Scale-aware Conditional Generative Adversarial Network for Image Dehazing

Prasen Kumar Sharma
IIT Guwahati
kumar176101005@iitg.ac.in

Priyankar Jain
IIT Guwahati
priyankar@iitg.ac.in

Arijit Sur
IIT Guwahati
arijit@iitg.ac.in

Abstract

Outdoor images are often deteriorated due to the presence of haze in the atmosphere. Conventionally, the single image dehazing problem aims to restore the haze-free image. Previous successful approaches have utilized various hand-crafted features/priors. However, such images suffer from color degradation and halo artifacts. By way of analysis, these artifacts, in general, prevail around the regions with high-intensity variation, such as edgy structures. This finding inspires us to consider the Laplacians of Gaussian (LoG) of the images which exceptionally retains this information, to solve the problem of single image haze removal. In this line of thought, we present an end-to-end model that learns to remove the haze based on the per-pixel difference between LoGs of the dehazed and original haze-free images. The optimization of the proposed network is further enhanced by using the adversarial training and perceptual loss function. The proposed method has been appraised on Synthetic Objective Testing Set (SOTS) and benchmark real-world hazy images using 16 image quality measures. Based on the Color Difference (CIEDE 2000), an improvement of $\sim$ 15.89% has been observed over the state-of-the-art method, Yang et al. [50]. An ablation study has been presented at the end to illustrate the improvements achieved by various modules of the proposed network.

1. Introduction

One may describe the process of image dehazing as a method to restore the haze-free images from their hazy counterpart, which are affected by the reduced contrast, dull colors and obscured visibility caused by the haze or fog veils. Outdoor images are often deteriorated due to the presence of haze. Haze is an *anthropogenic* atmospheric event in which the suspended *aerosol* particles of size in sub-micrometres such as dust, mist, fog, smoke, sand and other particulates (wet type) degrade the visibility of the scene. The particulates or droplets impair the visibility in one of the following two ways:(a) when the particles obscure the scene-objects behind them, and (b) as a consequence of the light scattering effect at a particular wavelength. Following the majority of existing works [14, 4, 25, 54, 57, 41], the degradation due to haze veils in the images can be analytically expressed as

$$\mathbf{H}(\mathbf{p}) = \mathbf{C}(\mathbf{p})T(\mathbf{p}) + \mathbf{L}(\mathbf{p})(1 - T(\mathbf{p})) \tag{1}$$

where, $\mathbf{H}$ is the hazy image, $\mathbf{C}$ is the haze-free background image, T is the transmission map, $\mathbf{L}$ is the global atmospheric light (homogeneous), demonstrating its intensity and $\mathbf{p}$ is the pixel location. The term $\mathbf{L}(\mathbf{p})(1 - T(\mathbf{p}))$ in the above equation is also known as *Airlight*. The transmission map T denotes an *Exponential Decay* with the scene depth and can be expressed as

$$T(\mathbf{p}) = \exp^{-\beta d(\mathbf{p})} \tag{2}$$

where, β denotes the attenuation coefficient of the atmosphere and d refers to the scene depth. To recover the dehazed image $\bar{\mathbf{C}}$, the estimated atmospheric light ($\bar{\mathbf{L}}$) and transmission ($\bar{T}$) maps can be used by inverting the Eq. 1 as

$$\bar{\mathbf{C}}(\mathbf{p}) = \frac{\mathbf{H}(\mathbf{p}) - \bar{\mathbf{L}}(\mathbf{p})(1 - \bar{T}(\mathbf{p}))}{\bar{T}(\mathbf{p})} \tag{3}$$

1.1. Background

Over the past few years, a significant number of researchers have delved into the domain of single image dehazing, proposed various filter and learning-based methods for the same with outstanding results. One of the early yet significant contributions towards single image dehazing was proposed by He *et al.* [14] that uses the Dark Channel Prior (DCP) to estimate the statistical distribution in clean images taken outdoors. It works based on the assumption that at least one of the three color channels (Red, Green and Blue) have a low-intensity value (*known as Dark Pixels*) that belongs to the haze-free image. Yu *et al.* [52] assumed the scattering model as $L(p)H(p)T(p) + L(p)(1 - T(p))$ and proposed a fast bilateral filter to smoothen the fine texture of the image for single image de-fogging. The term $L(p)H(p)T(p)$ is also known as *Direct Attenuation*. To remove the halo artifacts generated by the DCP [14], He *et al.*

[15] proposed the guided image filtering method which preserves the edges. However, it fails to enhance the contrast in the haze-free images. Meng *et al.* [30] (EIDBR) modelled the dehazing problem as an optimization task based on a weighted L1-norm based contextual regularization. Zhu *et al.* [64] devised a Color Attenuation Prior (CAP) to recover the depth of the hazy image. The transmission map is further estimated using the recovered depth scene to restore the dehazed image. Berman *et al.* [4] introduced a deterministic approach called Non-Local Image Dehazing (NLD) based on the haze-lines that directly estimates the haze-free images.

With the evolution of Convolutional Neural Networks (CNN's) [22], many deep learning-based schemes [37, 25, 54, 41, 50, 12, 6] have been introduced for single image dehazing task. Ren *et al.* [37] proposed a Multi-Scale Convolutional Neural Network (MSCNN) for single image dehazing. The proposed method in [37] directly maps the input hazy image to the transmission map using a coarse deep CNN. Li *et al.* [25] proposed a method AOD-Net which does not predict the transmission and airlight maps separately. Instead, it generates the haze-free image using a lightweight CNN, unifying transmission map and airlight estimation steps within a single unit known as the *K-Estimation Block*. Zhang *et al.* [54] proposed a method called DCPDN that estimates the transmission and atmospheric light maps by using a pyramid [13] densely connected CNN and a U-Net [38] respectively. The haze-free image is then recovered by using Eq. 3. The estimated haze-free image is further enhanced using a joint discriminator. Santra *et al.* [41] proposed a CNN based Patch Quality Comparator (PQC) to estimate the dehazed images. The method proposed in [10] uses unpaired training based on the Cycle-GAN framework [63] for image dehazing task. Yang *et al.* [50] leveraged the benefits of deep learning-based and prior-based methods in a single framework for haze removal problem.

1.2. Our contributions

Most of the methods discussed above have been successful in the single image dehazing task. However, a few of such haze-free images suffer from color degradation and halo artifacts that prevail around the high-intensity regions and edgy structures. Based on the inverse scattering model (Eq. 3) to recover $\mathbf{C}$, two essential parameters $T, \mathbf{L}$ have to be estimated. Existing approaches except a few [25, 10], separately predict the transmission and atmospheric light maps. Whereas the proposed scheme directly approximates the distribution of haze in the hazy image $\mathbf{H}$ and recover the haze-free image $\mathbf{C}$.

Scale-space invariant deep model. Deep learning models, in case of image denoising, may consider every object in an image at same scale-space. As a result, the denoised images may suffer from blurriness and halo artifacts. It is based on the fact that a traditional CNN model does not aware of scale-space of an object. Therefore, this paper makes the **first attempt** to study the behaviour of a scale-space aware deep learning-based model. For this, we define a *new* loss function which is based on the LoG of an image. LoG preserve the finer details of the edgy structures in the images at different scale-space which can be lost during the process of dehazing. It has been observed during our experimentation that difference of LoG's between clean and dehazed images can be used as a cost function to optimize the proposed model. Such incorporations may help the model to learn the scale-space [27] of every object in an image. An analogy, based on the use of perceptual loss function [19] in a deep network to recover the high-frequency details of an image, may support this argument. Therefore, the contributions of this paper can be summarised as follows:

1. A novel scale-space aware Conditional Generative Adversarial Network (CGAN) based method has been proposed for single image dehazing. In addition to the adversarial training, the perceptual loss function has been used to enhance the visual quality of the dehazed images.
2. We introduce the LoG difference between clean and dehazed images as a cost function to optimize the proposed CGAN-based model and wipe out the halo artifacts in the dehazed images by retaining the edgy structures more precisely.
3. A random data augmentation has been done when training to improve the efficiency of the proposed model further. A brief study of the same has been given in this paper followed by an ablation study, which is presented at the end of this paper, along with extensive experiments.

The rest of this manuscript is organized as follows: A brief introduction of the adopted modules in the proposed method is presented in Section 2. Section 3 describes the proposed method. The experimental details and results are elaborated in Section 4 and finally, the paper is concluded in Section 5.

2. Priliminaries

2.1. Laplacians of Gaussian (LoGs)

Laplacians are isotropic measure of second-order spatial derivative filters used to find the regions of rapid intensity variations (edgy structures) in an image. These filters are sensitive to the noise present in the image, and hence, it is common first to smooth the given image using the Gaussian filter with standard deviation σ - the combined filter, in general, known as Laplacian of Gaussian (LoG). In general, it

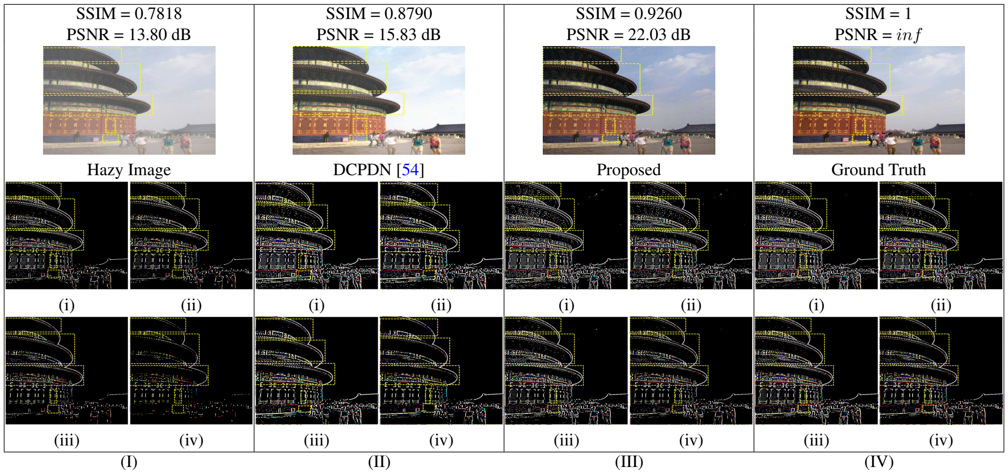


Figure 1. Sample Laplacians of Gaussian (LoG) filters of (I) Hazy, (II) Dehazed by using [54], (III) Proposed and (IV) Clean images. For each in I,II,III and IV, (i) $\mathsf{G}(\mathbf{m}, \mathbf{n}, k\sigma) - \mathsf{G}(\mathbf{m}, \mathbf{n}, \sigma)$, (ii) $\mathsf{G}(\mathbf{m}, \mathbf{n}, k^2\sigma) - \mathsf{G}(\mathbf{m}, \mathbf{n}, k\sigma)$, (iii) $\mathsf{G}(\mathbf{m}, \mathbf{n}, k^3\sigma) - \mathsf{G}(\mathbf{m}, \mathbf{n}, k^2\sigma)$, and (iv) $\mathsf{G}(\mathbf{m}, \mathbf{n}, k^4\sigma) - \mathsf{G}(\mathbf{m}, \mathbf{n}, k^3\sigma)$. A detailed description of the notations has been provided in subsection 2.1.

is written as

$$\begin{aligned} \mathbf{L}(\mathbf{m}, \mathbf{n}) &= \nabla^2 \mathbf{F}(\mathbf{m}, \mathbf{n}) = \frac{\partial^2 \mathbf{F}}{\partial \mathbf{m}^2} + \frac{\partial^2 \mathbf{F}}{\partial \mathbf{n}^2} \\ &= -\frac{1}{\pi\sigma^4}\Big[1 - \frac{\mathbf{m}^2 + \mathbf{n}^2}{2\sigma^2}\Big] \exp\Big(-\frac{\mathbf{m}^2 + \mathbf{n}^2}{2\sigma^2}\Big) \end{aligned} \tag{4}$$

where, $\mathbf{F}$ is a 2D signal (an image in this case) with pixel location $(\mathbf{m}, \mathbf{n})$. Theoretically, the Difference of Gaussian can be used to closely approximate the LoG [27, 28] with various σ values as

$$\mathsf{G}(\mathbf{m}, \mathbf{n}, k\sigma) - \mathsf{G}(\mathbf{m}, \mathbf{n}, \sigma) \approx (k-1)\sigma^2 \Delta^2 \mathsf{G} \tag{5}$$

where, $\sigma^2 \Delta^2 \mathsf{G}$ denotes scale-normalized LoG and k, σ are typically set to $\sqrt{2}, 1.6$ respectively [27]. The 2D Gaussian kernel is defined as

$$\mathsf{G}(\mathbf{m}, \mathbf{n}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{\mathbf{m}^2 + \mathbf{n}^2}{2\sigma^2}\Big) \tag{6}$$

In this paper, we have used four LoG filters using five Gaussian kernels of size 7×7 with σ, $k\sigma$, $k^2\sigma$, $k^3\sigma$ and $k^4\sigma$ as their standard deviations. Finally, the *per-pixel* difference between LoGs of clean and the dehazed image generated by the proposed model is used as a cost function to train our model. A few of the LoG filters of a sample hazy image, the corresponding dehazed images generated by using the proposed method and Zhang *et al.* [54], are shown in Figure 1. It can be clearly observed that the LoG filters of the dehazed image (generated by using the proposed model), have retained the edgy information better than the same of the dehazed image (estimated by using Zhang *et al.* [54]). Thus, the use of LoG difference (between *dehazed* and *clean* images) as a cost function to train a deep CNN can be useful for solving the single image dehazing problem.

2.2. Generative Adversarial Networks (GANs)

The proposed model is based on the Conditional GAN [31] framework which consists of two main sub-networks called "Generator" and "Discriminator" which are denoted as ϕ_G, ϕ_D respectively in our case. While ϕ_G aims to dehaze the given hazy image, ϕ_D learns to distinguish between the real (actual clean) and fake (dehazed) images. The proposed ϕ_G learns from its adversary ϕ_D until a Nash Equilibrium [34] is achieved by playing a 2-player minimax game based on the following equation

$$\min_{\phi_G} \max_{\phi_D} L_{GAN} \tag{7}$$

where, ϕ_G, ϕ_D learn by stochastically descending, asscending their parameters respectively and L_{GAN} can be written as

$$\begin{aligned} L_{GAN} = &\lambda_A.(\mathbb{E}_{\mathbf{H}\sim P_{haze}}\, log(1 - \phi_D(\phi_G(\mathbf{H})))) + \\ &\mathbb{E}_{\mathbf{C}\sim P_{clean}}\, log(\phi_D(\mathbf{C})) \end{aligned} \tag{8}$$

with λ_A as a weight constant.

3. Proposed approach

In this section, we first present the architecture of the proposed model as shown in Figure 2, which is based on a Conditional GAN framework followed by the cost functions incorporated for the single image dehazing task. The proposed model comprises of two main sub-networks, Generator (ϕ_G) and Discriminator (ϕ_D). The regimes of operations of ϕ_G and ϕ_D are as follows.

Generator (ϕ_G) model takes input as hazy image $\mathbf{H}$ in RGB color space and predicts the corresponding dehazed image $\bar{\mathbf{C}}$. It consists of an encoder-decoder [38] architecture which has been useful in various image restoration

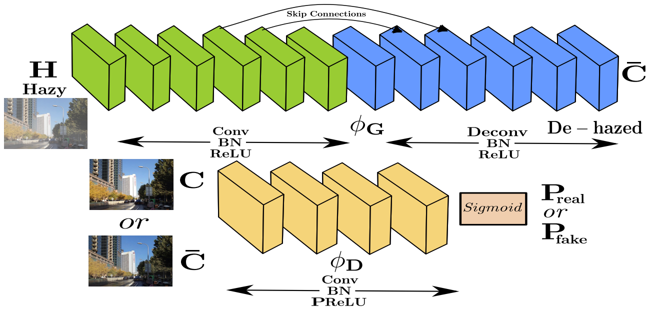


Figure 2. An overview of the proposed model for the single image dehazing problem.

tasks. The encoder part consists of 6 convolutional layers, each with 64 kernels followed by Batch Normalization (BN) [17] and ReLU activation function. Each kernel has a spatial dimension of 3×3 with stride and padding of 1. On the other hand, the decoder part comprises of 6 transpose convolutional layers which are also known as *Deconvolution*. Each deconvolution layer consists of 64 kernels except the last, with the spatial dimension of 3×3 and stride, padding of 1 followed by BN and ReLU. Orhan *et al.* [35] have shown that skip connections may reduce the singularities, such as elimination, overlap and those caused by the linear dependence of the nodes, that slow down the training process of deep CNN. Therefore, the skip connections have been assigned between layers 4, 5 of the encoder to layers 3, 2 of the decoder. The proposed ϕ_G directly estimates the dehazed image $\bar{\mathbf{C}} = \phi_G(\mathbf{H})$ from the input hazy image $\mathbf{H}$. Unlike Zhang *et al.* [54], the proposed scheme preserves the spatial dimension of the input ($\mathbf{H}$) and output ($\bar{\mathbf{C}}$) images, thereby achieving the shape invariant nature.

Discriminator (ϕ_D) learns to maximize the probability of precisely classifying the input samples into real or fake dehazed images. As discussed in sub-section 2.2, this, in turn, helps the ϕ_G to generate natural dehazed images. The proposed ϕ_D consists of 4 convolution layers with 8, 16, 32, and 3 kernels respectively, followed by BN and PReLU [16] activation function. Each kernel in ϕ_D has a spatial dimension of 3×3 with stride and padding of 1. The output of the ϕ_D is the mean sigmoid over the feature maps from the last convolution layer.

3.1. Loss function

Let $\phi_G(\mathbf{H}) \in [0,1]^{c\times w\times h}$ be the dehazed image estimated by the proposed model with c, w, h as channels, width and height respectively. The conventional *per-pixel* loss (L_E) between dehazed and ground truth ($\mathbf{C}$) images can be written as

$$L_E = \sum_{c_i,w_i,h_i} \left\| \phi_G(\mathbf{H})^{c_i,w_i,h_i} - \mathbf{C}^{c_i,w_i,h_i} \right\|_2^2 \tag{9}$$

In general, the noise present in an image exhibit high-frequency nature. During image de-noising by using a traditional CNN, the use of Euclidean distance as a cost function may incur the loss of high-frequency details of the image along with the noise removal [62]. As a result, the de-noised images appear to be blurry and degraded. The perceptual loss function proposed by Johnson *et al.* [19] has been used in the majority of the image de-noising and restoration problems [40, 56, 51, 45, 55, 60, 23] in recent times to overcome this drawback by retaining the high-frequency details of an image. The perceptual cost function is defined as a difference between high-level features of predicted and target images extracted by using a pre-trained CNN. In this case, initial five layers (l) of a pre-trained VGG16 [44] model (V) have been used to extract the features [1]. The perceptual loss function (L_P) can be expressed as

$$L_P = \sum_{l} \sum_{c_i,w_i,h_i} \left\| V_l(\phi_G(\mathbf{H}))^{c_i,w_i,h_i} - V_l(\mathbf{C})^{c_i,w_i,h_i} \right\|_2^2 \tag{10}$$

Adversarial training has been beneficial in many of the denoising tasks [5, 24, 53, 21]. In this case, the proposed generator ϕ_G can learn from its adversary ϕ_D based on the adversarial loss for a set of N training samples defined as

$$L_A = -\frac{1}{N}\sum_{i=1}^{N} log\ \phi_D(\phi_G(\mathbf{H})_i) \tag{11}$$

Intuitively, the use of Euclidean, Perceptual and Adversarial losses to train the proposed model may have given the visually appealing results. However, it is observed that a few of the estimated dehazed images suffer from halo and checkerboard artifacts. The Laplacians of Gaussian (LoG) filters (f), as discussed in sub-section 2.1, capture these finer details and can be used as a cost function based on the following equation [2].

$$L_{LoG} = \sum_{f,c_i,w_i,h_i} \left\| \mathbf{L}_f(\phi_G(\mathbf{H}))^{c_i,w_i,h_i} - \mathbf{L}_f(\mathbf{C})^{c_i,w_i,h_i} \right\|_2^2 \tag{12}$$

Therefore, the aggregated loss to train the proposed generator model ϕ_G can be written as

$$L_{\phi_G} = \lambda_E.L_E + \lambda_A.L_A + \lambda_P.L_P + \lambda_G.L_{LoG} \tag{13}$$

where, $\lambda_E, \lambda_A, \lambda_P$ and λ_G are the cost weights. Moreover, the final objective function of the proposed model for single image dehazing task defined in Eq. 7 becomes

$$\underset{\phi_G}{min}\ \underset{\phi_D}{max}\ \{L_{GAN} + \lambda_E.L_E + \lambda_P.L_P + \lambda_G.L_{LoG}\} \tag{14}$$

[1] https://github.com/pytorch/examples/tree/master/fast_neural_style

[2] Parameters of the laplacian model are not updated when training.

Measure	Behaviour	Input	DCP [14] TPAMI '11	EIDBR [30] ICCV '13	CAP [64] TIP '15	DEFADE [7] TIP '15	MSCNN [37] ECCV '16	NLD [4] CVPR '16	AOD-Net [25] ICCV '17	DCPDN [54] CVPR '18	PQC [41] TIP '18	PLD [50] ECCV '18	DSIE [12] CVPRW '19	Proposed
SSIM	▲	0.8148	0.8254	0.7665	0.7584	0.8072	0.8389	0.8210	0.9062	0.8722	0.8791	0.8889	0.7580	0.8941
PSNR	▲	15.95	17.33	15.47	18.12	18.89	19.48	17.97	20.43	18.04	19.28	19.45	15.57	20.57
VIF	▲	0.7348	0.5529	0.5766	0.5047	0.5380	0.5492	0.6003	0.6638	0.7502	0.7892	0.8150	0.4969	0.6890
MSE	▼	2.084	1.566	2.850	1.216	1.238	0.822	1.413	0.695	1.279	1.021	0.938	2.062	0.642
UQI	▲	0.7765	0.8581	0.8210	0.7565	0.8314	0.8649	0.8511	0.8962	0.8434	0.8552	0.8653	0.8015	0.8754
LPIPS	▼	0.1038	0.1604	0.1855	0.1348	0.1425	0.1155	0.1465	0.0795	0.0885	0.0751	0.0792	0.2320	0.0728
MS-SSIM	▲	0.9263	0.8859	0.8399	0.8796	0.8948	0.9101	0.8920	0.9435	0.9435	0.9524	0.9521	0.8296	0.9456
TV-Error	▼	0.7802	1.1151	1.7031	0.9753	1.3706	1.0559	1.4894	1.1302	0.8623	1.2920	1.1211	1.3097	1.1720
NIQE	▼	2.9053	2.8033	3.2193	1.5827	2.5663	2.4603	3.1550	2.4240	2.2705	2.7443	2.7819	4.0628	2.1452
FSIM	▲	0.9448	0.9445	0.8769	0.9405	0.9441	0.9598	0.9361	0.9579	0.9635	0.9665	0.9701	0.8830	0.9687
CIEDE 2000	▼	23.80	15.50	18.82	17.88	16.30	14.83	16.04	12.11	17.66	14.61	14.22	21.08	11.96
Haar PSI	▲	0.8616	0.7866	0.6382	0.8219	0.8180	0.8582	0.7633	0.8768	0.8418	0.8804	0.9037	0.6552	0.8752
GMSD	▼	0.0522	0.0643	0.1162	0.0643	0.0561	0.0438	0.0677	0.0462	0.0572	0.0425	0.0333	0.1137	0.0325
BRISQUE	▼	15.79	23.43	27.21	24.18	22.13	22.63	22.32	12.95	23.40	16.06	15.19	17.25	14.67
SpEED-QA	▼	8.66	13.88	28.61	17.84	16.53	15.21	17.33	12.99	11.96	12.00	11.15	25.23	10.17
fom	▲	-	0	0	1	0	0	0	3	0.6	1.8	3	0	5.6

Table 1. Quantitative comparison on the SOTS (Outdoor) dataset. Best and second best results are shown in blue and red colors respectively. A figure of merit (fom) decides the final score as number of (0.6×Best+0.4×Second Best)/Total Metrics. TV-Error is 10^7.

Measure	Behaviour	Input	DCP [14]	EIDBR [30]	CAP [64]	DEFADE [7]	MSCNN [37]	NLD [4]	AOD-Net [25]	DCPDN [54]	Cycle-Dehaze [10]	MAMF [6]	MS-PPD [57]	PQC [41]	PLD [50]	DSIE [12]	Proposed
NIQE	▼	4.5196	3.6763	5.6602	3.3444	3.6283	4.4219	4.7583	4.1259	5.9101	4.5506	8.6962	5.5536	4.4316	4.5568	8.1091	**4.3041**
BRISQUE	▼	18.49	21.80	22.55	19.11	22.98	20.29	20.19	18.73	20.57	13.23	21.62	19.73	19.78	22.42	21.03	**18.65**
BLINDS II	▲	5.87	7.20	11.34	6.44	10.08	10.27	10.65	9.70	14.00	10.48	12.18	12.18	10.48	6.50	12.27	**13.44**

Table 2. Quantitative results on the *Benchmark* images provided by Fattal *et al.* [11].

Baseline	Input	**M**-L_E	**M**-L_P	**M**-$L_E + L_P$	**M**-$L_E + L_A$	**M**-$L_E + L_A + L_P$	**M**-NDA	Proposed
SSIM	0.8148	0.7780	0.8220	0.8321	0.7937	0.8602	0.8750	0.8941
PSNR	15.95	17.78	16.80	17.83	17.49	18.58	19.28	20.57

Table 3. Quantitative comparison of the proposed method with the baseline configurations on the SOTS (Outdoor) test set.

4. Experiments and results

This section illustrates the details of the experimental setup and dataset used for the training and testing of the proposed model. A concise description of image assessment metrics chosen, followed by an ablation study and comparison with the existing methods on both synthetic and real-world hazy images are given.

4.1. Datasets and training details

We have chosen the training dataset provided by Zhang *et al.* [54], which consists of 4000 indoor images. In addition, we have also included 45 hazy outdoor images provided by Ancuti *et al.* [2]. During training, following [61], we have augmented the input pairs by using (1) *Random rotation*, (2) *Vertical flip*, (3) *Horizontal flip* and, (4) *Random cropping*. Each data augmentation technique has a probability of 0.5 to be applied to the input pair. Whereas, the input pair will be augmented with the expectation (p_{dat}) of 0.5. Input pair is *rotated* with the degree randomly chosen between [1°, 359°]. *Cropping* is done with the size $u \times u; u \in [8, 256]$ randomly chosen at a random location in the input pair. For each input pair when training, a randomly selected transformation between {*Horizontal flip, Vertical flip*}, in addition to *Rotation* and *Cropping*, is applied in random order. For testing, we have used synthetic dataset (SOTS) provided by Li *et al.* [26] which consists of 500 outdoor and indoor images. We have also evaluated our proposed work on the benchmark test set[3] provided by Fattal *et al.* [11] and real-world hazy images.

The proposed network is trained on a Nvidia Tesla GPU using the Torch framework [9] for 104 epochs. For training, we have experimentally chosen $\lambda_E = \lambda_A = \lambda_P = \lambda_G = 1$ for the losses in estimating the dehazed image. With the batch size of 5 images, Adam [20] optimization algorithm with a fixed learning rate of 2×10^{-4} has been used when training. The training samples are resized to 496×496.

4.2. Evaluation metrics

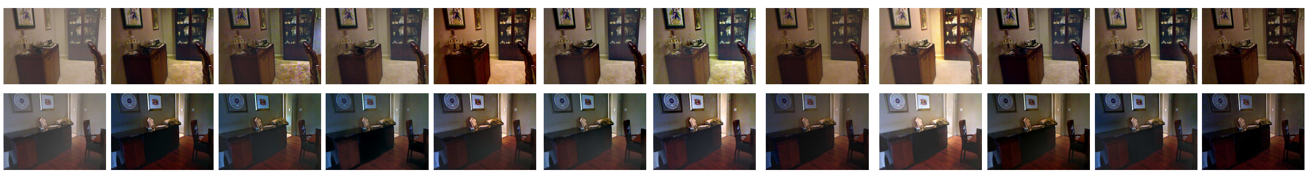
The proposed scheme has been compared with existing approaches using following 16 full-reference and no-reference image quality metrics: *Full-reference* - SSIM [47], PSNR, Visual information fidelity (VIF) [43], Universal image quality index (UQI) [46], Learned perceptual image patch similarity (LPIPS) [59], Mean squared error (MSE), Multi-scale structural similarity (MS-SSIM) [48], Feature similarity (FSIM) index [58], Color difference (CIEDE 2000) [42], Haar wavelet-based perceptual similarity index (HaarPSI) [36], Gradient magnitude similarity deviation (GMSD) [49] and SpEED-QA: Spatial Efficient Entropic Differencing for Image and Video Quality [3]. *No-reference* - Total variation error (TV-Error) [1], Naturalness image quality evaluator (NIQE) [33], Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain (BLINDS II)[4] [39], and Blind/referenceless image spatial quality evaluator (BRISQUE) [32]. In this paper, the behaviour of these evaluation norms are described by using following symbols: ▲ (denotes higher is better)

[3] http://www.cs.huji.ac.il/~raananf/projects/dehaze_cl/results/

[4] Images are resized to 512 × 512 to reduce the computation time.

Measure	Behaviour	Input	DCP [14] TPAMI '11	EIDBR [30] ICCV '13	CAP [64] TIP '15	DEFADE [7] TIP '15	MSCNN [37] ECCV '16	NLD [4] CVPR '16	AOD-Net [25] ICCV '17	DCPDN [54] CVPR '18	PQC [41] TIP '18	PLD [50] ECCV '18	DSIE [12] CVPRW '19	Proposed
SSIM	▲	0.6942	0.8595	0.7682	0.8171	0.7565	0.7955	0.7775	0.8260	0.7283	0.8513	0.8445	0.6791	0.8598
PSNR	▲	11.97	20.04	16.13	18.97	17.20	17.12	17.29	19.07	13.19	20.25	20.28	13.78	19.68
VIF	▲	0.6858	0.7179	0.6782	0.6468	0.6656	0.7305	0.7808	0.6698	0.7148	0.7785	0.7765	0.4563	0.6945
MSE	▼	4.856	0.776	2.024	0.983	1.531	1.592	1.427	1.144	3.723	0.722	0.728	3.114	0.914
UQI	▲	0.6435	0.8492	0.7681	0.7879	0.7326	0.7577	0.7698	0.8019	0.6774	0.8228	0.8144	0.7029	0.8406
LPIPS	▼	0.1999	0.1033	0.1584	0.1108	0.1449	0.1116	0.1432	0.1014	0.1531	0.0771	0.0779	0.2690	0.1128
MS-SSIM	▲	0.8771	0.9266	0.8767	0.9095	0.8924	0.9269	0.8968	0.9199	0.8915	0.9432	0.9409	0.7858	0.9267
TV-Error	▼	0.4808	0.7683	0.9635	0.6685	0.8218	0.6657	0.9738	0.6626	0.6302	0.8343	0.7430	1.0559	0.8786
NIQE	▼	1.8977	1.2869	1.7616	1.3165	1.6364	1.7327	1.7397	1.6264	2.3240	1.6780	1.7007	3.7639	0.9852
FSIM	▲	0.9112	0.9418	0.8978	0.9309	0.9289	0.9473	0.9203	0.9410	0.9253	0.9569	0.9573	0.8637	0.9613
CIEDE 2000	▼	34.73	11.85	18.76	15.24	18.91	21.39	18.34	16.14	29.00	12.41	13.48	23.36	12.14
Haar PSI	▲	0.7599	0.8313	0.6751	0.8138	0.7764	0.8514	0.7275	0.8336	0.7180	0.8663	0.8797	0.6332	0.8155
GMSD	▼	0.0879	0.0668	0.1129	0.0793	0.0805	0.0656	0.0935	0.0652	0.1134	0.0582	0.0547	0.1398	0.0517
BRISQUE	▼	38.92	34.86	33.48	36.33	33.71	35.44	33.42	34.74	41.74	33.78	34.10	29.87	32.95
SpEED-QA	▼	15.26	15.83	24.14	18.61	17.49	15.16	20.00	14.00	18.46	14.16	12.92	29.54	10.77
fom	▲	-	2	0	0	0	0	0.6	0.4	0.6	3	3.6	0.6	4.2

Table 4. Quantitative comparison on the SOTS (Indoor) dataset. Best and second best results are shown in blue and red colors respectively. A figure of merit (fom) decides the final score as number of (0.6×Best+0.4×Second Best)/Total Metrics. TV-Error is 10^7.



Input DCP [14] EIDBR [30] CAP [64] DEFADE [7] MSCNN [37] NLD [4] AOD-Net [25] DCPDN [54] PQC [41] PLD [50] Proposed

Figure 3. Subjective comparison of the proposed method with the existing state-of-the-art schemes on the SOTS (Indoor) test images.

Outdoor

Measure	Behaviour	Cycle-Dehaze [10]	MAMF [6]	MS-PPD [57]	Proposed
SSIM	▲	0.7850	0.7502	0.8119	0.8941
PSNR	▲	12.93	17.81	17.23	20.57
SpEED-QA	▼	10.64	21.50	14.84	10.17

Indoor

Measure	Behaviour	Cycle-Dehaze [10]	MAMF [6]	MS-PPD [57]	Proposed
SSIM	▲	0.7748	0.7269	0.7687	0.8598
PSNR	▲	17.18	17.16	16.67	19.68
SpEED-QA	▼	16.86	24.36	18.97	10.77

Table 5. Comparison with other existing methods on SOTS.

and, ▼ (denotes lower is better). The detailed description of adopted evaluation metrics has been given in the supplementary material.

4.3. Ablation study

This sub-section presents an ablation study of the proposed method. We have compared the proposed model with the baseline configurations (**M**-X, where X denotes the proposed model is trained using only loss X) and **M**-NDA refers to the proposed model trained without using adopted data augmentation based on the Eq. 14. It can be observed from the Table 3 that the inclusion of L_P in addition to L_E and L_A has shown a significant improvement. Further, the use of L_{LoG} has contributed an average improvement of $\sim$ 1.72%, $\sim$ 3.76% in SSIM and PSNR, respectively, over the model **M**-$L_E + L_A + L_P$. A noticeable increment of $\sim$ 2.18%, $\sim$ 6.69% in SSIM and PSNR, respectively, is further observed when the data augmentation techniques, summarised in sub-section 4.1, are used during the training.

4.4. Comparison with State-of-the-Art Methods

Evaluation on synthetic dataset. Tables 1, 4, and 5 present the quantitative comparison of the proposed scheme with 14 state-of-the-art methods using 15 image quality metrics as mentioned in earlier subsection 4.2. Based on the proposed figure of merit (fom) in Tables 1 and 4, it can be observed that the proposed scheme has shown a significant improvement over the existing methods [50, 41, 54, 25]. Despite the fact that SSIM value achieved by [25] on SOTS (outdoor) test set is $\sim$ 1.35% higher, the proposed scheme outperforms [25] by a noticeable margin of $\sim$ 86% in overall ranking (fom). One of the important aspect of the single image haze removal problem is color restoration. To evaluate this, we have employed CIEDE which essentially measures the color difference between two images. As reported in Table 1, the proposed scheme has outperformed the existing methods [14, 64, 37, 4, 25, 50, 54, 41] with the *lowest* CIEDE value of 11.96.

Qualitative analysis on outdoor and indoor test sets, as shown in Figures 4, 6 respectively, proves the supremacy of the proposed scheme over other methods. Unlike [12, 10, 57, 54], the proposed scheme does not suffer from color degradation. As shown in Figure 4(c), results obtained by using [37, 25, 41] still contain the hazy part and obscured edgy structures. Whereas, the result obtained by using the proposed scheme is free from such artifacts.

The primary reason behind such improvement may be the use of perceptual loss [19] and the introduced LoG difference as the cost functions. Especially, the LoG loss,

	DCP [14]	EIDBR [30]	CAP [64]	DEFADE [7]	MSCNN [37]	NLD [4]	AOD-Net [25]	DCPDN [54]	PQC [41]	PLD [50]	DSIE [12]	Proposed
Platform	MATLAB [29]	MATLAB [29]	MATLAB [29]	MATLAB [29]	MATLAB [29]	MATLAB [29]	Pycaffe [18]	Torch [9]	Keras [8]	MATLAB [29]	Torch [9]	Torch [9]
Time	16.37	2.64	0.78	34.84	1.71	5.05	0.48	0.13†	29	1.68	6.10‡	0.05

Table 6. Average running time (in seconds) on the test set SOTS (Indoor). † Tested with images of size 512×512. ‡ On CPU.

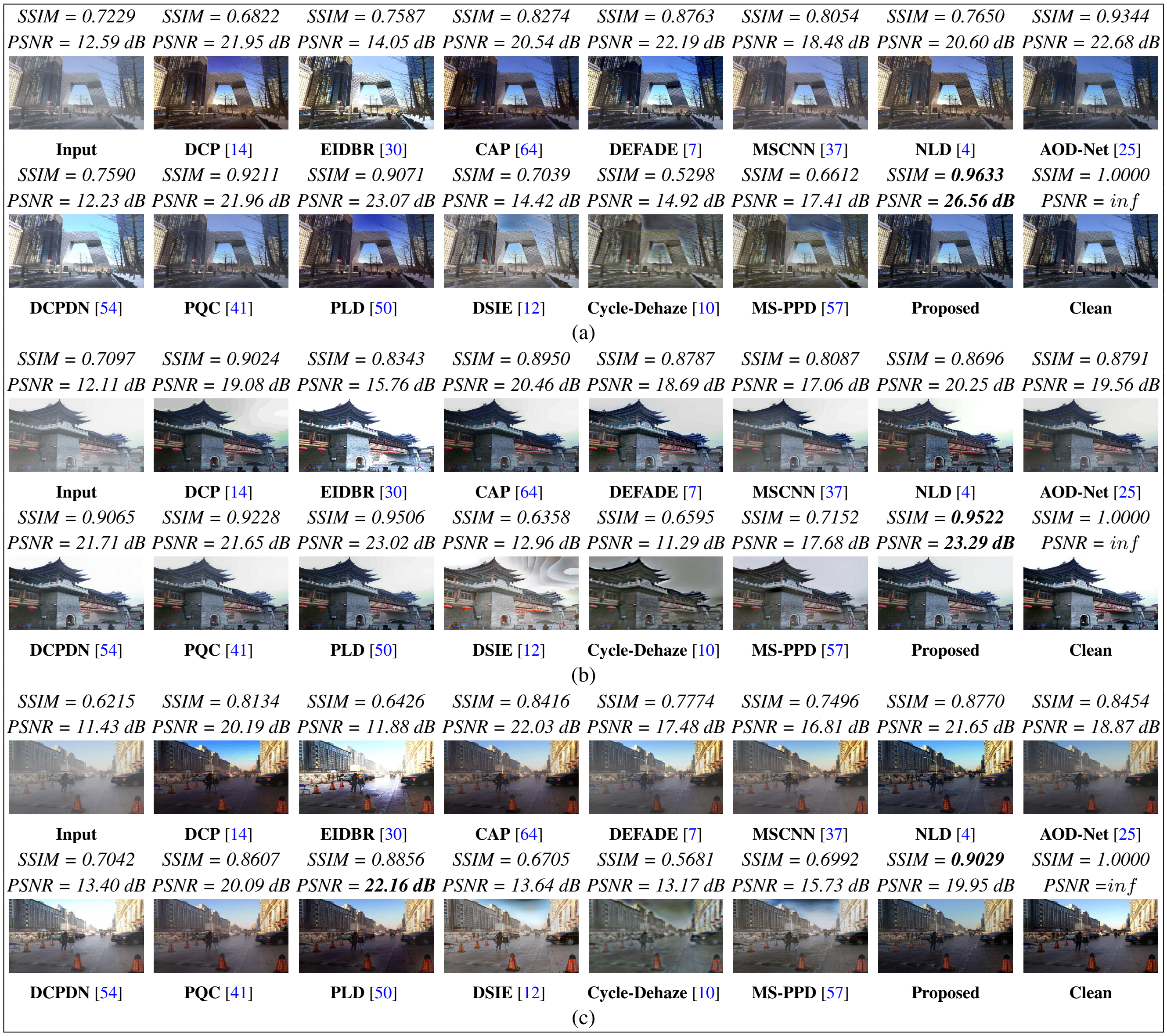


Figure 4. Subjective evaluation of the proposed method with existing schemes in terms of SSIM and PSNR(dB) on SOTS (Outdoor) images.

which may have improved the efficiency of the proposed model by considering the scale-space of the objects from the initial epoch. The proposed method has also been tested on the *benchmark images* provided by the Fattal *et al.* [11] and results are tabulated in the Table 2.

Evaluation on real-world dataset. The proposed model has been evaluated on several real-world hazy images, as shown in Figure 5. It can be observed that the earlier existing approaches such as [14, 64] tend to under dehaze the given images whereas schemes such as [30, 7, 4] have produced the dehazed images with oversaturated tones. It may be because these methods have used a hand-crafted feature such as dark channel prior, to estimate the haze distribution in the images. As a result, the models may not have generalized well on a variety of hazy images. Recent deep learning based approaches such as [37, 25, 41, 50] have been successful compared to the previous models. However, such methods have failed to address

Figure 5. Subjective comparison of the proposed model with the existing methods on the real-world hazy images.

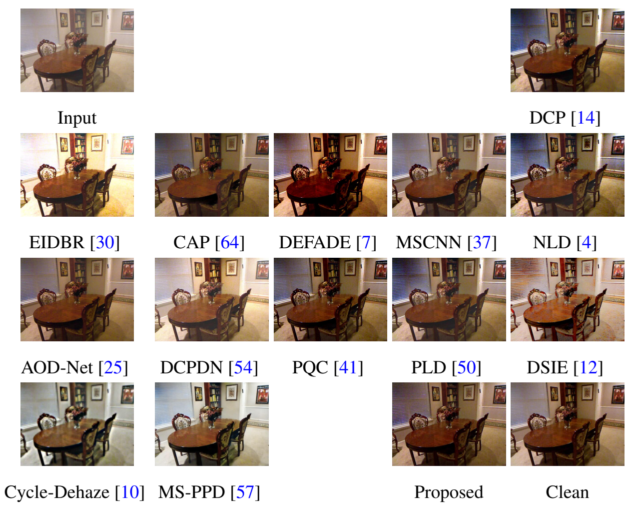


Figure 6. Comparison with the existing schemes on a synthetic hazy image (Indoor).

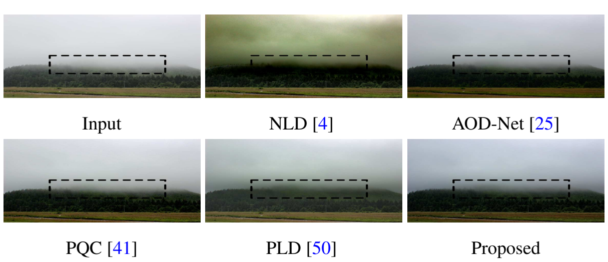


Figure 7. **Failure case**. The proposed model does not perform well on the images with dense haze.

the perceptual quality of the dehazed images. The proposed scheme has produced visually appealing results compared to other existing methods.

Run-time comparison and failure case. The runtime comparison of the proposed scheme with existing methods has been shown in Table 6. It can be observed that the proposed model takes about ~ 0.05 seconds to test an image with an average size of 620×460. The proposed method fails to address the images with dense haze, as shown in Figure 7. However, the perceptual quality of the dehazed image recovered by using the proposed scheme is better than the same by using the existing methods [4, 25, 41, 50].

5. Conclusions

In this work, we have presented an end-to-end deep learning based approach for the single image haze removal problem. The proposed scheme is built upon the conditional GAN framework and directly estimates the dehazed image. We have shown the better preservation of the edgy structures in the LoGs of the hazy images, which inspired us to consider the LoG difference as a cost function. The generalization of the proposed model has been verified by using three benchmark test sets, namely: SOTS (Indoor and Outdoor), Fattal *et al.* [11] and, real-world hazy images. Despite the fact that the proposed model fails to address the images with dense haze, it has been evaluated using 15 image quality assessment metrics, and extensive comparison with existing methods proves it's primacy.

6. Acknowledgement

We acknowledge the Department of Biotechnology, Govt. of India for the financial support for the project BT/COE/34/SP28408/2018.

References

[1] https://www.tensorflow.org/api_docs/python/tf/image/total_variation.

[2] C. O. Ancuti, C. Ancuti, R. Timofte, and C. D. Vleeschouwer. O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images. In *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 754–762, 2018.

[3] C. G. Bampis, P. Gupta, R. Soundararajan, and A. C. Bovik. Speed-qa: Spatial efficient entropic differencing for image and video quality. *IEEE Signal Processing Letters*, 24(9):1333–1337, Sep. 2017.

[4] D. Berman, T. treibitz, and S. Avidan. Non-local image dehazing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[5] J. Chen, J. Chen, H. Chao, and M. Yang. Image blind denoising with generative adversarial network based noise modeling. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[6] Y. Cho, J. Jeong, and A. Kim. Model-assisted multiband fusion for single image enhancement and applications to robot vision. *IEEE Robotics and Automation Letters*, 3(4):2822–2829, Oct 2018.

[7] L. K. Choi, J. You, and A. C. Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. *IEEE Transactions on Image Processing*, 24(11):3888–3901, Nov 2015.

[8] F. Chollet et al. Keras. https://keras.io, 2015.

[9] R. Collobert, K. Kavukcuoglu, and C. Farabet. Torch7: A matlab-like environment for machine learning. In *BigLearn, NIPS Workshop*, 2011.

[10] D. Engin, A. Genç, and H. K. Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018.

[11] R. Fattal. Dehazing using color-lines. *ACM Trans. Graph.*, 34(1):13:1–13:14, Dec. 2014.

[12] T. Guo, X. Li, V. Cherukuri, and V. Monga. Dense scene information estimation network for dehazing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[13] D. Han, J. Kim, and J. Kim. Deep pyramidal residual networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6307–6315, July 2017.

[14] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, Dec 2011.

[15] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(6):1397–1409, June 2013.

[16] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, pages 1026–1034, Washington, DC, USA, 2015. IEEE Computer Society.

[17] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 448–456. JMLR.org, 2015.

[18] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[19] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, 2016.

[20] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

[21] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8183–8192, June 2018.

[22] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, Dec 1989.

[23] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[24] C. Ledig, L. Theis, F. Huszr, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, July 2017.

[25] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[26] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019.

[27] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1994.

[28] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov 2004.

[29] MATLAB. The MathWorks Inc., Natick, Massachusetts.

[30] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.

[31] M. Mirza and S. Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.

[32] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, Dec 2012.

[33] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a completely blind image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, March 2013.

[34] J. F. Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[35] E. Orhan and X. Pitkow. Skip connections eliminate singularities. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[36] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand. A haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication*, 61:33 – 43, 2018.

[37] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, 2016.

[38] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[39] M. A. Saad, A. C. Bovik, and C. Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE Transactions on Image Processing*, 21(8):3339–3352, Aug 2012.

[40] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[41] S. Santra, R. Mondal, and B. Chanda. Learning a patch quality comparator for single image dehazing. *IEEE Transactions on Image Processing*, 27(9):4598–4607, Sep. 2018.

[42] G. Sharma, W. Wu, and E. N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30.

[43] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, Feb 2006.

[44] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[45] C. Wang, C. Xu, C. Wang, and D. Tao. Perceptual adversarial networks for image-to-image transformation. *IEEE Transactions on Image Processing*, 27(8):4066–4079, Aug 2018.

[46] Z. Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002.

[47] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[48] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, Nov 2003.

[49] W. Xue, L. Zhang, X. Mou, and A. C. Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):684–695, Feb 2014.

[50] D. Yang and J. Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, editors, *Computer Vision – ECCV 2018*, pages 729–746, Cham, 2018. Springer International Publishing.

[51] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, 37(6):1348–1357, June 2018.

[52] J. Yu, C. Xiao, and D. Li. Physics-based fast single image fog removal. In *IEEE 10th International Conference on Signal Processing Proceedings*, pages 1048–1052, Oct 2010.

[53] Y. Yuan, S. Liu, J. Zhang, Y. Zhang, C. Dong, and L. Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 814–81409, June 2018.

[54] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[55] H. Zhang and V. M. Patel. Density-aware single image deraining using a multi-stream dense network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[56] H. Zhang, V. Sindagi, and V. M. Patel. Image de-raining using a conditional generative adversarial network. *CoRR*, abs/1701.05957, 2017.

[57] H. Zhang, V. Sindagi, and V. M. Patel. Multi-scale single image dehazing using perceptual pyramid deep network. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1015–101509, June 2018.

[58] L. Zhang, L. Zhang, X. Mou, and D. Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, Aug 2011.

[59] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, June 2018.

[60] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[61] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, June 2018.

[62] H. Zhao, O. Gallo, I. Frosio, and J. Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, March 2017.

[63] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial net-

works. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

[64] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, Nov 2015.