\to c}$ for back propagation, which effectively helps the teacher network to teach the student network how to generalize well on real-world data and provides external knowledge to teacher. The mutual learning paradigm explores the closed upper and lower bound constraints of the student network in solution space, which significantly improves the performance of $G_{r\to c}$.

C. Homogeneous Architecture for Self-distillation Network

Most previous dehazing approaches [7], [15] attempt to improve the performance by complicated network design with heavy parameters, resulting in difficult deployment. Thus, we aim to devise a compact cascaded dehazing network (CCD-Net) as our distillation network, which only has about 4M parameters, far less than previous prevalent methods, such as MSBDN (31.35M) [7] and DMT-Net (51.79M) [15].

To this end, we use 4 Res2Block groups with 64 channels to extract the features after the down-sample by Head module, which is the convolution operation with the kernel size of 3×3 and stride of 2. In order to fully make use of features from different depth of network, we conduct 3×3 convolution and concatenation operation of channel dimension to reduce the dimension to 64 channel in the feature aggregation stage. In Fig. 3, we adopt the pixel shuffle operation to up-sample

(a) Original (b) p=0.5 (c) p=0.7 (d) p=0.9 (e) p=1.1 (f) p=1.4

Fig. 4: Examples of the proposed HDA module. (a) are the original hazy images. (b)-(f) are generated hazy images with different haze density by adjusting the factor p.

the features to the original resolution, and the Tail module is devised to fuse the extracted features by 3×3 convolution. The tanh function is considered as the activation of the output to generate final image. From Fig. 3 and Table I that the proposed self-distillation framework avoids complicated hand-craft on purpose but still works well in the synthetic and real domain.

D. Sample-cycle with Haze Density Augmentation

In order to further improve the generalization ability of the student network and enrich the sample complexity of teacher for better domain compatibility, we propose a HDA module to adjust the haze density of real-world images effectively. Different from previous data augmentation for image dehazing methods such as random crop size and random flip which only perform on synthetic datasets or original training samples, we focus on the nonlinear adjustment of transmission in real-world samples for harder hazy samples.

Specifically, we suppose that the real-world dehazed images produced by the student network are clear and satisfactory enough after certain training epochs. Based on this assumption, we can reserve the real-world hazy images and dehazed results to compose the paired samples: $\{x_r, y_r\}_{r=1}^{N}$, then the pre-trained transmission estimation network T-Net is used to acquire the accurate transmission map $t_r(x)$. The detailed architecture of T-Net is illustrated in Fig. 5.

To nonlinear augment the hazy density controllably, we set the random factor p to adjust the intensity of transmission and control $t_r^{'}(x)$ in $[0, 0.99]$:

$$t_r^{'}(x) = t_r^{p}(x_r) \tag{2}$$

Finally, the classic atmospheric scattering model is used to rebuild the hazy images with diverse haze density:

$$x_r^{'}(x) = t_r^{'}(x) y_r(x) + A(1 - t_r^{'}(x)) \tag{3}$$

where A is the atmospheric light estimated by the DCP [8], and T-Net is optimized ceaselessly during the training phrase. Our HDA module effectively rebuilds real-world hazy images and adjusts the haze density controllably, which significantly enrichs the training samples for the teacher network. In Fig. 4, we give some examples with different haze density generated by our proposed HDA module via adjusting the factor p. The proposed HDA module not only enhances the density of original hazy image, but also provides more "soft" samples for better convergence in training phase.

After augmentation of haze density, we re-compose $x_r^{'}(x)$ and $y_r(x)$ to form paired training samples. These paired

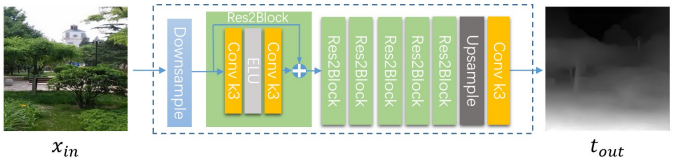

Fig. 5: Overview of T-Net in HDA module.

pseudo real-world samples $\{x_r^{'}, y_r\}$ are conductive for the teacher network to further explore the useful knowledge, which can help the teacher network teach the student network better. In the training stage, the re-composed paired samples are randomly selected to replace the synthetic samples of a batch and the above process likes a special cycle with samples that boosts the generalization ability of the student network.

E. Training Losses

In the proposed self-distillation dehazing framework, we adopt the following loss functions to train the network.

Reconstruction Loss: We utilize the Charbonnier loss [4] as our reconstruction loss function:

$$\mathcal{L}_{rc} = \mathcal{L}_{char}(G_{s\to c}(x_{in}^{s}), y_{gt}^{s}) + \lambda_{HDA}\mathcal{L}_{char}(G_{s\to c}(x_r^{'}), y_r). \tag{4}$$

where y_{gt}^{s} stands for ground truth, λ_{HDA} is the balance weight for controlling the contribution of the augmented samples, and $\mathcal{L}_{char}$ can be presented by:

$$\mathcal{L}_{\text{char}} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\|X^i - Y^i\|^2 + \epsilon^2}. \tag{5}$$

where ϵ is empirically set as $1e-3$ in our experiments.

Perceptual Loss: Besides the pixel-wise supervision, we also utilize the perceptual loss based on the VGG-19 [18] pre-trained on ImageNet [6]:

$$\mathcal{L}_{\text{per}} = \sum_{j=1}^{3}\frac{1}{C_j H_j W_j}\left\|\phi_j\left(y_{gt}^{s}\right) - \phi_j(G_{r\to c}(x_{in}^{s}))\right\|_2^2. \tag{6}$$

where H_j, W_j and C_j denote the height, width, and channel of the feature map in the j-th layer of the backbone network, ϕ_j is the activation of the j-th layer.

Adversarial Loss: The avdversarial loss $\mathcal{L}_{adv}$ is defined based on the probabilities of the discriminator $D(G_{r\to c}(x_r))$ over the real-world hazy samples as:

$$\mathcal{L}_{adv} = \sum_{n=1}^{N} -\log D\left(G_{r\to c}\left(x_{in}^{r}\right)\right). \tag{7}$$

DCP Loss: The dark channel [8] is formulated as:

$$DC(x_{in}) = \min_{y\in N(x)}\left[\min_{c\in\{r,g,b\}} x_r^{c}(y)\right]. \tag{8}$$

where x and y are pixel coordinates of image x_r^{c}, x^{c} denotes c-th color channel of x_r, and $N(x)$ represents the local neighbor window centered at x_r. Previous work [8] has demonstrated that the most intensity of the dark channel image are close to zero. Thus, we utilize the dark channel loss to enforce that the dark channel of the dehazed images are in consistence with that of haze-free images:

$$\mathcal{L}_{DC} = \left\|DC\left(G_{r\to c}(x_{in})\right)\right\|_1. \tag{9}$$

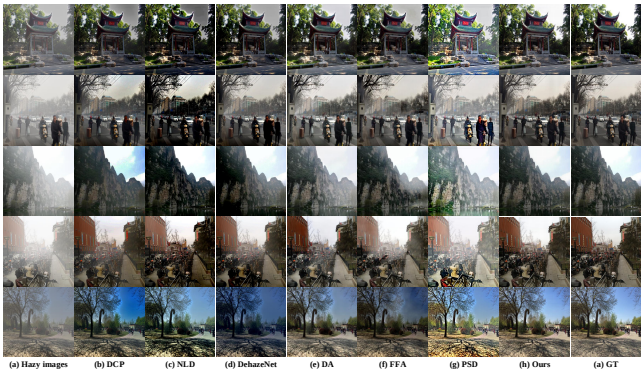

Fig. 6: Visual comparisons on synthetic hazy images.

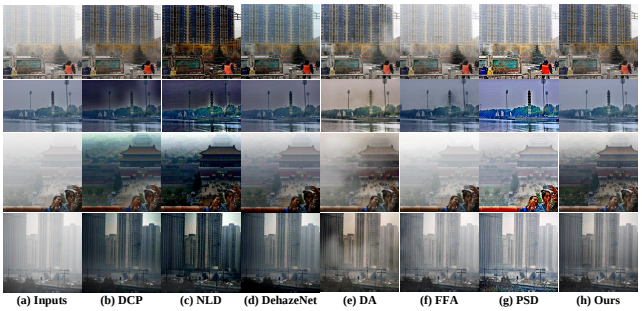

Fig. 7: Visual comparisons on real-world hazy images.

Overall Loss Function: With all the four members, the overall loss function is defined as follow:

$$\mathcal{L} = \lambda_{rc}\mathcal{L}_{rc} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{DC}\mathcal{L}_{DC} + \lambda_{per}\mathcal{L}_{per} \qquad (10)$$

where $\lambda_{rc}, \lambda_{adv}, \lambda_{DC}, \lambda_{per}$ are trade-off weights.

II. Experiments

A. Implementation Details

Datasets. The synthetic and real-world images from RESIDE dataset [12] are selected for training. Concretely, 6000 paired synthetic hazy images, 3000 from ITS and 3000 from OTS, are chose for training. In addition, we also randomly select 1000 images from URHI (Unannoteated Real-world Hazy Images) in the training stage for improving the generalization ability to real domain. To evaluate the effectiveness of the proposed framework, we compare the proposed framework with other state-of-the-art dehazing methods on two synthetic datasets, i.e. SOTS [12] and Haze4k [15], and a real-world dataset, i.e. I-HAZE [1].

Training details. We augment the training dataset with randomly rotated by 90,180,270 degrees and horizontal flip. And randomly crop all the synthetic and real images to 256×256 for the training. We utilize Adam optimizer with initial learning rate of 1×10^{-4}, and adopt the CyclicLR to adjust the learning rate in training. The decay parameter of EMA is 0.999. The trade-off weights are set as: $\lambda_{rc} = 1$,$\lambda_{adv} = 0.2$,$\lambda_{DC} = 10^{-2}$,$\lambda_{per} = 0.2$ and $\lambda_{HDA} = 0.5$.

Inference. In the testing phase, we only feed the hazy image into the student network and treat the predicted image from the student network as the final dehazed result.

TABLE I: Quantitative comparisons with the state-of-the-art dehazing methods on three datasets (PSNR(dB)/SSIM). Best results are **bold**. Second-best results are underlined.

Method	Haze4k [15]		SOTS [12]		I-HAZE [1]	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
(TPAMI'10)DCP [8]	14.01	0.76	15.09	0.76	14.43	0.75
(CVPR'16)NLD [2]	15.27	0.67	17.27	0.75	14.12	0.65
(TIP'16)DehazeNet [3]	19.12	0.84	21.14	0.85	10.10	0.60
(ICCV'17)AOD-Net [11]	17.15	0.83	19.06	0.85	13.98	0.73
(ICCV'19)GDN [14]	23.29	0.93	23.29	0.95	16.62	0.78
(CVPR'20)DA [17]	24.61	0.90	27.76	0.93	17.73	0.73
(AAAI'20)FFA-Net [16]	26.96	0.95	26.88	0.95	12.70	0.54
(CVPR'21)PSD-FFA	16.56	0.73	13.99	0.697	15.04	0.69
(ACMMM'21)DID-Net [15]	27.81	0.95	28.30	0.95	-	-
(ACMMM'21)DMT-Net [15]	28.53	0.96	**29.42**	**0.97**	-	-
Ours	**28.99**	**0.97**	28.84	**0.97**	**18.42**	**0.81**

B. Comparisons on Synthetic Hazy Images

The visual comparisons on synthetic hazy images from Haze4k [15] in Fig. 6. From Fig. 6, traditional prior-based dehazing methods, such as DCP [8] and NLD [2], may suffer from the obvious halo artifacts nearby depth discontinuity regions. There still exists some remaining hazes in the results of DehazeNet [3] and FFA [16] because of its insufficient dehazing ability. Although the recent domain adaptation methods, i.e. DA [17] and PSD [5], can improve the quality of degraded images, the obtained dehazed results look unrealistic. Compared with the above algorithms, our method can provide better dehazed results in terms of visual effects, which are closer to the ground truths. Moreover, the quantitative comparisons on synthetic datasets (i.e. Haze4k [15] and SOTS [12]) are revealed in Table I. The proposed approach also achieves best or second-best scores of PSNR and SSIM, which demonstrates the superior performance on synthetic hazy images.

C. Comparisons on Real-world Hazy Images

To evaluate the generalization ability on the real domain, the comparisons on real-world hazy images selected from URHI dataset [12] are depicted in Fig. 7. In Fig. 7, DCP [8] and NLD [2] severely suffer from color distortion in sky regions and the results generated by prior-based methods look unnatural. When handling dense hazy images, the dehazed results of DehazeNet [3] and FFA [16] still have remaining hazes in the last row of Fig. 7. The domain adaptation methods, such as DA [17] and PSD [5], can improve the dehazing performance on real-world hazy images to some extent, but the obvious haze artifacts still remain in the dehazed results, as seen in the first row of Fig. 7 and the obtained results from PSD [5] have the drawback of over-enhancement in the second row of Fig. 7. In comparison, our approach provides a visual pleasant dehazed results with more details and color rendition.

D. Ablation Study

For the ablation study, we choose Haze4K [15] as the training dataset and the performance of different networks are evaluated on Haze4K test-set [15] and I-Haze test-set [1]. In Table II and III, T and S respectively denote the teacher and student network and the arrow stands for the direction of knowledge transfer.

TABLE II: Comparison between different training paradigms of self-distillation dehazing network.

Training Paradigms	Domain of T	Domain of S	PSNR (Synthetic)	PSNR (Real)
T $\rightarrow$ S	Synthetic	Synthetic	27.19	17.23
T $\rightarrow$ S	Synthetic	Real	27.06	17.82
T $\leftrightarrows$ S	Synthetic	Synthetic	**28.05**	17.75
T $\leftrightarrows$ S	Synthetic	Real	27.73	**18.03**

TABLE III: Comparison between different settings of haze density augmentations.

Training Paradigms	Augmentations	Starting Epochs	PSNR (Synthetic)	PSNR (Real)
T $\leftrightarrows$ S	No	-	27.73	18.03
T $\leftrightarrows$ S	HDA	0	28.54	18.24
T $\leftrightarrows$ S	HDA	20	28.72	18.31
T $\leftrightarrows$ S	HDA	50	28.99	18.42
T $\leftrightarrows$ S	HDA	80	28.82	18.24

Effectiveness of mutual learning paradigm: We conduct experiments on different kinds of training paradigms: 1) single path knowledge transfer from teacher to student in the same domain, 2) single path knowledge transfer from teacher to student in separated domains, 3) mutual knowledge transfer on the same domain and 4) mutual knowledge transfer on the separated domains. Table II reveals that single-path knowledge transfer paradigm is inferior to the mutual learning paradigm and the design of feeding the inputs from synthetic domain to the teacher network and feeding the real domain inputs to the student network ensure a better generalization performance in real domain.

Effectiveness of Haze Density Augmentations: Table III shows the performance of different settings of HDA module. We can conclude that: 1) HDA can boost the performance on both synthetic datasets and real-world datasets with a large margin. 2) Different settings of the augmentation strategy influence the performance of overall network. This is because the haze density augmentation adopts an online sampling strategy which directly estimates transmission map from the student network. The inaccurate initialization of the student network may result in the unsatisfactory clean output, which interferes with the estimation of transmission map.

III. Conclusions

In this letter, we have proposed a novel mutual learning paradigm for domain adaptation. Based on this paradigm, our method designs a self-distillation framework consisting of a teacher network and a student network to address the domain shift problem of image dehazing via knowledge mutual transfer within two networks. Moreover, to further improve the generalization ability on the real domain, a novel HDA strategy is presented to augment the haze density of training samples. Experiments on synthetic datasets and real-world images verify the effectiveness of the proposed approach.

References

[1] C. O. Ancuti, C. Ancuti, R. Timofte, and C. D. Vleeschouwer, "I-haze: a dehazing benchmark with real hazy and haze-free indoor images," in *arXiv:1804.05091v1*, 2018.

[2] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1674–1682.

[3] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[4] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proceedings of 1st International Conference on Image Processing*, vol. 2. IEEE, 1994, pp. 168–172.

[5] Z. Chen, Y. Wang, Y. Yang, and D. Liu, "Psd: Principled synthetic-to-real dehazing guided by physical priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 7180–7189.

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[7] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2157–2167.

[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[10] M. Ju, C. Ding, C. A. Guo, W. Ren, and D. Tao, "IDRLP: image dehazing using region line prior," *IEEE Trans. Image Process.*, vol. 30, pp. 9043–9057, 2021.

[11] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4770–4778.

[12] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[13] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, Y. Song, and Y. Ming-Hsuan, "Semi-supervised image dehazing," *IEEE Transactions on Image Processing*, vol. 19, pp. 2766–2779, 2019.

[14] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7314–7323.

[15] Y. Liu, L. Zhu, S. Pei, H. Fu, J. Qin, Q. Zhang, L. Wan, and W. Feng, "From synthetic to real: Image dehazing collaborating with unlabeled real data," in *ACM Multimedia*. ACM, 2021, pp. 50–58.

[16] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 11 908–11 915.

[17] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[19] P. Wang, H. Zhu, H. Huang, H. Zhang, and N. Wang, "Tms-gan: A twofold multi-scale generative adversarial network for single image dehazing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2760–2772, 2022.

[20] X. Zhang, J. Wang, T. Wang, and R. Jiang, "Hierarchical feature fusion with mixed convolution attention for single image dehazing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 2, pp. 510–522, 2022.

[21] X. Zhang, T. Wang, W. Luo, and P. Huang, "Multi-level fusion and attention-guided cnn for image dehazing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 11, pp. 4162–4173, 2021.

[22] D. Zhao, L. Xu, L. Ma, J. Li, and Y. Yan, "Pyramid global context network for image dehazing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 8, pp. 3037–3050, 2021.

[23] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3522–3533, Nov. 2015.