Self-supervised Learning and Adaptation for Single Image Dehazing

Yudong Liang[1], **Bin Wang**[1], **Wangmeng Zuo**[2], **Jiaying Liu**[3] and **Wenqi Ren**[4]
[1]School of Computer and Information Technology, Institute of Big Data Science and Industry, Shanxi University, Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi, China
[2]School of Computer Science at Harbin Institute of Technology, Haerbin, China
[3]Wangxuan Institute of Computer Technology, Peking University, Beijing, China
[4]School of Cyber Science and Technology, Shenzhen Campus, Sun Yat-sen University, Shenzhen, China
liangyudong@sxu.edu.cn, 202022407046@email.sxu.edu.cn, wmzuo@hit.edu.cn, liujiaying@pku.edu.cn, rwq.renwenqi@gmail.com

Abstract

Existing deep image dehazing methods usually depend on supervised learning with a large number of hazy-clean image pairs which are expensive or difficult to collect. Moreover, dehazing performance of the learned model may deteriorate significantly when the training hazy-clean image pairs are insufficient and are different from real hazy images in applications. In this paper, we show that exploiting large scale training set and adapting to real hazy images are two critical issues in learning effective deep dehazing models. Under the depth guidance estimated by a well-trained depth estimation network, we leverage the conventional atmospheric scattering model to generate massive hazy-clean image pairs for the self-supervised pre-training of dehazing network. Furthermore, self-supervised adaptation is presented to adapt pre-trained network to real hazy images. Learning without forgetting strategy is also deployed in self-supervised adaptation by combining self-supervision and model adaptation via contrastive learning. Experiments show that our proposed method performs favorably against the state-of-the-art methods, and is quite efficient, i.e., handling a 4K image in 23 ms. The codes are available at https://github.com/DongLiangSXU/SLAdehazing.

1 Introduction

Image dehazing [Tan, 2008; Chen *et al.*, 2021] is a challenging low-level vision problem with great theoretical and practical values. Under the hazy condition, the visibility decreases as the distance increases due to the light scattering effect, making hazy images often suffer from color shifts and poor contrast. Image dehazing has been a touchstone for machine learning algorithms and extensive methods [Cai *et al.*, 2016; Wu *et al.*, 2021] have been proposed and evaluated. Benefited from the development of deep networks, notable progress has been made in learning-based image dehazing models.

Supervised learning is usually adopted in deep image dehazing, but requires a large number of hazy-clean image pairs that are too expensive or difficult to collect. One conventional solution is to apply a synthetic dehazing dataset of which the hazy images are synthesized by leveraging the atmospheric scattering model [Israël and Kasten, 1959; Ren *et al.*, 2016] to generate haze images. In general, the data synthesis process applies the atmospheric light and the transmission map, which is generated according to the depth information. Due to the lack of depth information, the sizes of the existing synthetic datasets are limited, and the diversity of images is quite deficient. The real hazy images are quite diverse and hard to describe by a simple physical model. Thus, the synthesized images are always inconsistent with the real-world hazy images, and the trained deep models by synthesized data degrade heavily in the real hazy image cases. Although more complex deep architectures have been introduced for supervised dehazing, the generalization ability from synthetic data to real hazy images remains a bottleneck. Albeit a huge real hazy image dataset is beneficial to overcome the limitation of supervised dehazing, it is not available so far and alternative solution is thus required.

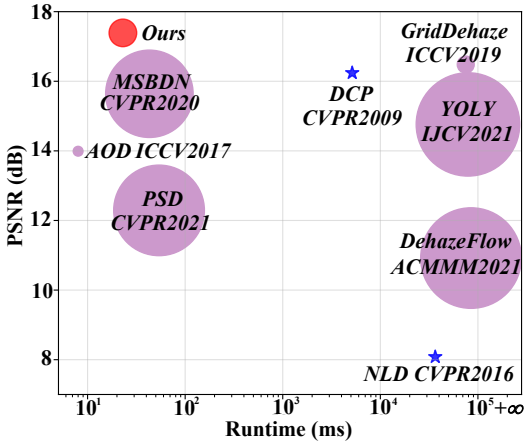

Figure 1: Visual comparisons of PSNR, runtime and parameter numbers on 4KID dataset. The area of the circle correlates with the parameter numbers except for the non-deep learning methods, which are indicated by stars.

To alleviate the domain shift between synthetic and real hazy images, single image dehazing methods with domain adaptation or unsupervised learning have been explored. Shao *et al*. [Yuanjie *et al.*, 2020] applied the translation network to reduce the domain gap between the synthetic and real domains. Nevertheless, the dehazing methods by domain adaptation still rely on the supervised data. The unsupervised image dehazing methods [Ulyanov *et al.*, 2018] usually measure the properties between dehazed images and hazy input. Unsupervised deep dehazing methods [Kar *et al.*, 2021; Li *et al.*, 2020] have exhibited better generalization ability. However, the dehazing results would easily get biased due to the lack of ground-truth hazy-free images, as the unsupervised loss would only describe some aspects of the restorations, such as restricting the contrast or lightness of the restorations. It is unreliable as no guidelines exist for the selections and combinations of the unsupervised loss for unsupervised image dehazing, giving some leeway for better formulating unsupervised image dehazing.

To alleviate the domain shift and bias issues of unsupervised dehazing learning, we propose a novel self-supervised learning and adaptation framework for single image dehazing. Benefited from the significant progress in single image depth estimation, the well-trained depth estimation model can be utilized to estimate the depth information of clear images. Under the depth guidance, we leverage the conventional atmospheric scattering model to generate the hazy images, which are then utilized for self-supervised learning to restore the corresponding clear images. Our self-supervised learning makes it feasible to exploit various clear images to construct large scale training set, and can achieve state-of-the-art performance with a lightweight dehazing network. To make a step further, real-world hazy images are utilized to adapt the deep dehazing network in a self-supervised manner, which stitches the generated hazy images and real hazy images together to restore the stitched corresponding dehazed images via contrastive learning. Better visual results and generalization ability are achieved without changing the dehazing model. Note that both stages get rid of the dependence on the real hazy-clean image pairs, which brings better adaptation ability for the data diversity. Benefited from our self-supervised image dehazing framework, the proposed lightweight network outperforms the state-of-the-art unsupervised image dehazing performances and even surpasses the recent supervised image dehazing methods as Figure 1. Besides, our model is quite efficient which can handle 4K images in $23ms$. The codes are available at https://github.com/DongLiangSXU/SLAdehazing. To summarize,

- We propose a two-stage self-supervised learning and adaptation framework for image dehazing, which includes a self-supervised learning stage under depth guidance and a self-supervised adaptation stage.
- In the self-supervised adaptation stage, we incorporate learning without forgetting with model adaptation via contrastive learning for adapting pre-trained dehazing network to real hazy images.
- Our proposed method outperforms the state-of-the-art methods and is lightweight and quite efficient, *i.e*., handling a 4K image within 23 ms.

2 Related Work

The most widely applied physical model in image dehazing is the atmospheric scattering model [He *et al.*, 2010], which explains the formulation of the hazy image $I(x)$ as:

$$I(x) = J(x)t(x) + A(1 - t(x)), \tag{1}$$

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

where $J(x)$ denotes the hazy-free image, and A denotes the global atmospheric light. Due to the light scattering effect, the hazy image $I(x)$ can be obtained as the spatially variant combination of $J(x)$ and A guided by the transmission map $t(x)$. x indicates the location and the parameter β is the scattering coefficient of the atmosphere for a certain image. The transmission map $t(x)$ decides the portion of the light that reaches the camera, which is attenuated as the distance of depth $d(x)$ increases. However, the transmission map and the global atmospheric light are necessary for synthesizing hazy images, but cannot be acquired in real applications.

Most recently, unsupervised deep image dehazing methods have been extensively studied. Golts *et al*. [Golts *et al.*, 2019] proposed to minimize the Dark Channel Prior (DCP) energy function for unsupervised dehazing. Li *et al*. [Li *et al.*, 2020] explained a "zero-shot" image dehazing framework that just uses the observed single hazy image for learning and inference. Li *et al*. [Li *et al.*, 2021a] performed unsupervised image dehazing without image collection, and obtained promising performances, however, the processing speed was too slow to meet the real-time requirements. Chen *et al*. [Chen *et al.*, 2021] started from a well pre-trained dehazing model, then applied the real hazy images to finetune the model in an unsupervised manner. In contrast, our method proposes a fully unsupervised image dehazing process by incorporating self-supervised learning with self-supervised adaptation.

3 Proposed Method

As shown in Figure 2, we propose a two-stage self-supervised image dehazing framework including a self-supervised learning stage under estimated depth guidance and an self-supervised adaptation stage for adapting to real hazy images. Both stages share the same model that applies to a hazy image generation network under the estimated depth guidance and a physical model-based dehazing network. In this section, we first describe the hazy image generation, and then present the details of the two stages.

3.1 Hazy Image Generation

Our method does not require any datasets of hazy-clear pairs, but collects various clear images for training. In particular, the proposed hazy image generation method applies a random global atmospheric light A and a scattering coefficient β as well as the depth information to generate the hazy image. For the existing benchmarks, the acquisition of the image depth information mainly relies on professional equipment and costs relatively high labors and time, which greatly limits the sizes of the benchmarks. Considering that significant

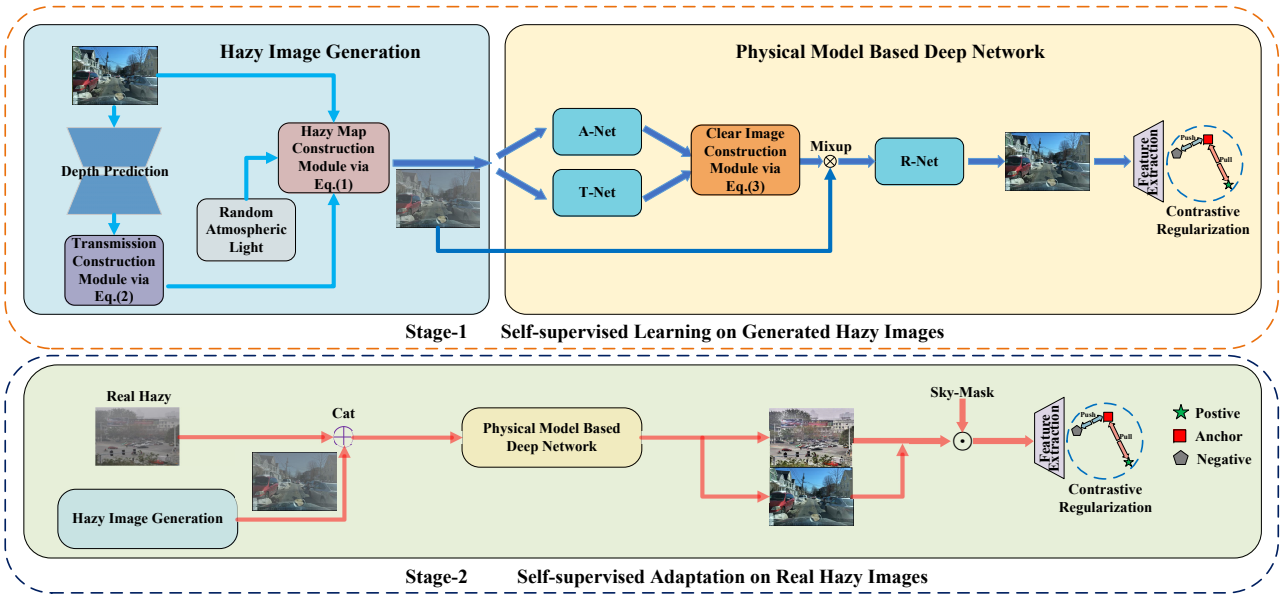

Figure 2: The architecture of the proposed two-stage framework.

progress has been made in single image depth prediction, the result by off-the-shelf depth prediction model [Alhashim and Wonka, 2018] can provide a reasonable estimation of image depth information, which can be then employed to guide hazy image generation. We note that both the estimated depth and the atmospheric scattering model are not precise in characterizing the formation of real hazy images. Nonetheless, with the estimated depth and the model in Eqn. (1), we can generate sufficient number of synthetic hazy-clean image pairs. Moreover, our physical model-based dehazing network and self-supervised adaptation are beneficial to alleviate the effect of domain shift between synthetic and real hazy images. In this work, we collect 2,000 images of diverse contents from websites to generate hazy images, which is surprisingly effective for self-supervised dehazing. The proposed hazy image generation is fixed during the training process, which is quite efficient and can be safely abandoned during the inference.

3.2 Physical Model Based Dehazing Network

Following the atmospheric scattering model [He *et al.*, 2010], we design a physical model-based dehazing network involving three major modules, *i.e.*, T-Net, A-Net, and R-Net. As Figure 2, both T-Net and A-Net take a hazy image as the input to estimate the transmission map and global atmospheric light, respectively. Given the estimated transmission map t and global atmospheric light A, the preliminary estimation can then be obtained based on atmospheric scattering model,

$$\hat{J}(x) = \frac{I - A}{t} + A. \tag{3}$$

However, estimation error is inevitable when estimating t and A by T-Net and A-Net. Artifacts, color shifts, or some remaining hazes may be introduced in the preliminary estimation $\hat{J}(x)$. Thus, a refinement module denoted as R-Net is further applied to better predict the dehazed images by incorporating the generated hazy image and preliminary dehazing result. The adaptive mixup operation [Wu *et al.*, 2021] is utilized to fuse the generated hazy images and preliminary dehazing results $\hat{J}(x)$, which can preserve the information from the hazy images and avoid the bias introduced by the preliminary dehazing result $\hat{J}(x)$.

For the T-Net and A-Net, we adopt similar encoder-decoder network structures. The encoder network includes five convolution layers to extract the features with reduced sizes, while the decoder network contains feature upsampling and mapping processes. Two deconvolution layers are utilized to upsample the feature maps and three convolution layers are applied. Feature fusion attention (FFA) [Qin *et al.*, 2019] blocks are applied for the feature mapping process in the middle of the encoder and decoder. Considering that it is more difficult to estimate the transmission map than the global atmospheric light, six FFA blocks are applied for the transmission map estimation and three FFA blocks are used for the global atmospheric light estimations. The adaptive mixup operation [Wu *et al.*, 2021] for the generated hazy images and preliminary dehazing result is given by:

$$Mix(I(x), \hat{J}(x)) = \sigma(\theta) * I(x) + (1 - \sigma(\theta) * \hat{J}(x) \tag{4}$$

where θ is a learnable parameter to control the fusion of the hazy map $I(x)$ and the preliminary estimation $\hat{J}(x)$. The number of channels is 32 for the encoder-decoder network which is quite efficient.

The refinement R-Net adopts a lightweight U-Net [Ronneberger *et al.*, 2015] structure with the number of channels to be 64. Nine FFA blocks are then applied to decode the fused features to get the refined dehazing result $\hat{J}_r(x)$. In the following, we describe the self-supervised learning and adaption for training the physical model based dehazing network.

3.3 Self-supervised Learning with Synthesis Data

The first stage, *i.e.*, self-supervised learning based image dehazing, applies the physical model based dehazing network to the generated hazy images. The L_1 reconstruction losses are deployed to the estimated transmission map and dehazing result. In addition, contrastive learning [Wu *et al.*, 2021] is utilized to make the dehazing result more similar to the hazy-free image other than the hazy image. The overall learning objective is then given as,

$$L = L_t(\hat{t}, t) + L_J(\hat{J}_r, J) + \lambda \cdot CR(f(J), f(\hat{J}_r), f(I)), \quad (5)$$

where the position index x is omitted for simplicity, t, J, I respectively denote the generated transmission map, the hazy-free image and the synthesised hazy image. CR is the contrastive loss, λ is a hyper-parameter and f extracts features for measurements. Specifically, the contrastive loss is defined on the intermediate features from the fixed pre-trained model *e.g.*, VGG features, to pull the anchors *i.e.*, dehazing results, towards the positive examples *i.e.*, hazy-free images, and to push away from the negative example, *i.e.*, hazy images. Thus, it is defined as,

$$CR(J, \hat{J}_r, I) = \sum_k \omega_k \cdot \frac{D(f_k(J), f_k(\hat{J}_r))}{D(f_k(I), f_k(\hat{J}_r))}, \quad (6)$$

where f_k denotes the k-th layer feature extracted using the fixed pre-trained model, ω_k denotes the weight and D denotes the metric to measure the difference.

Note that we do not impose the reconstruction loss on the preliminary dehazing result $\hat{J}$, thereby allowing more tolerance and flexibility on the preliminary estimation by T-Net. Fortunately, R-Net is introduced to correct the estimation error of the preliminary dehazing result and to be robust against the potential bias with the atmospheric scattering model or depth estimation. Benefited from the network architecture and learning objective, our whole dehazing network can achieve reasonable generalization ability to real hazy images which can not be fully described by the atmospheric scattering model.

3.4 Self-supervised Adaptation on Real Hazy Images

The second stage aims to adapt the learned model to better handle real hazy images. Due to the lack of ground-truth clear images, existing unsupervised image dehazing methods may get biased toward the applied unsupervised loss. Multiple unsupervised losses are empirically combined by the existing methods [Li *et al.*, 2020; Chen *et al.*, 2021], as single unsupervised loss only describes a certain aspect of the images, such as the lightness or contrast. However, it is difficult to find a suitable combination of unsupervised losses for avoiding the bias between supervised and unsupervised losses. Instead, we suggest to incorporate a learning without forgetting strategy with contrastive learning for self-supervised model adaptation. In particular, a real hazy image and a generated hazy image are concatenated to be fed into the dehazing network. Then, contrastive learning is performed on the dehazing result for self-supervised adaptation. Thanks to the concatenation of generated and real hazy images, learning without forgetting can be enforced during the adaptation process, largely alleviating the learning drifts issue. As no reconstruction loss is applied, some quantitative assessment metrics (*e.g.*, PSNR) may decrease after the second stage of training. Nonetheless, the visual quality on real hazy image can be significantly improved.

To apply contrastive learning, we utilize Contrast Limited Adaptive Histogram Equalization (CLAHE) to generate positive examples for real hazy images. Chen *et al.* [Chen *et al.*, 2021] also suggested not to treat CLAHE results as supervision directly since it may introduce the inherent flaws of CLAHE to the learned model. Thus, we use the results by CLAHE on real hazy image as the positive examples in the contrastive loss,

$$L_{wf} = CR(\{J_s, J_{CLAHE}\}, \{\hat{J}_s, \hat{J}_{real}\}, \{I_s, I_r\}). \quad (7)$$

where $\{I_s, I_r\}$ denotes the concatenation of the generated hazy image and real hazy image as negative examples. $\{\hat{J}_s, \hat{J}_{real}\}$ denotes the dehazing results of the generated and real hazy images, and $\{J_s, J_{CLAHE}\}$ denotes the corresponding positive examples. Moreover, the sky regions of real images indicated by 'Sky-Mask' in Figure 2 are not adapted by the real hazy images as the CLAHE method always fails in the sky regions. And the sky mask is generated by roughly estimating the dark channel prior [He *et al.*, 2010].

4 Experiments

We collect 2,000 pictures of diverse contents from websites to generate the hazy images, which proves to be surprisingly effective for self-supervised learning of dehazing network. Real hazy images from URHI (Unannotated Real Hazy Images) of RESIDE [Li *et al.*, 2018] are applied for self-supervised adaptation stage, which has complex scenes and a wide range of brightness. The images are randomly cropped to 256×256 and normalized to $0 \sim 1$. Our method is evaluated on two synthetic datasets, *i.e.*, SOTS and 4KID [Zheng *et al.*, 2021], as well as a dataset of real hazy images, *i.e.*, URHI [Li *et al.*, 2018]. Following [Li *et al.*, 2021a], the SOTS of the RESIDE v0 dataset in our experiments contains "SOTS-indoor" and "SOTS-outdoor" dataset, *i.e.*, 500 indoor hazy images and 500 outdoor hazy images. 4KID is a dataset containing large-size 4K (*i.e.*, 3840×2160) synthetic hazy images established by [Zheng *et al.*, 2021]. According to [Zheng *et al.*, 2021], we randomly selected 200 hazy images in the dataset for testing. We also test the running time of the competing dehazing methods on this dataset.

The proposed method is implemented with Pytorch 1.4.0 and experiments are conducted on a PC with one NVIDIA TITAN XP GPU. The models are trained by Adam optimizer with exponential decay rates β_1 and β_2 of 0.9 and 0.999, respectively. The initial learning rate and batch size are set to 0.0002 and 8, respectively. In the first training stage, the cosine annealing strategy is applied to adjust the learning rate and the total number of iterations is 50k. As for self-supervised adaptation, only 5k iterations are required. In general, the dehazing network after the first stage of self-supervised learning can achieve the state-of-the-art quantitative performance in terms of PSNR or SSIM. In contrast, the second stage of self-supervised adaptation is effective in

Figure 3: Visual comparisons of different methods on the SOTS dataset.

Figure 4: Visual comparison of different dehazing methods on the 4KID dataset.

	SOTS-indoor		SOTS-outdoor	
	PSNR	SSIM	PSNR	SSIM
Haze	11.44	0.3734	12.56	0.3695
DCP [He *et al.*, 2010]	16.62	0.8179	17.16	0.8514
AOD [Li *et al.*, 2017]	19.06	0.8504	19.58	0.8711
NLD [Berman *et al.*, 2016]	17.29	0.7489	13.06	0.7449
YOLY [Li *et al.*, 2021a]	19.41	0.8327	22.41	0.9023
PSD [Chen *et al.*, 2021]	12.45	0.6812	15.42	0.7624
Ours-Stage1	**20.49**	**0.8578**	**24.33**	**0.9323**

Table 1: Comparison of the quantitative performance by different dehazing methods on the SOTS dataset

4KID	PSNR / SSIM	T(ms)	Parms(Mb)
Haze	10.14 / 0.7230	-	-
DCP[He *et al.*, 2010]	16.24 / 0.8280	5150	-
AOD[Li *et al.*, 2017]	13.99 / 0.7907	**5**	**0.002**
NLD[Berman *et al.*, 2016]	8.08 / 0.5661	36540	-
GridDehaze[Liu *et al.*, 2019]	16.48 / 0.8174	-	0.96
MSBDN[Dong *et al.*, 2020]	15.64 / 0.8498	43	31.35
DehazeFlow[Li *et al.*, 2021b]	10.92 / 0.7037	-	39.61
YOLY[Li *et al.*, 2021a]	14.76 / 0.8338	-	41.53
PSD[Chen *et al.*, 2021]	12.29 / 0.6651	54	33.11
Ours-Stage1	**17.39 / 0.8392**	23	3.11

Table 2: Comparison of the generalization ability and running time by different methods on the 4KID dataset, '-' indicates non-deep learning methods or methods cost too much GPU storage to execute.

improving the visual quality and generalization to real hazy images, but may result in lower PSNR or SSIM values on the synthetic hazy images.

4.1 Results on Synthetic Datasets

The performances of different methods on SOTS dataset are provided in Table 1 and Figure 3, It can be seen that our method outperforms the competing methods with a large margin. Several state-of-the-art supervised methods are also tested for the adaptivity or generalization ability of the model in Table 2 and Figure 4, namely, GridDehaze [Liu *et al.*, 2019], MSBDN [Dong *et al.*, 2020], and DehazeFlow [Li *et al.*, 2021b]. The performances, runing time and parameters of different methods are listed in Table 2. Our method achieves the best quantitative performances with much fewer parameters and is very efficient, *i.e.*, 23ms for a 4K image.

4.2 Results on Real Hazy Datasets

To evaluate our method on the real hazy images, we conduct experiments on the URHI benchmark established by [Li *et al.*, 2018]. The comparison results are shown in Figure 5. In particular, severe color shifts can be observed from the results by DCP and NLD, while supervised methods suffer from the problem of incomplete removal of hazes in the distant areas or resulting in dark dehazing results. Moreover, dark regions can also be observed in the results by YOLY, and the iterative optimization strategy restricts its application in real-time scenarios. PSD is the latest unsupervised dehazing method usually producing visually pleasing results. Our method can compete with PSD and achieve better dehazing results in some distant areas. Instead of resorting to the well trained supervised model as PSD, our method does not require any supervised image dehazing model and existing synthetic dataset. Moreover, our method is quite efficient in inference and has much fewer model parameters.

4.3 Ablation Study

Effect of Dehazing Network Architectures

Three types of architectures are investigated in Table 3: (a) An end-to-end network without following the physical model, (b) a network based on the physical model without the refine-

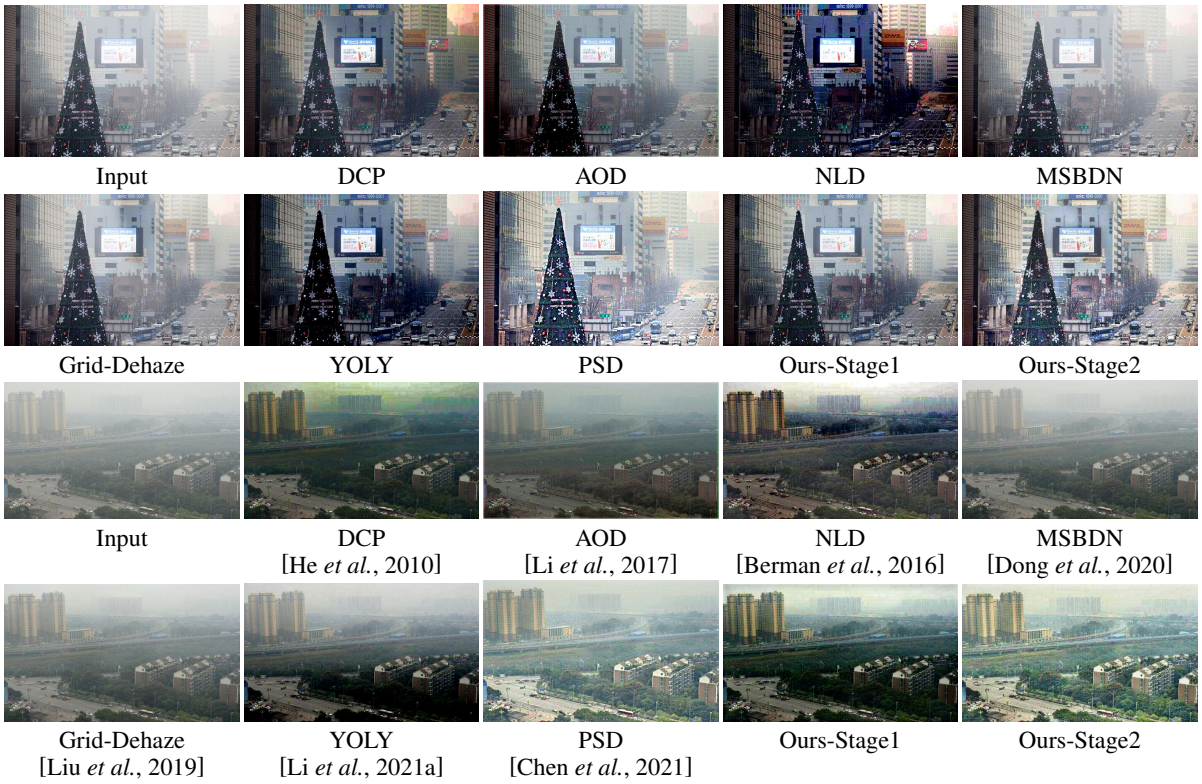

Figure 5: Visual comparison of different dehazing methods for real hazy images.

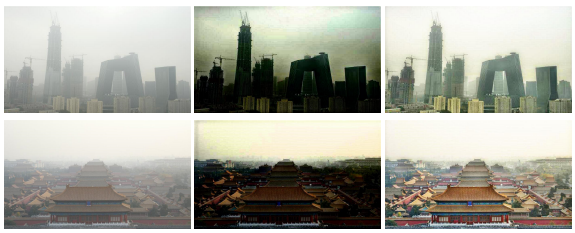

Figure 6: Visual comparison of dehazing results for real hazy images before and after self-supervised adaptation.

	PSNR	SSIM
(a) w/o the physical model	21.67	0.8822
(b) w/o R-Net	20.02	0.8668
(c) the proposed	22.41	0.8950

Table 3: Performance comparison on SOTS for dehazing networks with different architectures

ment module R-Net and (c) our proposed structure. From Table 3, one can see that it is critical to apply both the physical model and the error refinement module for attaining promising dehazing performance. Besides, visual comparison is provided in the supplementary material. Without the physical model, the removal of hazes are incomplete. Without refinement module, severe color shifts and artifacts may happen.

Effect of Self-supervised Adaptation

The visual restorations before and after self-supervised adaptation are compared for real hazy images in Figure 6, denoted as "Stage1" and "Stage2" respectively. As real hazy images have a wider range of lighting changes, the dehazing results by the first stage which only applies clear images to synthesize hazy images are opt to generate incorrect lightness, resulting in dark images. In comparison, self-supervised adaptation is effective in addressing this issue and improving the dehazing performance on real hazy images.

5 Conclusion

In this paper, a self-supervised learning and adaptation framework has been proposed to alleviate the bias and domain shift issues in unsupervised image dehazing. Our proposed method involves two stages of training, *i.e.*, (i) self-supervised image dehazing under depth guidance and (ii) self-supervised adaptation for real hazy images. In the second stage of training, the learning without forgetting strategy is combined with self-supervised learning and model adaptation via contrastive learning to avoid the bias issues for real hazy images. Our method performs favorably against the state-of-the-art methods and is quite efficient in inference, *i.e.*, $23ms$ for a 4K image.

Acknowledgments

This work is partially supported by National Natural Science Foundation of China (Nos. 61802237, U19A2073, 621724090), the Natural Science Foundation of Shanxi

Province (201901D211176), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (STIP) (2019L0066), the Science and Technology Major Project of Shanxi Province (202101020101019), the Key R&D program of Shanxi Province (International Cooperation, 201903D421041).

References

[Alhashim and Wonka, 2018] Ibraheem Alhashim and Peter Wonka. High quality monocular depth estimation via transfer learning. *arXiv e-prints*, abs/1812.11941, 2018.

[Berman *et al.*, 2016] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016.

[Cai *et al.*, 2016] 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.

[Chen *et al.*, 2021] Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. Psd: Principled synthetic-to-real dehazing guided by physical priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7180–7189, 2021.

[Dong *et al.*, 2020] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2157–2167, 2020.

[Golts *et al.*, 2019] Alona Golts, Daniel Freedman, and Michael Elad. Unsupervised single image dehazing using dark channel prior loss. *IEEE transactions on Image Processing*, 29:2692–2701, 2019.

[He *et al.*, 2010] 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, 2010.

[Israël and Kasten, 1959] Hans Israël and Fritz Kasten. Koschmieders theorie der horizontalen sichtweite. In *Die Sichtweite im Nebel und die Möglichkeiten ihrer künstlichen Beeinflussung*, pages 7–10. Springer, 1959.

[Kar *et al.*, 2021] Aupendu Kar, Sobhan Kanti Dhara, Debashis Sen, and Prabir Kumar Biswas. Zero-shot single image restoration through controlled perturbation of koschmieder's model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16205–16215, 2021.

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

[Li *et al.*, 2018] 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, 2018.

[Li *et al.*, 2020] Boyun Li, Yuanbiao Gou, Jerry Zitao Liu, Hongyuan Zhu, Joey Tianyi Zhou, and Xi Peng. Zero-shot image dehazing. *IEEE Transactions on Image Processing*, 29:8457–8466, 2020.

[Li *et al.*, 2021a] Boyun Li, Yuanbiao Gou, Shuhang Gu, Jerry Zitao Liu, Joey Tianyi Zhou, and Xi Peng. You only look yourself: Unsupervised and untrained single image dehazing neural network. *International Journal of Computer Vision*, 129(5):1754–1767, 2021.

[Li *et al.*, 2021b] Hongyu Li, Jia Li, Dong Zhao, and Long Xu. Dehazeflow: Multi-scale conditional flow network for single image dehazing. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 2577–2585, 2021.

[Liu *et al.*, 2019] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7314–7323, 2019.

[Qin *et al.*, 2019] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. *arXiv preprint arXiv:1911.07559*, 2019.

[Ren *et al.*, 2016] 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.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

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

[Ulyanov *et al.*, 2018] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9446–9454, 2018.

[Wu *et al.*, 2021] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021.

[Yuanjie *et al.*, 2020] Shao Yuanjie, Li Lerenhan, Ren Wenqi, Gao Changxin, and Sang Nong. Domain adaptation for image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2020.

[Zheng *et al.*, 2021] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Xiaobin Hu, Tao Wang, Fenglong Song, and Xiuyi Jia. Ultra-high-definition image dehazing via multi-guided bilateral learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16185–16194, 2021.