Self-Supervised Dehazing Network Using Physical Priors

Gwangjin Ju[1], Yeongcheol Choi[2], Donggun Lee[1], Jee Hyun Paik[2], Gyeongha Hwang[3], and Seungyong Lee[1⋆]

[1] POSTECH, Pohang, Korea
{gwangjin, dalelee, leesy}@postech.ac.kr
[2] POSCO ICT, Pohang, Korea
{ycchoi, jeehyun100}@poscoict.com
[3] Yeungnam University, Gyeongsan, Korea
ghhwang@yu.ac.kr

Abstract. In this paper, we propose a lightweight self-supervised dehazing network with the help of physical priors, called *Self-Supervised Dehazing Network (SSDN)*. SSDN is a modified U-Net that estimates a clear image, transmission map, and atmospheric airlight out of the input hazy image based on the Atmospheric Scattering Model (ASM). It is trained in a self-supervised manner, utilizing recent self-supervised training methods and physical prior knowledge for obtaining realistic outputs. Thanks to the training objectives based on ASM, SSDN learns physically meaningful features. As a result, SSDN learns to estimate clear images that satisfy physical priors, instead of simply following data distribution, and it becomes generalized well over the data domain. With the self-supervision of SSDN, the dehazing performance can be easily finetuned with an additional dataset that can be built by simply collecting hazy images. Experimental results show that our proposed SSDN is lightweight and shows competitive dehazing performance with strong generalization capability over various data domains.

1 Introduction

Vision systems are applied to many tasks like autonomous driving, factory safety surveillance, etc. However, haze artifacts reduce the scene visibility, and the performance of vision systems could be degraded by reduced visibility. Dehazing can help vision systems become robust by reducing haze from the scene.

One of the popular approaches in the dehazing task is prior-based. Most of the existing prior-based dehazing methods are based on Atmospheric Scattering Model (ASM) [17]. Prior-based methods remove haze by estimating the transmission map, which reflects the amount of haze in the image. Since the methods

⋆ ORCID IDs: {0000-0003-1006-8351}, {0000-0003-2525-873X}, {0000-0001-6482-8892}, {0000-0001-5485-2602}, {0000-0001-9791-8130}, {0000-0002-8159-4271}

(a) (b) (c)

Fig. 1. Dehazed results of (a) input images by (b) DehazeNet [4] and (c) our method. While our method is trained on the indoor dataset in a self-supervised manner, it shows comparative performance to DehazeNet which is trained on the outdoor dataset in a supervised manner.

are based on physical properties, they show stable performance over a variety of domains with different image contents.

Recently, deep learning based dazing methods are proposed, which utilize powerful CNNs. Early method [4] estimates the transmission map of ASM using CNN. Most deep learning-based dehazing methods are image-to-image translation approaches like pix2pix [11] and estimate the clear image directly from a hazy image. More recently, an unsupervised image-to-image translation approach [5] has been proposed using cycle consistency from CycleGAN [27]. Deep learning-based methods show high performance in the trained domains and run fast by GPU acceleration.

Although many approaches are proposed for dehazing, they have limitations. Most prior-based methods are not designed to utilize GPU, so they run slowly compared to deep learning-based methods. They also show lower performance compared to data-driven methods. Deep learning-based methods show high performance but their performance degrades a lot when the data domain changes. Moreover, building datasets is difficult because acquiring both hazy and clear images with the same view is almost infeasible.

In this paper, inspired by [18] and [17], we propose *Self-Supervised Dehazing Network (SSDN)*. SSDN estimates disentangled clear image, transmission map, and atmospheric airlight out of a hazy image. In the training phase, SSDN learns to satisfy the ASM by reconstructing the hazy image from outputs. To make the output of SSDN realistic, we exploit physical prior knowledge such as Dark Channel Prior [10], total variation [21], the relation between hazy and clear images in color space, and local variance.

To the best of our knowledge, this research is the first attempt to merge physical prior knowledge and self-supervision on the dehazing task. SSDN has several advantages that compensate limitations of other approaches. First, SSDN dehazes images based on ASM, which is physically meaningful and explainable than other deep learning-based methods. Second, SSDN is implemented as a lightweight CNN, enabling application to real-time systems. Third, SSDN is well generalized over data domains thanks to physical prior knowledge. Lastly, it is easy to build a dataset for SSDN training by simply collecting hazy images without corresponding clear images. With these advantages, SSDN can be applied to practical vision systems, especially real-time ones due to its lightweight structure and stable performance over various data domains.

In summary, our method merges physical prior knowledge with self-supervised learning to resolve limitations of previous dehazing methods, such as long execution time, different performance over domains, and laborious dataset building.

2 Related Work

2.1 Prior-Based Methods

Prior-based methods perform dehazing based on the observed characteristics of hazy images. Most of those method utilizes Atmospheric Scattering Model [17] that models hazy image based on physics-based knowledge.

One of the most representative methods is the Dark Channel Prior (DCP) [10]. DCP assumes that at least one of the R, G, and B channel values tends to be very low for clear images. The Color Attenuation Prior (CAP) [28] assumes that the difference between saturation and value channel in HSV space is proportional to the amount of haze. In [3], the Non-local Color Prior (NCP) is proposed, which is based on the observation that a clear image consists of a small number of clusters in the RGB color space.

Most prior-based methods are straightforward and more accessible than deep learning-based approaches. Furthermore, those methods show strong domain generalization capability. However, the dehazing performance of prior-based methods is lower than most of the supervised deep learning-based methods.

2.2 Supervised Learning Methods

Recently, the dehazing task was successfully performed using supervised deep learning methods. These methods exploit a paired dataset of hazy and clear images to learn to generate clear images from given hazy images.

DehazeNet [4] takes a hazy image as input and estimates the transmission map. The estimated transmission map is then used with ASM to dehaze the hazy image. Feature Fusion Attention Network (FFANet) [19] improved the dehazing performance by proposing specialized modules for dehazing, such as channel attention and pixel attention modules. In the supervised learning methods, the dehazing performance has been improved thanks to the recent advances of deep learning architectures [13, 16, 25].

Supervised deep learning methods have demonstrated high dehazing performance and showed fast inference speed based on GPU acceleration. However, recent deep learning-based methods are getting to use larger and more complex networks, and performing them on low-end devices, e.g., embedded systems for autonomous driving, would not be easy. In addition, it is challenging to build a paired dataset needed for supervised learning methods. Lastly, these methods may show poor generalization capability due to the dependency on the dataset.

2.3 Unsupervised Learning Methods

Recently unsupervised learning methods have been proposed to perform dehazing without a paired dataset to overcome the difficulty of dataset construction. Cycle-Dehaze [5] uses cycle consistency loss [27] to train a dehazing network on an unpaired dataset. D^4 (Dehazing via Decomposing transmission map into Density and Depth) [24] utilizes depth estimation with ASM to build cycle consistency. These methods need no paired dataset but still, it needs both clean and hazy image sets.

DDIP [7] and You Only Look Yourself (YOLY) [15] perform dehazing based on ASM. They optimize a network from scratch for each hazy image, without using any dataset. As these methods perform optimization for a single image, they are free from domain change and show better performance than other prior-based methods. However, they take seconds or even minutes to remove haze from an image, which makes them hard to be used for real-time vision systems.

3 Physical Priors

In this section, we describe physical priors used for our proposed framework. We first describe priors proposed in previous works and then propose several priors for the dehazing task based on ASM.

3.1 Conventional Priors

Atmospheric Scattering Model ASM [17] represents the scattering of scene radiance by the particle in the atmosphere as the following equation:

$$I = JT + A(1 - T), \tag{1}$$

where I, J, T, and A denote hazy image, clear scene radiance, transmission map, and atmospheric airlight, respectively. For a hazy image $I \in \mathbb{R}^{H \times W \times C}$, we assume the transmission is a multi-channel map $T \in \mathbb{R}^{H \times W \times C}$ and the airlight A is homogeneous, i.e. $A \in \mathbb{R}^{1 \times 1 \times C}$.

Dark Channel Prior DCP [10] estimates the transmission map based on the observation that the pixel intensity of at least one channel among RGB channel is close to 0 in a clear image. It can be represented as follow:

$$T_{DCP} = 1 - \min_{c}(\min_{y \in \Omega(x)}(\frac{I^c(y)}{A^c})), \tag{2}$$

where c denotes a RGB color channel and $\Omega(x)$ is the window of each pixel.

Total Variation Prior Total variation prior [21] has been used for various image restoration tasks. It is based on the observation that a clear image tends to have a quite low total variation (nearly 0) which can be represented as follow:

$$TV(I) = \sum_{i,j} |I_{i+1,j} - I_{i,j}| + |I_{i,j+1} - I_{i,j}| \approx 0, \tag{3}$$

where $I_{i,j}$ denotes the pixel value of I at coordinate (i, j).

3.2 Our Proposed Priors

In ASM, a hazy image is represented as an interpolation between a clear image and airlight, and it is known that airlight is close to white. We propose several priors on hazy and clear images with this property.

In RGB color space, the pixel intensity of the hazy image I is higher than that of the clear image J as I is closer to white than J due to the interpolation in ASM:

$$J < I. \tag{4}$$

In HSV color space, we derive priors on value (brightness) and saturation channels. For value (brightness) channel, the pixel value of a hazy image is higher than that of the clear image, again due to the interpolation in ASM. For saturation channel, the saturation of hazy image is lower than that of a clear image, as mixture with white reduces the saturation.

$$J_{value} < I_{value}, \tag{5}$$

$$J_{sat} > I_{sat}, \tag{6}$$

where x_{value} is value channel and x_{sat} is saturation channel in HSV space.

The homogeneous airlight A makes the contrast of a hazy image lower than a clear image. Hence, the local variance of the hazy image is lower than that of the clear image as follow:

$$var(J) > var(I), \tag{7}$$

where $var(x)$ is a local variance operator that computes the variance of each color channel in a local window centered at x.

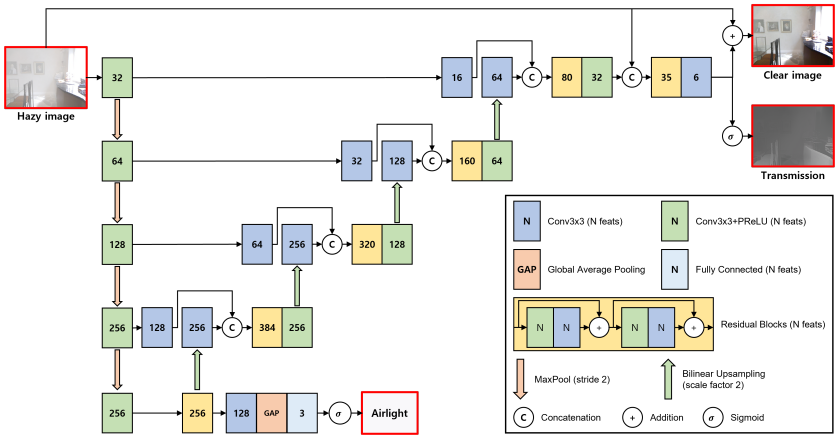


Fig. 2. Overall framework of our proposed SSDN. Given a hazy image, SSDN estimates clear image, transmission map, and airlight separately.

4 Self-Supervised Dehazing Network

Our self-supervised learning framework for dehazing has been inspired by CVF-SID [18] but contains important differences to exploit ASM and physical priors of hazy images. Unlike the denoising problem that CVF-SID handles by simply exploiting reconstruction loss and variance norm, the dehazing task needs to consider a complicated haze structure. So we exploit several physical priors described in Sec. 3 to train our Self-Supervised Dehazing Network (SSDN) based on ASM.

4.1 Overall Framework

Our proposed SSDN is a multi-variate function, and the structure of SSDN is a modified U-Net [20] as shown in Fig. 2. The output of SSDN is given by:

$$\hat{J}, \hat{T}, \hat{A} = SSDN(I), \tag{8}$$

where

- $\hat{J} \in \mathbb{R}^{H \times W \times C}$ is the estimated clear image.
- $\hat{T} \in \mathbb{R}^{H \times W \times C}$ is the estimated transmission map.
- $\hat{A}$ is the estimated airlight. It assumed to be homogeneous, i.e. $\hat{A} \in \mathbb{R}^{1 \times 1 \times C}$. It is estimated by the lowest level feature map followed by global average pooling, a fully-connected layer, and a Sigmoid layer.

4.2 Self-Supervised Dehazing Losses

In this section, we describe the training objectives used in the proposed framework. The framework performs dehazing by exploiting the reconstruction

loss using ASM, the physical priors based loss, and the regularization loss. In our framework, GAN loss is not available since it is trained only with hazy images. Excluding GAN loss makes the quantitative performance slightly lower but helps the domain generalization which is important in practical applications. $Sl1(x)$ in the following sections is smooth L1 loss introduced in [8] and defined as below.

$$Sl1(x) = \begin{cases} 0.5(x)^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{9}$$

ASM Losses We impose the constraint so that the outputs $\hat{J}$, $\hat{T}$, and $\hat{A}$ of SSDN satisfy ASM as follow:

$$L_{rec} = Sl1(I - (\hat{J}\hat{T} + \hat{A}(1 - \hat{T}))). \tag{10}$$

If we only use Eqn. (10), the dehazing network will give a trivial solution, i.e. $\hat{J} = I$. To avoid the trivial solution, we propose auxiliary reconstruction loss exploiting Eqn. (2):

$$L_{DCPrec} = Sl1(I - (\hat{J}T_{DCP} + A_{max}(1 - T_{DCP}))), \tag{11}$$

where A_{max} denotes the maximum pixel value of I of each channel and T_{DCP} is described in Eqn. (2). When the input of SSDN is $\hat{J}$, the clear image output should be the same as the input and the transmission map output should be 1 as follow:

$$\begin{aligned} \hat{J}_J, \hat{T}_J, \hat{A}_J &= SSDN(\hat{J}), \\ L_{Jrec} &= Sl1(\hat{J}_J - \hat{J}) + Sl1(\hat{T}_J - 1). \end{aligned} \tag{12}$$

Since $\hat{A}$ is the homogeneous airlight, if we use it as the input of SSDN, the airlight output is the same to the input and the transmission output should be 0.

$$\begin{aligned} \hat{J}_A, \hat{T}_A, \hat{A}_A &= SSDN(\hat{A}), \\ L_{Arec} &= Sl1(\hat{A}_A - \hat{A}) + Sl1(\hat{T}_A). \end{aligned} \tag{13}$$

Additionally, we propose a self-augmentation loss inspired by [18] as follow:

$$\begin{aligned} T' &= \hat{T} + N(0, \sigma_T^2), \\ A' &= \hat{A} + N(0, \sigma_A^2), \\ I' &= \hat{J}T' + A'(1 - T'), \\ \hat{J}_{aug}, \hat{T}_{aug}, \hat{A}_{aug} &= SSDN(I'), \\ L_{aug} &= Sl1(I' - (\hat{J}_{aug}\hat{T}_{aug} + \hat{A}_{aug}(1 - \hat{T}_{aug}))), \end{aligned} \tag{14}$$

where $N(\mu, \sigma^2) \in \mathbb{R}^{1\times1\times C}$ is a Gaussian noise for homogeneous changes of T and A, while σ_T^2 and σ_A^2 are hyperparameters. Applying homogeneous changes helps the self-augmented data $\hat{J}_{aug}$ to be physically plausible augmentations.

Overall, the loss exploiting ASM is represented as follow:

$$L_{ASM} = \lambda_{rec}L_{rec} + \lambda_{DCPrec}L_{DCPrec} + \lambda_{Jrec}L_{Jrec} + \lambda_{Arec}L_{Arec} + \lambda_{aug}L_{aug}, \tag{15}$$

where each λ is a weight hyperparameter for each loss term.

Prior Losses In this section, we define the loss functions reflecting the priors described in Sec. 3 so that the clear image output $\hat{J}$ of SSDN is close to the real clear image.

First, we use Eqn. (2) as a guidance for $\hat{T}$. To reduce halo artifact, Ω in Eqn. (2) is set to 1×1. Since DCP may make wrong estimation of the transmission map in a bright scene, such as a white wall in an indoor or city scene, we multiply T_{DCP} on loss term to impose low weight on those cases.

$$L_{DCP} = Sl1((\hat{T} - T_{DCP})T_{DCP}), \tag{16}$$

where T_{DCP} is defined in Eqn. (2). Secondly, we define the total variation prior described in Eqn. (3) to the clear image output $\hat{J}$.

$$L_{TV} = Sl1(TV(\hat{J})). \tag{17}$$

We implement the priors described in Eqns. (4), (5), and (6) as the following loss terms:

$$L_{PI} = Sl1(\max(\hat{J} - I, 0)), \tag{18}$$

$$L_{value} = Sl1(\max(\hat{J}_{value} - I_{value}, 0)), \tag{19}$$

$$L_{sat} = Sl1(\max(I_{sat} - \hat{J}_{sat}, 0)). \tag{20}$$

The local variance prior described in Eqn. (7) is implemented as a loss term as follow:

$$L_{var} = Sl1(\max(var(I) - var(\hat{J}), 0)). \tag{21}$$

We avoid overdehazing and make the output close to the clear image by applying $\max(*, 0)$ in Eqns. (18), (19), (20), and (21). Lastly, we define the loss for the airlight from the observation that the airlight is similar to the highest pixel value of the hazy image:

$$L_A = Sl1(\hat{A} - max(I)). \tag{22}$$

Overall, the loss function reflecting the priors is represented as follow:

$$\begin{aligned} L_{prior} =& \lambda_{DCP} L_{DCP} + \lambda_{TV} L_{TV} + \lambda_{PI} L_{PI} + \\ & \lambda_{value} L_{value} + \lambda_{sat} L_{sat} + \lambda_{var} L_{var} + \lambda_A L_A. \end{aligned} \tag{23}$$

where each λ is a weight hyperparameter for each loss term.

Regularization We apply the regularization so that the output of the network has appropriate values. First, we define the identity loss to prevent that the clear image output $\hat{J}$ is quite different from I as follow:

$$L_{idt} = Sl1(\hat{J} - I). \tag{24}$$

Secondly, we impose regularization on the estimated transmission map. In general, the transmission map is known to be smooth. Moreover, since the scattering

coefficient is similar in the range of the visual light, the transmission should be close to gray. The regularization for the transmission is given by:

$$L_{Tgray} = Sl1(\hat{T} - mean_c(\hat{T})), \tag{25}$$

$$L_{Tsmooth} = Sl1(\hat{T} - BoxFilter(\hat{T})), \tag{26}$$

where $mean_c(x)$ is a channelwise mean operator for pixel x.

Merging each regularization, the final regularization loss function is defined as:

$$L_{reg} = \lambda_{idt} L_{idt} + \lambda_{Tgray} L_{Tgray} + \lambda_{Tsmooth} L_{Tsmooth}, \tag{27}$$

where each λ is a weight hyperparameter for each loss term.

Total Loss The final objective function is given by:

$$L = L_{ASM} + L_{prior} + L_{reg}. \tag{28}$$

5 Experiments

5.1 Implementation Details

We implement SSDN based on U-Net but replaced normal convolutions and ReLUs with residual blocks and PReLUs, respectively. For a lightweight framework, we reduce the number of channels and apply channel reductions to skip connections and decoder outputs as shown in Fig. 2.

In the training phase, we randomly crop each image from the dataset to a patch with size of 128x128. To obtain the effect of extending the dataset, we apply random horizontal flip as data augmentation. For the update of network parameters, Adam optimizer [12] with a learning rate of 1e-4 is used with batch size of 32 on two TITAN XPs.

There are hyperparameters in Eqn. (28). We empirically set them: $\{\lambda_{rec}, \lambda_{DCPrec}, \lambda_{Tgray}, \lambda_{Tsmooth}, \lambda_{var}\}$ to 1, $\{\lambda_{Jrec}, \lambda_{Arec}, \lambda_{aug}, \lambda_{DCP}, \lambda_{idt}\}$ to 0.1, $\{\lambda_{TV}, \lambda_{PI}, \lambda_{value}, \lambda_{sat}, \lambda_A\}$ to 0.01. In Eqn. (14), we define the values of hyperparameters as $\sigma_T^2 = 0.3$ and $\sigma_A^2 = 0.2$.

5.2 Datasets

We conduct experiments on both indoor and outdoor scenes. For the indoor case, we train SSDN with RESIDE standard [14] indoor training set. Indoor set of the Synthetic Objective Testing Set (SOTS) from RESIDE standard is selected as the indoor test set. We refer to SSDN trained on RESIDE standard as *ours-indoor*.

For the outdoor case, we train SSDN with RESIDE beta [14] Outdoor Training Set (OTS). The synthetic set of Hybrid Subjective Testing Set (HSTS) from RESIDE standard is selected as the outdoor test set. We refer to SSDN trained on OTS as *ours-outdoor*.

Table 1. Comparison in terms of computational cost on 620x460 images.

Metrics	Supervised Method			
	DehazeNet[4]	MSBDN-DFF[9]	DW-GAN[6]	TBDN[26]
FLOPs (GMACs)	-	183.62	150.61	396.96
Params (M)	-	31.35	51.51	50.35
Exec time (s)	1.05	0.03	0.05	0.06
Metrics	Prior-Based	Unsupervised		Self-Supervised
	DCP	DDIP	YOLY	Ours
FLOPs (GMACs)	-	-	-	**94.53**
Params (M)	-	-	-	**16.62**
Exec time (s)	0.05	$\sim$ 600	$\sim$ 30	**0.008**

(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 3. Dehazing results of ours and other methods. The first two rows from SOTS indoor, the other two from the HSTS. (a) Input hazy images, (b) DehazeNet, (c) DCP, (d) DDIP, (e) YOLY, (f) *ours-indoor*, (g) *ours-outdoor*, and (h) GT clean images.

5.3 Results

Baselines To compare the performance of our method, we select diverse approaches to dehazing. For the supervised method, we select DehazeNet [4], MSBDN-DFF [9], DW-GAN [6], and TBDN [26]. The performance of [6] and [26] on HSTS are not reported on their paper. We train them on OTS and test on HSTS as *ours-outdoor*. Hence, the results of [6] and [26] on Table 3 may not be the best for them. For the prior-based method, we select DCP [10], which is the representative method in that area. Lastly, for unsupervised methods, we select DDIP [7] and YOLY [15], which optimizes a neural network on each image.

Experiment Results Comparisons on computational cost are shown in Table 1. For an image with the size of 620×460, YOLY and DDIP take about few minutes and supervised methods take fractions of a second. On the other hand, our method takes about 0.008 seconds on a TITAN XP since it requires

Table 2. Quantitative results of other methods and ours on the synthetic indoor dehazing test set (SOTS).

Metrics	Supervised Method				
	DehazeNet[4]	MSBDN-DFF[9]		DW-GAN[6]	TBDN[26]
PSNR	21.14	33.79		35.94	**37.61**
SSIM	0.847	0.984		0.986	**0.991**
Metrics	Prior-Based	Unsupervised		Self-Supervised	
	DCP	DDIP	YOLY	*ours-indoor*	*ours-outdoor*
PSNR	16.62	16.97	19.41	19.56	19.51
SSIM	0.818	0.714	0.833	0.833	0.827

Table 3. Quantitative results of other methods and ours on the synthetic outdoor dehazing test set (HSTS).

Metrics	Supervised Method				
	DehazeNet[4]	MSBDN-DFF[9]		DW-GAN[6]	TBDN[26]
PSNR	24.48	**31.71**		30.67	27.22
SSIM	0.915	0.933		**0.973**	0.916
Metrics	Prior-Based	Unsupervised		Self-Supervised	
	DCP	DDIP	YOLY	*ours-indoor*	*ours-outdoor*
PSNR	14.84	20.91	23.82	19.47	19.84
SSIM	0.761	0.884	0.913	0.859	0.851

(a) (b) (c) (d) (e) (f)

Fig. 4. More dehazing results on SOTS dataset by our proposed SSDN. (a), (d) input hazy images, (b), (e) output dehazed images, and (c), (f) GT clear images. For upper two rows, the images are sampled from SOTS indoor set and dehazed using *ours-indoor*, SOTS outdoor set and *ours-outdoor* for lower two rows.

Table 4. Result of ablation study. For each experiment setting, PSNR and SSIM is measured between the output and SOTS indoor dataset. The first row, *hazy-clear* shows the PSNR and SSIM between hazy and clear images in the dataset.

Experiment Name	Experiment Setting	PNSR	SSIM
-	hazy-clear	11.97	0.6934
Model 1	L_{rec} (Eqn. (10))	11.59	0.6896
Model 2	L_{ASM} (Eqn. (15))	9.08	0.2821
Model 3	L_{ASM}, L_{prior} (Eqns. (15), (23))	8.60	0.3012
Model 4	L (Eqn. (28))	19.35	0.8304

(a) (b) (c) (d) (e) (f)

Fig. 5. Results of ablation study. (a) input hazy image, (b) *Model 1*, (c) *Model 2*, (d) *Model 3*, (e) *Model 4* and (f) the GT clear image. The top row shows the RGB image, and the bottom shows the estimated transmission map.

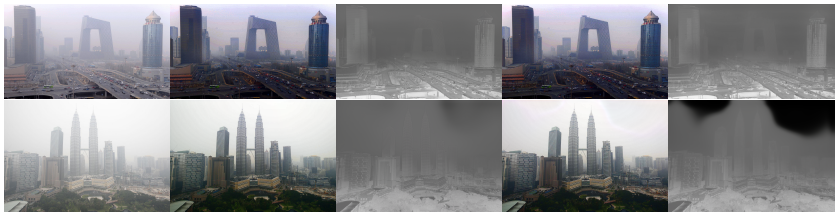
only a single CNN forward operation. It shows that SSDN can be combined into real-time vision systems without losing real-time property while others cannot.

The quantitative comparisons of each baseline and our methods are shown in Tables 2, 3 and Fig. 3. Our methods show worse performance than supervised methods and similar performance compared to DCP, DDIP, and YOLY. However, our methods show a good generalization property since SSDN learns physical prior knowledge that can be commonly applied to general hazy images.

More dehazed results on SOTS indoor and outdoor dataset by our proposed SSDN are shown in Fig. 4.

Ablation Study In this section, we show the effect of loss terms (Eqns. (10), (15), and (23)) by ablation study. The results are shown in Table 4 and Fig. 5. *Model 1* outputs an image same to the input, a trivial solution. *Model 2* avoids a trivial solution but it does not properly learn dehazing, especially for bright scene such as white wall (Fig. 5 (c)). *Model 3* handles bright scene and dehazes better than *Model 2* by applying low weights to white regions and using other prior losses such as L_{prior} (Eqn. (23)) (Fig. 5 (d)). L_{reg} (Eqn. (27)) makes *Model 4* avoid over-dehazing of *Model 3* to produce a clear dehazed image.

Real World Examples As shown in Fig. 6, the proposed method works on real-world hazy images although it has been trained only on a synthetic dataset. In Fig. 6 (b), *ours-indoor* shows the strong generalization capability.



(a) (b) (c) (d) (e)

Fig. 6. Dehazing results on real hazy images using SSDN. Both cases show competitive performance. (a) input hazy images, (b), (c) output RGB images and transmission maps of *ours-indoor*, and (d), (e) output RGB images and transmission maps of *ours-outdoor*.

(a) (b) (c) (d) (e)

Fig. 7. Dehazing results on real extremely hazy images using SSDN. *ours-outdoor* does not work properly while finetuned *ours-outdoor* shows better performance. (a) input hazy images, (b), (c) output RGB images and transmission maps of *ours-outdoor*, and (d), (e) output RGB images and transmission maps of finetuned *ours-outdoor*.

Despite the generalization capability of SSDN, extremely hazy images is too different from trained images. In this case, the performance of SSDN can be easily improved by finetuning process with additional hazy images. Fig. 7 (d) shows the results of finetuned *ours-outdoor* with 110 real-world extremely hazy images.

Vision System Application To show that SSDN can make the vision system robust to haze, we choose the depth estimation task as an exemplar case. We use LapDepth [23], a depth estimation framework trained on Make3D dataset [22]. To simulate hazy weather, we synthesize hazy images out of Make3D test set. To synthesize hazy images, we extract the transmission maps and airlights based on DCP from O-HAZE [2] and NH-HAZE [1]. Then, we apply ASM on RGB images of Make3D. In this strategy, we successfully transfer realistic haze from O-HAZE and NH-HAZE to Make3D dataset with this process as shown in Fig. 8.

In Table 5, LapDepth shows large performance degradation on hazy images. SSDN can compensate for the performance degradation while it had little effect on execution time, taking about 0.005 seconds even on images with the size of 1704×2272 with a TITAN XP. For comparison, we finetune LapDepth on our synthesized hazy images. It shows better performance on hazy images, while it

(a) (b) (c) (d) (e) (f)

Fig. 8. Haze synthesis results on Make3D dataset and its dehazed result using SSDN. (a), (d) original clear images, (b), (e) synthesized hazy images, and (c), (f) dehazed results by *ours-outdoor*.

Table 5. Depth estimation results of LapDepth on each experiment setting.

Experiment Setting	Dataset	SSDN	RMSE	SSIM
LapDepth	Clear images		8.76	0.94
LapDepth	Hazy images		11.32	0.92
LapDepth	Hazy images	✓	9.85	0.92
Finetuned LapDepth	Clear images		9.16	0.93
Finetuned LapDepth	Hazy images		9.09	0.94

makes a degraded performance on clear images. On the other hand, SSDN does not affect the performance for the clear weather case. It shows that our proposed SSDN is a practical method for assisting existing vision systems.

6 Conclusion

SSDN disentangles a hazy image into a clear image, transmission map, and atmospheric airlight based on ASM. To the best of our knowledge, it is the first approach to merging physical prior knowledge and self-supervision for the dehazing task. The proposed method shows competitive dehazing performance to other prior-based methods or unsupervised methods while running extremely faster than them. Our proposed SSDN shows strong generalization capability and it can be more stable over various domains by finetuning with simply gathered additional hazy images. We also showed that SSDN can make existing vision systems robust to hazy images. Experimental results show that our proposed SSDN is a practical dehazing method for real-time vision systems.

Acknowledgements We would like to thank the anonymous reviewers for their constructive comments. This work was supported by IITP grants (SW Star Lab, 2015-0-00174; AI Innovation Hub, 2021-0-02068; AI Graduate School Program (POSTECH), 2019-0-01906), KOCCA grant (R2021040136), NRF grant (NRF-2021R1F1A1048120) from Korea government (MSIT and MCST), and POSCO ICT COMPANY LTD.

References

1. Ancuti, C.O., Ancuti, C., Timofte, R.: NH-HAZE: an image dehazing benchmark with non-homogeneous hazy and haze-free images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2020)
2. Ancuti, C.O., Ancuti, C., Timofte, R., De Vleeschouwer, C.: O-haze: A dehazing benchmark with real hazy and haze-free outdoor images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2018)
3. Berman, D., treibitz, T., Avidan, S.: Non-local image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
4. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: An end-to-end system for single image haze removal. IEEE Transactions on Image Processing **25**(11), 5187–5198 (2016)
5. Engin, D., Genc, A., Kemal Ekenel, H.: Cycle-dehaze: Enhanced cyclegan for single image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2018)
6. Fu, M., Liu, H., Yu, Y., Chen, J., Wang, K.: Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2021)
7. Gandelsman, Y., Shocher, A., Irani, M.: "double-dip": Unsupervised image decomposition via coupled deep-image-priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
8. Girshick, R.: Fast r-cnn. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) (2015)
9. Hang, D., Jinshan, P., Zhe, H., Xiang, L., Xinyi, Z., Fei, W., Ming-Hsuan, Y.: Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
10. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1956–1963 (2009). https://doi.org/10.1109/CVPR.2009.5206515
11. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations (ICLR) (2015)
13. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 4780–4788 (2017). https://doi.org/10.1109/ICCV.2017.511
14. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing **28**(1), 492–505 (2019)
15. Li, B., Gou, Y., Gu, S., Liu, J.Z., Zhou, J.T., Peng, X.: You Only Look Yourself: Unsupervised and Untrained Single Image Dehazing Neural Network. International Journal of Computer Vision **129**(5), 1754–1767 (2021)
16. Liu, J., Wu, H., Xie, Y., Qu, Y., Ma, L.: Trident dehazing network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 430–431 (2020)

17. Narasimhan, S.G., Nayar, S.K.: Vision and the atmosphere. International Journal of Computer Vision **48**(3), 233–254 (2002)
18. Neshatavar, R., Yavartanoo, M., Son, S., Lee, K.M.: Cvf-sid: Cyclic multi-variate function for self-supervised image denoising by disentangling noise from image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17583–17591 (2022)
19. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: Ffa-net: Feature fusion attention network for single image dehazing. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 11908–11915 (2020)
20. Ronneberger, O., P.Fischer, Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS, vol. 9351, pp. 234–241. Springer (2015)
21. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D Nonlinear Phenomena **60**(1-4), 259–268 (1992). https://doi.org/10.1016/0167-2789(92)90242-F
22. Saxena, A., Sun, M., Ng, A.Y.: Make3d: Learning 3d scene structure from a single still image. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(5), 824–840 (2009). https://doi.org/10.1109/TPAMI.2008.132
23. Song, M., Lim, S., Kim, W.: Monocular depth estimation using laplacian pyramid-based depth residuals. IEEE transactions on circuits and systems for video technology **31**(11), 4381–4393 (2021)
24. Yang, Y., Wang, C., Liu, R., Zhang, L., Guo, X., Tao, D.: Self-augmented unpaired image dehazing via density and depth decomposition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2037–2046 (2022)
25. Yu, Y., Liu, H., Fu, M., Chen, J., Wang, X., Wang, K.: A two-branch neural network for non-homogeneous dehazing via ensemble learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 193–202 (2021)
26. Yu, Y., Liu, H., Fu, M., Chen, J., Wang, X., Wang, K.: A two-branch neural network for non-homogeneous dehazing via ensemble learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2021)
27. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2017)
28. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Transactions on Image Processing **24**(11), 3522–3533 (2015). https://doi.org/10.1109/TIP.2015.2446191