

Zero-Shot Image Dehazing Using Pseudo Atmospheric Light Image

Eunsung Jo^{*}, Eunpil Park[†], and Jae-Young Sim^{*†}

^{*} Graduate School of Artificial Intelligence, UNIST, Ulsan, Korea

[†] Department of Electrical Engineering, UNIST, Ulsan, Korea

E-mails: {esjo93, cosmos, jysim}@unist.ac.kr

Abstract—Abundant training data for deep neural networks improve the performance of single image dehazing (SID) substantially, however they suffer from the domain-shift problem of the discrepancy between the training set and the test set. In this paper, we propose a zero-shot SID method to overcome the domain-shift problem in a self-supervised learning manner. We employ two generator networks to estimate the transmission and the original scene radiance, respectively, from an input hazy image. We also synthesize the pseudo atmospheric light image (PALI) to train the transmission generator to assign zero transmission values to PALI. Since the pseudo-sky patches and the dense-hazy region rarely have the structural textures, the network learns the dense-hazy property from the PALI in a self-supervised learning manner. The experimental results show that the proposed method faithfully restores the scene radiance image, and the PALI loss is effective to train the deep neural network.

Index Terms—single image dehazing, zero-shot learning, self-supervised learning

I. INTRODUCTION

As shown in Figure 1, the particles floating in the air, e.g. haze and fog, cause the scattering and absorption of the light. Therefore, the images captured under the aforementioned environment exhibit low contrast, blurriness, and color distortion, since the scene radiance is attenuated along the light path to the camera and the atmospheric light unrelated with the scene is additionally observed. Single image dehazing (SID) is an ill-posed problem, which estimates the original scene radiance image from an input hazy image.

There are three approaches for SID: model-driven, domain-driven, and image-driven methods. The model-driven methods [1]–[3] solve the inverse problem of the image formation model (IFM) to estimate the scene radiance image from the hazy image. The IFM is defined as

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

where $I(\mathbf{x})$, $J(\mathbf{x})$, and $t(\mathbf{x})$ denote the observed intensity, the scene radiance, and the transmission at pixel \mathbf{x} , respectively. A means the atmospheric scattered light accumulated from infinitely far region in the hazy image, and $t(\mathbf{x}) = e^{-\beta d(\mathbf{x})}$ is the transmission, where β is the attenuation coefficient and $d(\mathbf{x})$ is the distance between the camera and the scene. The

This work was supported in part by the Samsung Research Funding and Incubation Center of Samsung Electronics under Project SRFC-IT1802-08, and in part by IITP grant funded by Korea government(MSIT) (No.2020-0-01336, Artificial Intelligence Graduate School Program(UNIST)).

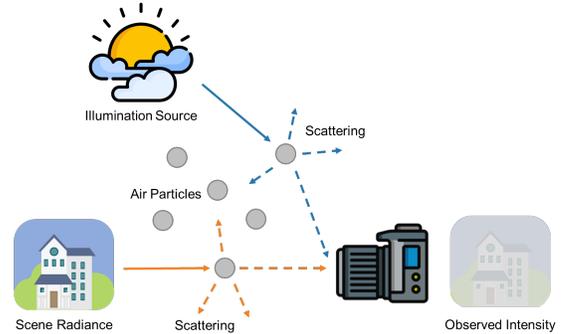


Fig. 1: The imaging model with haze.

model-driven SID methods estimate unknown parameters of A and t to recover J from I based on (1), and thus provide limited performance due to the mismatch between the imaging model and real situation.

Recently, deep neural networks have been making breakthrough for SID that learn arbitrary mapping functions from the hazy image domain to the clean image domain [4]–[8]. However, the domain-driven methods often suffer from the domain-shift problem between training image set and test image set. To overcome this problem, the image-driven methods were recently proposed that train deep neural networks using only an input hazy image without additional training datasets. But the existing methods do not completely utilize the advantages of abundant data [9], [10] or the prior knowledge of IFM [11].

In this paper, we propose an image-driven SID method based on the zero-shot and self-supervised learning framework. We design two generators to estimate the transmission map t and the scene radiance image J , respectively, from an input hazy image. We find the atmospheric light A and synthesize the pseudo atmospheric light image (PALI) of dense haze by adding the random noise to the image of the constant color of A . The PALI should be assigned the zero transmission values, since it has no structural textures and is assumed to be associated with infinitely far regions. Then we train the transmission generator with self-supervision such that the PALI is assigned small transmission values close to zero. Experimental results demonstrate that the proposed method yields better performance compared with the state-of-the-art SID methods.

II. RELATED WORKS

We briefly summarize the three approaches for SID.

A. Model-driven Approach

The model-driven methods solve the inverse problem of (1) to estimate the scene radiance $\mathbf{J}(\mathbf{x})$ such that

$$\mathbf{J}(\mathbf{x}) = \frac{\mathbf{I}(\mathbf{x}) - A(1 - \mathbf{t}(\mathbf{x}))}{\mathbf{t}(\mathbf{x})}. \quad (2)$$

Therefore the goal of the model-driven method is to estimate the unknown parameters of \mathbf{t} and A from \mathbf{I} . Several priors for the parameter estimation are used: the maximization of local image contrast or the image sharpness [2], the surface shading and the scene transmission are uncorrelated in local patches [3], and the dark channel prior (DCP) that the minimum intensity among the RGB channels of the scene radiance is generally zero [1].

B. Domain-driven Approach

To train deep neural networks, a dataset called RESIDE [12] has been synthesized from a RGBD image dataset based on the IFM in (1). To overcome the domain gap problem, recent works employed real hazy images by simulating the haze using fog generators [13]–[16]. The supervised learning methods require paired datasets of hazy and clean images. The supervised learning methods estimate the parameters of the IFM following the model-driven methods [4], [5], or directly estimate the scene radiance from an input hazy image [6]–[8]. However, the paired image datasets are hard to be obtained in general. Goltz *et al.* [17] proposed the unsupervised learning method by employing the DCP prior as a loss function [1]. However, the discrepancy between the domain of training images and the domain of test images causes the domain-shift problem, and therefore the domain-driven methods often cause limited performance on test images.

C. Image-driven Approach

Recently, the image-driven methods, or the zero-shot learning methods, have been proposed to avoid the domain-shift problems, that train deep neural networks using the input hazy image only without any training dataset. The Double-DIP (DDIP) [9] solves the low-level vision problems with the layer decomposition methodology. Inspired by that the entropy of the blended image with two different images is higher than the sum of the entropy of each image, Gandelsman *et al.* established two separate generators where each reconstructs an exclusive individual image. The Zero-shot Image Dehazing (ZID) [10] combines various prior-based losses, such as the DCP. The Zero Restore (ZR) [11] synthesizes the hazy images using the target hazy image and the IFM (1).

III. PROPOSED METHOD

We propose a SID method based on the zero-shot and self-supervised learning framework. The overall architecture of the proposed method is described in Figure 2. We first estimate the transmission map \mathbf{t} and the scene radiance image \mathbf{J} from

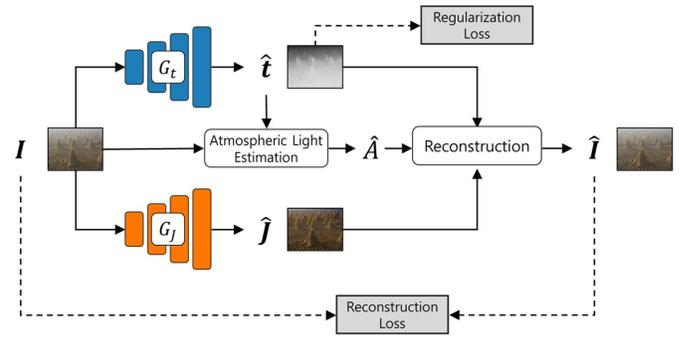


Fig. 2: Overall architecture of the proposed method.

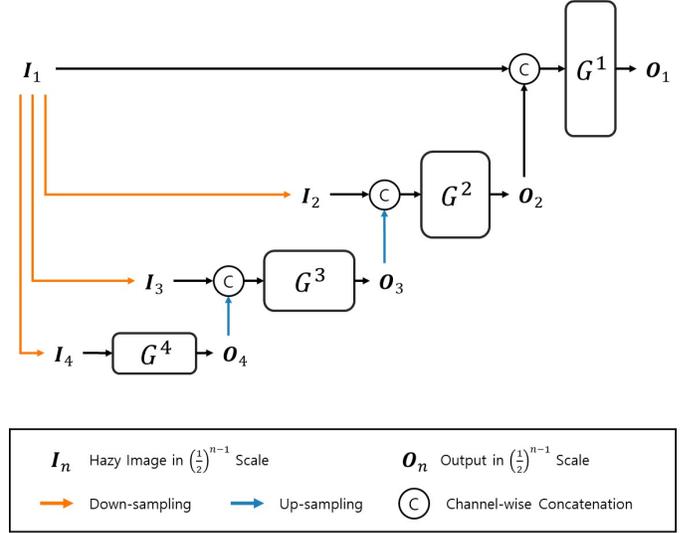


Fig. 3: The generator structure. G^i 's are the sub-networks at multiple scales.

an input hazy image \mathbf{I} using two generators of G_t and G_J , respectively, and estimate the atmospheric light \hat{A} . Then we generate $\hat{\mathbf{I}}$ according to the IFM (1), which is then compared to the input image \mathbf{I} in the unsupervised manner. We also synthesize PALI to train G_t in the self-supervised manner, assuming that PALI is useful to train the zero-shot SID since A is similar to $\mathbf{I}(\mathbf{x})$ with extremely low $\mathbf{t}(\mathbf{x})$.

A. Training by Parameter Estimation

The two generators of G_t and G_J have the identical structure, where the detailed structure is illustrated in Figure 3. It is composed of the four sub-networks at different scales, and each sub-network has the four modules of *Conv-BN-LReLU*.

The atmospheric light A is estimated from the input image by also using the estimated transmission map. Since the atmospheric light A is the additive intensity due to the scattering of light accumulated from the infinitely far region, we assume the region of A to be the brightest region in a hazy image and assigned zero transmission values. Specifically, we find \hat{A}

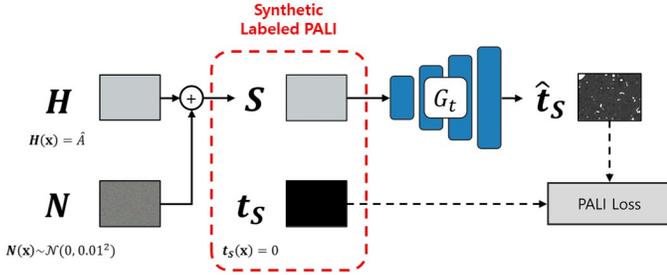


Fig. 4: Self supervision by using the pseudo atmospheric light image.

as the maximum intensity at pixel \mathbf{x} among the pixels having the lowest 1% values of the transmission.

To train the network, we employ the reconstruction loss \mathcal{L}_{rec} and the regularization loss \mathcal{L}_{reg} . Note that the generators G_t and G_J have the results at four different scales. For the simplicity, we define the loss functions first and apply them equivalently to each scale. We reconstruct a hazy image \hat{I} from the estimated \hat{J} , \hat{A} and \hat{t} based on the IFM in (1). The reconstruction loss is computed as

$$\mathcal{L}_{\text{rec}} = \frac{1}{N} \sum_{\mathbf{x}} \sum_c F(I_c(\mathbf{x}) - \hat{I}_c(\mathbf{x})), \quad (3)$$

where N is the number of pixels and F is the smoothed L1 distance given by

$$F(e) = \begin{cases} 0.5e^2, & \text{if } |e| < 1. \\ |e| - 0.5, & \text{otherwise.} \end{cases} \quad (4)$$

We also assume smooth transmission such that the distance $d(\mathbf{x})$ and the attenuation coefficient β is homogeneous in an image, and define the regularization loss as

$$\mathcal{L}_{\text{reg}} = \frac{1}{N} \sum_{\mathbf{x}} \left\{ \|\nabla_x \hat{t}(\mathbf{x})\|_1 + \|\nabla_y \hat{t}(\mathbf{x})\|_1 \right\}, \quad (5)$$

where ∇_x and ∇_y denote the partial derivative operators.

B. Pseudo Atmospheric Light Image

The transmission estimator G_t may assign small transmission values to foreground objects with the similar colors to A . We propose to synthesize the PALI S using the estimated atmospheric light and train the network using the PALI in a self-supervised learning manner as shown in Figure 4. Specifically, we first construct an image H by setting the intensity of all the pixels to A . Then we select random noises from the Gaussian distribution with zero mean and the standard deviation of 0.01 which are then added to H to obtain the PALI S .

Based on the assumption that the desired transmission values for the PALI are almost zero, we train G_t to generate a transmission image \hat{t}_S corresponding to the PALI S such that \hat{t}_S is encouraged to be close to zero transmission image t_S . Note that G_t is trained to avoid estimating the structured pixel regions with the similar colors to A as the far region, since the PALI has no structured textures such as edges or corners.

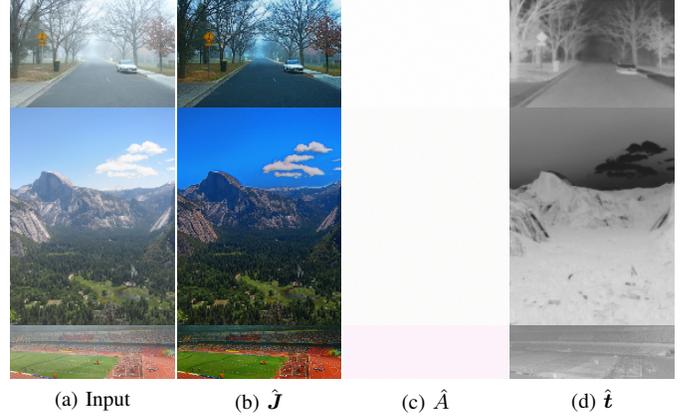


Fig. 5: The results of parameter estimation.

To this end, we design the PALI loss given by

$$\mathcal{L}_{\text{PALI}} = \frac{1}{N} \sum_{\mathbf{x}} F(\hat{t}_S(\mathbf{x})), \quad (6)$$

where \mathbf{x} is the pixel and $\hat{t}_S = G_t(S)$. Note that the PALI S is regarded as the pure haze image with no valid scene radiance information, and therefore G_t can learn the characteristics of the haze more faithfully via $\mathcal{L}_{\text{PALI}}$. Finally, the total loss is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{rec}} + \lambda_1 \mathcal{L}_{\text{PALI}} + \lambda_2 \mathcal{L}_{\text{reg}}, \quad (7)$$

where λ_1, λ_2 are the hyper-parameters empirically set to 10^{-2} .

IV. EXPERIMENTAL RESULTS

A. Experiment Setup

a) *Datasets*: We use the unpaired real hazy images from <https://www.wisdom.weizmann.ac.il/~vision/BlindDehazing/> and the paired images from the test set of the I-HAZE [13] and O-HAZE [14], which have gaps from the model of (1) and are available for the quantitative evaluation as well.

b) *Training details*: At the beginning, we trained G_J to reconstruct I from I for the network parameter initialization purpose. After that, all networks are trained with $\mathcal{L}_{\text{total}}$ simultaneously. We used the Adam [18] optimizer to train the networks, and set the initial learning rate as 5.0×10^{-4} , which is then decayed by half for every 100 iterations. Due to the computational complexity, we down-sampled the images by 4 for the I-HAZE [13] and O-HAZE [14] images. We implemented the experiments with the PyTorch on a single TITAN RTX.

B. Results of the Proposed Method

a) *Parameter Estimation Results*: We show the results of the parameter estimation in Figure 5. The input hazy images have the low contrast and the color distortion problems, which are alleviated in the scene radiance images \hat{J} as shown in Figures 5(a) and (b). The other parameters associated with the IFM, \hat{A} and \hat{t} , are also reliably estimated as shown in Figures 5(c) and (d).

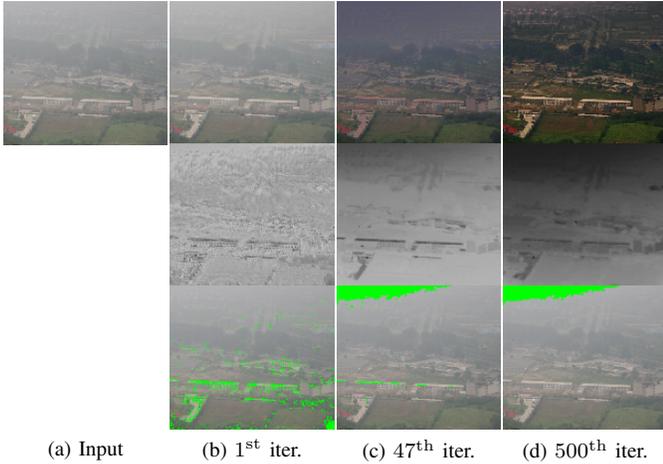


Fig. 6: The intermediate results of the parameter estimation during the training. (a) An input hazy image, and the intermediate results of the proposed method at the (b) 1st, (c) 47th, and (d) 500th iterations, respectively. From top to bottom, each row represents \hat{J} , \hat{t} , and the candidate regions for A .

Loss Function	I-HAZE		O-HAZE	
	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow
$\mathcal{L}_{\text{total}}$	15.7308	0.7048	17.8736	0.6754
w/o $\mathcal{L}_{\text{PALI}}$	15.1986	0.6951	17.6011	0.6424

TABLE I: The quantitative effect of the PALI loss.

b) Progressive Performance Improvement: We synthesize the PALI using the atmospheric light A and random Gaussian noise to learn the haze characteristics more faithfully. We also show the progressive improvement of performance of the transmission estimator G_t through the training iterations in Figure 6. At the 1st step, some non-sky regions are selected as the candidate region for A , and therefore the estimated transmission map \hat{t} becomes unreliable as shown in Figure 6(b). The main reason is that even the non-sky regions are bright and homogeneous having similar colors to the initial atmospheric light. Nevertheless, as we train the transmission estimator G_t with the synthesized PALI, the performance of G_t is incrementally improved and a correct sky-region is selected as the candidate region, then consequentially the quality of the estimated scene radiance is gradually improved.

c) Effect of PALI Loss: We conduct the ablation study on the proposed PALI loss. Figure 7 compares the qualitative results. When removing the PALI loss $\mathcal{L}_{\text{PALI}}$, the haze in the top region are not removed and the transmission values in that region are underestimated as shown in Figure 7(c). As we discussed in section IV-B0b, the synthetic PALI leads to the better performance of the transmission estimation.

Table I also shows the quantitative scores tested on I-HAZE [13] and O-HAZE [14], where we see that the total loss $\mathcal{L}_{\text{total}}$ including the PALI loss achieves the best performance.

C. Comparison With Existing Methods

In this section, we compare the proposed method with the

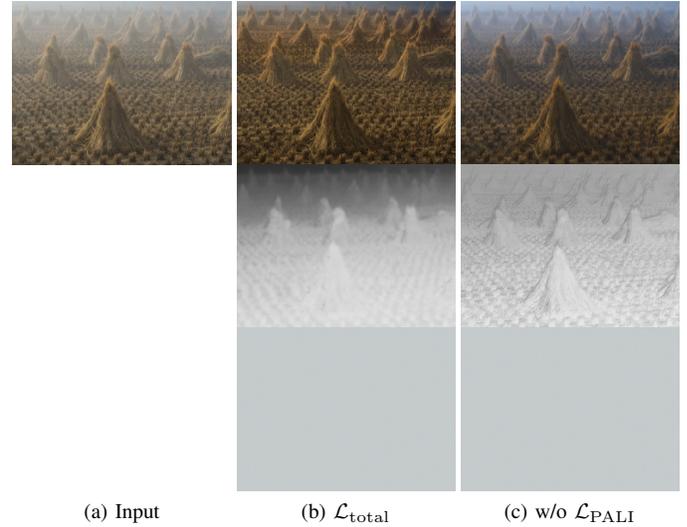
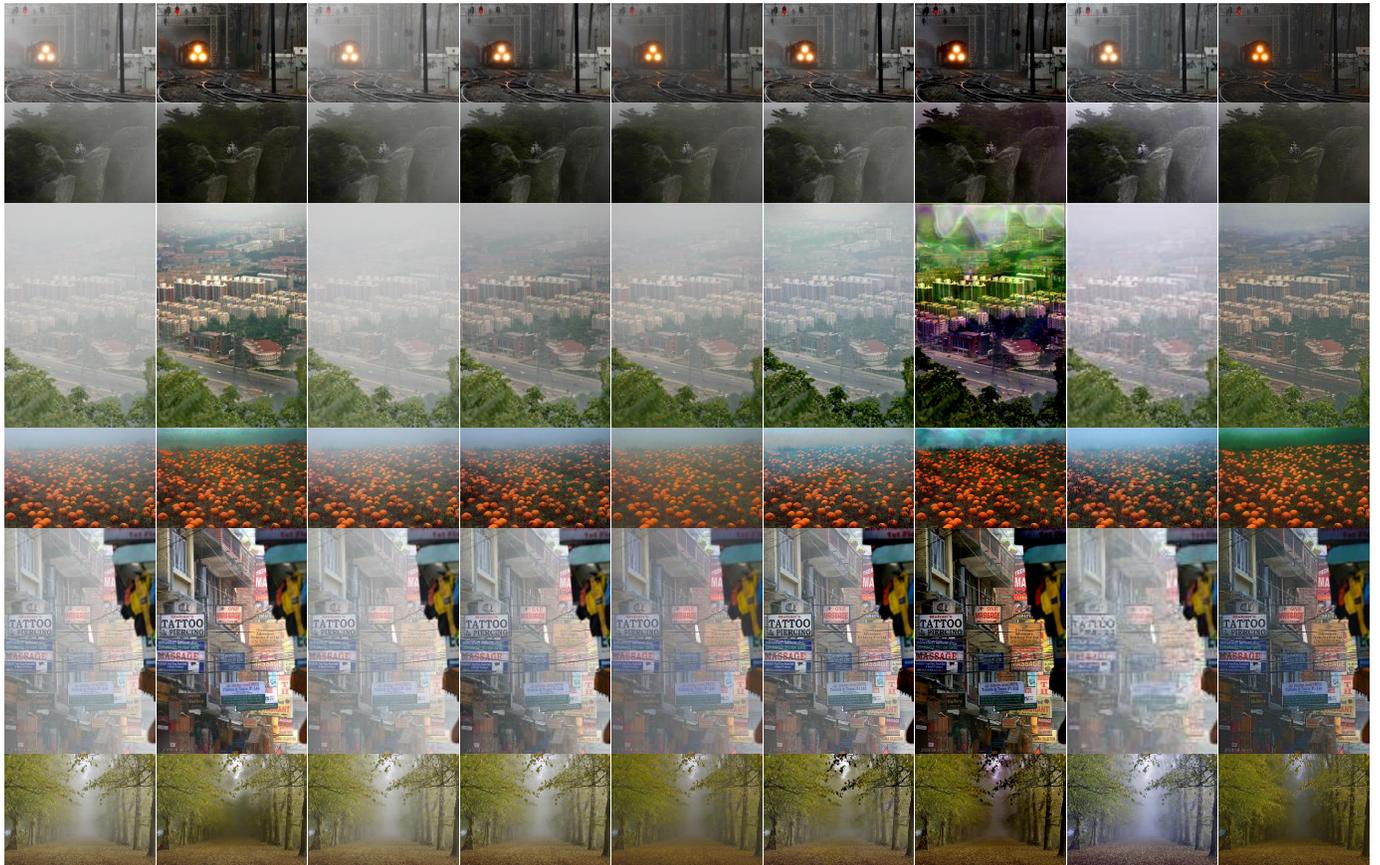


Fig. 7: The qualitative effect of the proposed PALI loss. (a) An input hazy image I . The results of the proposed SID method (b) with the PALI loss and (c) without the PALI loss, respectively. The images from top row to bottom are the estimated \hat{J} , \hat{t} , and \hat{A} .

existing SID methods: the model-specific method DCP [1], the supervised domain-driven methods AOD-Net [4] and MSBDN [6], and the unsupervised domain-driven method DCPLoss [17], and the image-driven methods DDIP [9], ZID [10], and ZR [11]. We compare our method with them on the conventional unpaired real hazy data. For AOD-Net [4], MSBDN [6], and DCPLoss [17], we downloaded pre-trained models provided by authors. For the image-driven methods [9]–[11], we trained the models from the scratch for each image with the shared code and the default settings.

a) Qualitative Comparison: We first provide the qualitative comparison results tested on the unpaired real hazy images in Fig 8. DCP [1] successfully removes the haze well, but as the DCP adopts the approximation and the regularization such as the soft matting, the haze remains at the boundaries of the objects, as shown in the last row in Fig 8(b). The supervised-learning based domain-driven methods, the AOD-Net [4] and the MSBDN [6], rarely remove the haze artifacts due to the domain-shift problems as shown in Fig 8(c-d). Note that the test images are not in the training datasets [12]–[14]. Similarly, the unsupervised-learning based domain-driven method, DCPLoss [17], also shows the limited performance due to the domain-shift problems as shown in Fig 8(e). We expect the image-driven methods shows the better performance than the domain-driven methods as it is free from the domain-shift problems. However, DDIP [9] tends to lose the color information, ZID [10] draws the artifacts due to the over-enhancement significantly ruining the naturalness of the restored images, and ZR [11] shows the under-enhanced results as shown in Figure 8(f-h). On the other hands, the proposed method faithfully restores the scene radiance without noisy artifacts as shown in Figure 8(i).

We also compare the methods on the paired real hazy



(a) Input (b) DCP [1] (c) AOD-Net [4] (d) MSBDN [6] (e) DCPLoss [17] (f) DDIP [9] (g) ZID [10] (h) ZR [11] (i) Ours

Fig. 8: Qualitative comparison of the SID results tested on the unpaired hazy images.

Method	I-HAZE		O-HAZE	
	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow
DCP [1]	12.1475	0.6272	16.5172	0.6775
AOD-Net [4]	15.3829	0.7121	15.0626	0.5782
MSBDN [6]	16.3930	0.7093	17.5609	0.4408
DCPLoss [17]	15.2877	0.7141	16.8780	0.5983
DDIP [9]	14.9828	0.6959	16.6616	0.6323
ZID [10]	11.8692	0.5943	13.2808	0.4994
ZR [11]	16.3311	0.6972	16.4143	0.4398
Ours	15.7308	0.7048	17.8736	0.6754

TABLE II: Quantitative comparison of the SID results tested on the I-HAZE and O-HAZE images. Numbers in red, blue, and bold denote the best, second-best and third-best scores.

images in Figure 9. In this case, the DCP shows artifacts as the under estimated transmission, which is the main drawbacks of the model-specific method. On the other hands, the domain-driven methods show the satisfactory performance as shown in Figure 9(c-e). However, please note that it is difficult to prepare the abundant images for the unseen future test images, and it is worth to note that the image-driven methods have comparable performance with the domain-driven methods except for the ZID.

b) Quantitative Comparison: For the quantitative comparison, we evaluate the PSNR and SSIM scores for the test

images in I-HAZE and O-HAZE and the results are shown in Table. II. The proposed method has the third-best PSNR score in I-HAZE [13] and showed the best PSNR and the second-best SSIM scores in O-HAZE [14]. It is worth to note that the proposed method have the best performance on the O-HAZE, which is even better than the domain-driven method trained by the abundant dataset, since the proposed method is softly constrained by (1) than mode-driven methods. Yet, the proposed method still have limitations on the more extreme cases, such as the non-homogeneous haze. Nevertheless, the experimental results suggest that the image-driven SID method can outperform the domain-driven method and thus there are rooms to research in a zero-shot and self-supervised learning scheme in the future.

V. CONCLUSION

In this paper, we proposed a zero-shot SID method with the self-supervised learning scheme. We employed two generators to estimate the transmission and the original scene radiance from an input hazy image, respectively. We also synthesize the PALI to train the deep neural networks in the self-supervised learning strategy. Using the PALI, the network learns the characteristics of dense haze and the performance is improved accordingly. The experimental results showed that



Fig. 9: Qualitative comparison of the SID results tested on the I-HAZE and O-HAZE images.

the proposed method is effective and has the comparable performance with the state-of-the arts SID methods when tested on the I-HAZE dataset and even outperforms them when tested on the O-HAZE dataset.

REFERENCES

- [1] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE TPAMI*, vol. 33, no. 12, pp. 2341–2353, 2011.
- [2] R. T. Tan, "Visibility in bad weather from a single image," in *CVPR*, 2008, pp. 1–8.
- [3] R. Fattal, "Single image dehazing," *ACM TOG*, vol. 27, no. 3, p. 1–9, aug 2008.
- [4] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *ICCV*, 2017, pp. 4780–4788.
- [5] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *ECCV*, 2016, pp. 154–169.
- [6] 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 *CVPR*, 2020, pp. 2154–2164.
- [7] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing," in *AAAI*, 2020, pp. 11 908–11 915.
- [8] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," in *ICCV*, 2019, pp. 7313–7322.
- [9] Y. Gandelsman, A. Shocher, and M. Irani, "'Double-dip': Unsupervised image decomposition via coupled deep-image-priors," in *CVPR*, 2019.
- [10] B. Li, Y. Gou, J. Z. Liu, H. Zhu, J. T. Zhou, and X. Peng, "Zero-shot image dehazing," *IEEE TIP*, vol. 29, pp. 8457–8466, 2020.
- [11] A. Kar, S. K. Dhara, D. Sen, and P. K. Biswas, "Zero-shot single image restoration through controlled perturbation of koschmieder's model," in *CVPR*, 2021, pp. 16 205–16 215.
- [12] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE TIP*, vol. 28, no. 1, pp. 492–505, 2019.
- [13] 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.
- [14] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer, "O-haze: A dehazing benchmark with real hazy and haze-free outdoor images," in *CVPRW*, 2018, pp. 867–875.
- [15] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images," in *ICIP*, 2019, pp. 1014–1018.
- [16] C. O. Ancuti, C. Ancuti, and R. Timofte, "Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *CVPRW*, 2020, pp. 1798–1805.
- [17] A. Golts, D. Freedman, and M. Elad, "Unsupervised single image dehazing using dark channel prior loss," *IEEE TIP*, vol. 29, pp. 2692–2701, 2020.
- [18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2015.