Towards Unsupervised Single Image Dehazing with Deep Learning

Lu-Yao Huang[1,2], Jia-Li Yin[2], Bo-Hao Chen[2], and Shao-Zhen Ye[1]

[1] Fuzhou University; [2] Yuan Ze University

Contribution: Unsupervised Learning for Haze Removal

Deep learning computation is often used in single-image dehazing techniques for outdoor vision systems. Its development is restricted by the difficulties in providing a training set of degraded and ground-truth image pairs. In this paper, we develop a novel model that utilizes cycle generative adversarial network through unsupervised learning to effectively remove the requirement of a haze/depth data set.

Modeling of hazy images

The most widely used model to explain the formation of a hazy image can be expressed as follows:

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

where $I(x)$ is the observed hazy image, $J(x)$ is the scene radiance, A is the global atmospheric light, and $t(x)$ is the scene transmission.

Reformulated model

Assuming that the haze mask $H(x)$ is an approximation of transmission map $t(x)$, we can re-express the formula in Eq. (1) as follows:

$$I(x) = H(x)\,J(x) + (1 - H(x))\,A, \tag{2}$$

then we have

$$I(x) = (J(x) - A - \varepsilon)\,H(x) + A, \tag{3}$$

and furthermore we have

$$J(x) = \frac{I(x) - A}{H(x)} + A + \varepsilon, \tag{4}$$

where ε is the constant bias with the empirically default value of 0.2 for the approximation of actual transmission map $t(x)$ and estimated haze mask $H(x)$. Because $H(x)$ denotes an approximation to the transmission map and plays an important role in the process of formulating $J(x)$ from $I(x)$, therefore, we aim to strip $H(x)$ from $I(x)$ to obtain $J(x)$.

Acknowledgment

This work was supported by the Ministry of Science and Technology, Taiwan, under Grant Nos. MOST 108-2221-E-155-034-MY3, and MOST 107-2221-E-155-052-MY2; Fuyao intelligent manufacture simulation project under Grant No.01001701; National Natural Science Foundation of China, China, under Grant No.61602120.

Main idea

This paper proposes an efficient end-to-end dehazing model that is capable of removing haze from degraded images by using only unsupervised learning. Specifically, our model is designed based on a reformulated atmospheric scattering model and utilizes cycle generative adversarial network (CycleGAN) architecture to remove the requirement of an explicit paired haze/depth data set. Qualitative and quantitative evaluations demonstrate the superiority of the proposed model over most existing state-of-the-art methods.

Loss function

Adversarial Loss. Adversarial loss is applied to the generators and the discriminator. For discriminator **D**, the loss function is formulated as:

$$\begin{aligned} & L_{\mathbf{D}}(\mathbf{D}, \mathbf{G}, \mathbf{F}, I, J, A, \varepsilon) \\ = \;& \mathbb{E}\left[\log \mathbf{D}(I)\right] \\ + \;& \mathbb{E}\left[\log\left(1 - \mathbf{D}\left((J - A - \varepsilon)\,\mathbf{F}(J) + A\right)\right)\right]. \end{aligned} \tag{5}$$

For generators **G** and **F**, we add two additional objective functions, namely a total variation (TV) loss L_{TV} and a mean absolute error (MAE) loss L_{MAE}, to encourage better generated images, which can be expressed as:

$$L_{TV}(\mathbf{G}, I) = \mathbb{E}\left[\sum_{\forall j} \sqrt{\left|\partial_x(\mathbf{G}(I))_j\right|^2 + \left|\partial_y(\mathbf{G}(I))_j\right|^2}\right], \tag{6}$$

and the corresponding L_{MAE} is given as

$$L_{MAE}(\mathbf{G}, I) = \mathbb{E}\left[\|\mathbf{G}(I) - I\|_1\right]. \tag{7}$$

Cycle consistency loss. The cycle consistency loss for minimizing the difference between the input image I and reconstructed image I' is formulated as:

$$L_C(\mathbf{G}, \mathbf{F}, I, J, A, \varepsilon) = \mathbb{E}\left[\|((J - A - \varepsilon)\,\mathbf{F}(J) + A) - I\|_1\right]. \tag{8}$$

Framework

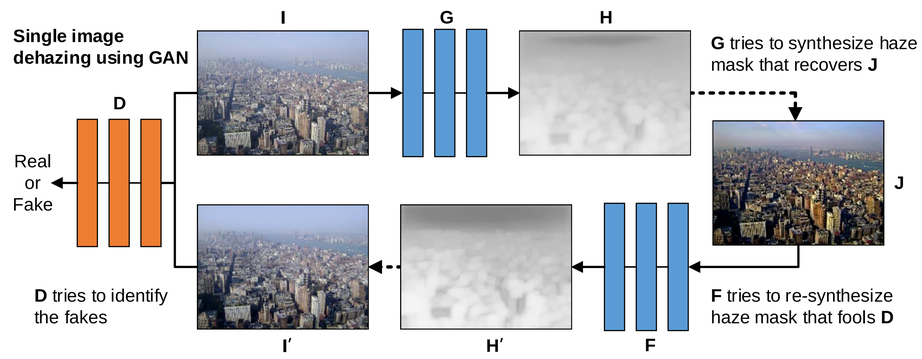

Figure 1: Illustration of the proposed model based on the CycleGAN architecture, where **G** & **F** are the generators and **D** is the discriminator.

Our model includes two mapping functions, $\mathbf{G} : I \xrightarrow{H} J$ and $\mathbf{F} : J \xrightarrow{H'} I$, which are modeled by using two generator networks, **G** and **F**, and a discriminator network **D** to distinguish between the input domain and reconstructed domain.

Results

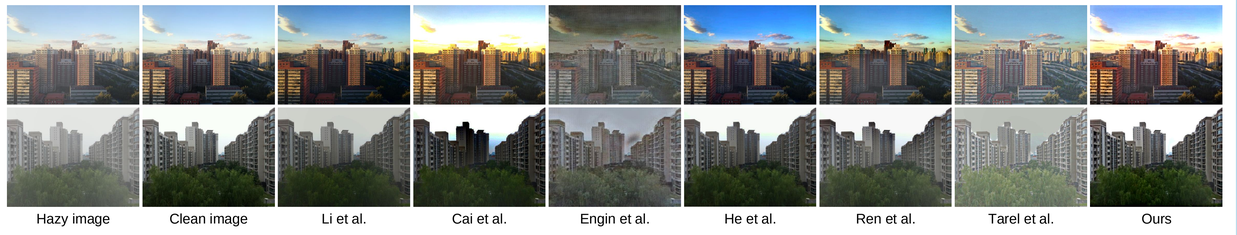

Figure 2: Dehazing results of different state-of-the-art dehazing competitors on synthesized haze images.

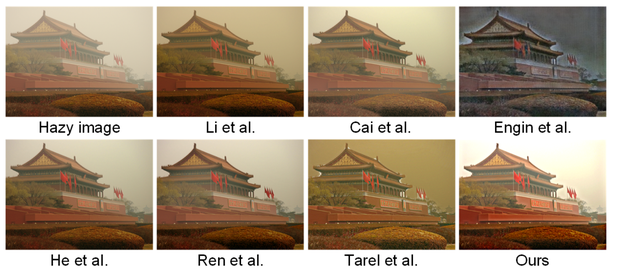

Figure 3: Dehazing results of different state-of-the-art dehazing competitors on real-world haze image "Tiananmen-square."

Figure 4: Dehazing results of different state-of-the-art dehazing competitors on real-world haze image "Traffic jam."

Table 1: Quantitative comparison using PSNR for synthetic testing dataset and FADE for real testing dataset

Dataset	Metrics	Proposed model	Li et al.	Cai et al.	Zhang et al.	Engin et al.	Ren et al.	Tarel et al.
SOTS	PSNR	**21.02**	20.66	15.24	12.65	18.90	19.59	12.80
URHIS	FADE	**0.7296**	0.8680	0.9987	0.7928	1.053	0.8899	0.8064