# DHSGAN: An End to End Dehazing Network for Fog and Smoke

Ramavtar Malav[1], Ayoung Kim[2][0000-0001-9829-2408], Soumya Ranjan Sahoo[1][0000-0002-6115-9889], and Gaurav Pandey[3][0000-0002-4838-802X]

[1] Indian Institute of Technology Kanpur, Kanpur, India
[2] Korea Advanced Institute of Science and Technology, Daejeon, South Korea
[3] Ford Motor Company, Palo Alto, CA, USA

**Abstract.** In this paper we propose a novel end-to-end convolution dehazing architecture, called De-Haze and Smoke GAN (DHSGAN). The model is trained under a generative adversarial network framework to effectively learn the underlying distribution of clean images for the generation of realistic haze-free images. We train the model on a dataset that is synthesized to include image degradation scenarios from varied conditions of fog, haze, and smoke in both indoor and outdoor settings. Experimental results on both synthetic and natural degraded images demonstrate that our method shows significant robustness over different haze conditions in comparison to the state-of-the-art methods. A group of studies are conducted to evaluate the effectiveness of each module of the proposed method.

**Keywords:** Dehazing · Desmoking · GAN · ConvLSTM.

## 1 Introduction

Reduced vision due to haze, smoke, fog, and under-water is a serious problem in many computer vision applications such as object detection, classification, identification, visual navigation, etc. A typical camera works on the principle of pin-hole model and the quality of the image obtained depends upon the amount of light reaching the image sensors through the lens. However, in a turbid medium (due to haze, fog, smoke, etc.) the image contrast and color fidelity reduces due to scattering of light by medium particles. This, in turn, reduces the amount of information content in the resulting hazy image, which cannot be used for any processing until they are dehazed.

The Atmospheric Scattering Model [22], given below, is widely used [3, 5, 10, 18, 27, 31, 33, 36] in computer vision to formulate such degraded/hazy images in dense particle conditions such as fog, smoke, and underwater.

$$I_h(u,v) = \underbrace{I_r(u,v)t(u,v)}_{\text{direct attenuation}} + \underbrace{A(1-t(u,v))}_{\text{airlight}}, \tag{1}$$

where $I_h(u,v)$ is hazed pixel intensity at location $(u,v)$ in the query image, $I_r(u,v)$ is the real intensity to be estimated, $t(u,v)$ is the medium transmission and $A$ is global atmospheric light. The transmission value is expressed as

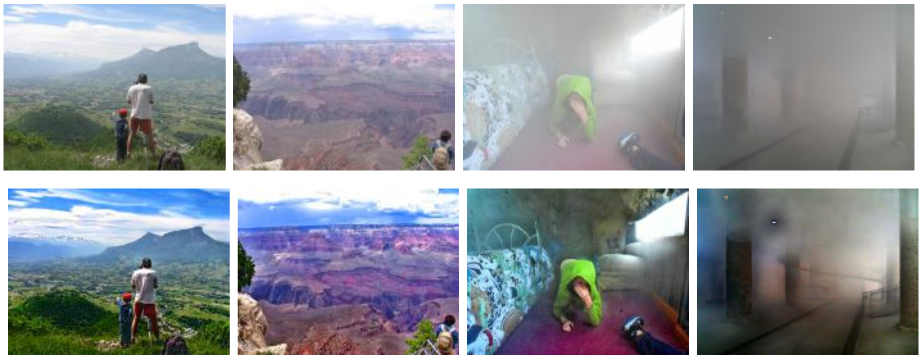


**Fig. 1.** Image enhancement by `DHSGAN` on real fog and smoke images. Here, top row shows the real world hazy image and bottom row display the enhancement by `DHSGAN`.

$t(u,v) = e^{-\beta d(u,v)}$, where $\beta$ is attenuation coefficient of the atmosphere and $d(u,v)$ is depth of scene point from the camera. The Eq. (1) indicates that after estimating the transmission $t(u,v)$ and atmospheric light $A$, the image can be recovered with:

$$\hat{I}_r(u,v) = \frac{I_h(u,v) - \hat{A}(1-\hat{t}(u,v))}{\hat{t}(u,v)}. \tag{2}$$

Most of the previous works estimate the transmission map and atmospheric light separately to recover the dehazed image. The transmission map estimation is successfully explored using both empirical prior based technique such as dark-channel prior [10], color attenuation prior [36], haze line prior [1,2] and learning based methods [3,18,27,33]. In comparison, atmospheric light is mainly calculated using empirical rules [2,10,31] with the exception of recent methods [18, 33] that use deep learning techniques. The most popular method of atmospheric light estimation [10] is based on the observation that the intensity of hazed pixel at very high depth should be equal to atmospheric light $A$. Therefore, the brightest pixel from the top deep pixels is selected as atmospheric light. It works very well for an image with high depth pixels or high atmospheric light (as maximum intensity pixel is selected), but its performance decreases as atmospheric light decreases, especially, for indoor images with low depth. Recently in [1,2], an atmospheric light estimation algorithm was formulated based on the observation that hazy image can be modelled by haze lines in RGB space which converge at the atmospheric light.

We further notice that although the dehazing is collectively referred to as the problem of image degradation due to fog, smoke, haze, underwater, etc., the recent learning methods [3,18,27,33] include training samples with only high atmospheric light. However, problems of reduced vision due to smoke in situations like fire-rescue emphasize the need of a general algorithm to address the full range of atmospheric light for including both smoke and bad weather conditions such as fog, haze, and mist.

In this paper, we propose an end-to-end deep learning framework for dehazing of images, we call it **DHSGAN**, which models the dehazed image as:

$$\hat{I}_r = \mathcal{G}(\hat{t}, I_h) = \mathcal{G}(\mathcal{T}(I_h), I_h), \tag{3}$$

where function $\mathcal{G}$ is approximated by a fully end-to-end convolution network, which also implements $\mathcal{T}$ (transmission map) as its inner module. This helps us in addressing the common limitation of most dehazing methods - the scene degradation where atmospheric scattering model (1) becomes invalid. Further, the function $\mathcal{G}$ is learned under a generative adversarial network (GAN) framework. The notion of GAN was first proposed by Goodfellow et al. [7] to learn underlying distribution of training data for generation of realistic images. The overwhelming success of GAN in image generation [7, 14] encouraged researchers to employ GAN for solving other low-level vision tasks such as style transfer [35], image enhancement [33, 34], and super-resolution [17]. Here we use GAN to generate a clear image from a hazy input and a transmission map which is again estimated by a CNN module within the framework (Fig 2(a)). The key contributions of the proposed work are:

- A novel end to end dehazing network is proposed which does not use the inverse atmospheric model (2) or any post processing step, rather the clean image is learned and directly generated by the final layer of a fully convolutional network. In addition, the model is trained under the generative adversarial network framework to synthesize realistic clean images. Furthermore, we employ a convolutional recurrent sub-architecture to take advantage of temporal redundancy in case of a video sequence.
- The network is learned on a large number of training samples with high variance to approximate a model that shows robustness on diverse conditions of scene degradation caused by smoke, fog, and haze in both indoor and outdoor settings.
- A group of studies are performed to demonstrate the importance of each sub-module of the purposed method. Further, extensive experiments are conducted on both synthetic and real dataset and comparisons are made with recent state of the art methods, showing the robustness of our model.

### 1.1 Related Work

A variety of approaches have been proposed in the literature to overcome the degradation caused by haze concerning both single image dehazing and video or multiple frames based dehazing [1, 2, 4, 5, 10, 18–20, 31, 33, 36].

Tan [31] proposed a Markov Random Field based dehazing method, which relies on the observation that clear-day images have more contrast than images affected by bad weather. In [5], R. Fattal proposed an image dehazing method based on the independent component analysis (ICA). He et al. [10] presented a very simple but promising method based on a key observation that local patches of haze-free images have low-intensity pixels in at least one color channel and in the corresponding hazy image, the intensity of these pixels is mainly contribution of airlight. Zhu et al. [36] identified a new prior based on the statistical observation that brightness and saturation of pixels vary sharply with haze thickness and difference of these two is positively correlated with the scene depth. This statistic was termed as color attenuation prior (CAP). Similarly, Berman et al. [1, 2] used a non-local prior (haze-line) to recover the clean image.

Recently, convolution neural networks (CNN) have achieved exemplary results in many application of computer vision such as image classification [12, 30], object detection [25] including single image dehazing [3, 18, 27, 33]. In [3], a neural network model is trained to retrieve transmission map from the hazy image. This transmission map is subsequently used with empirical rule [10] based atmospheric light estimation to get the clear image. Most recently, Zhang et al. [33] successfully used a joint architecture for both transmission map and atmospheric light estimation. The [33], uses a math-operation module where the inverse atmospheric model (2) is embedded within the network and a joint discriminator-based generative adversarial network is used to refine the transmission map and dehazed image. This work is closely related to ours as they use a discriminator network to train the sub-networks that estimates the transmission map and atmospheric light. However, estimating the two parameters, transmission and atmospheric light, separately can accumulate errors and potentially amplify each other in the final step of recovery (Eq. (2)). Earlier, Li et al. [18] address this problem by manipulating atmospheric scattering model to linearly embed both transmission map and atmospheric light into one variable and trained a light-weight CNN for its estimation. In the proposed method we do not use the inverse atmospheric model and let the generation network learn this model for a wide range of atmospheric light conditions.

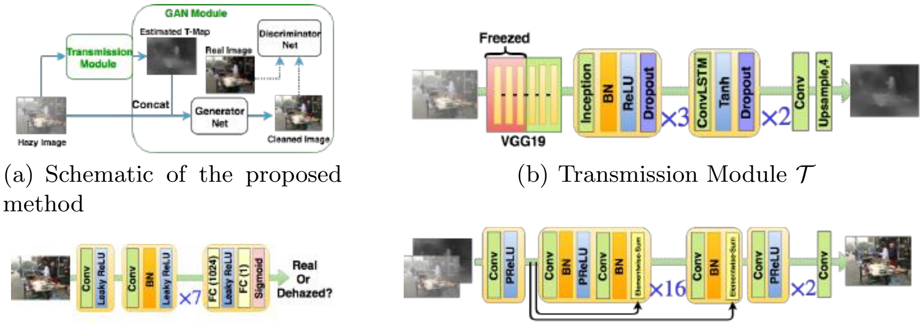
The rest of the paper is organised as follows. In section 2 we describe the proposed framework for dehazing of image sequences. In section 3 we present experimental results of the proposed technique and compare it with various state-of-the-art methods. In section 4 we present some concluding remarks and future works.

## 2 DHSGAN

In this section, the proposed DHSGAN is explained. Our model can be classified into two sub-modules (see figure 2(a)): T the Transmission Module and the GAN Module, both of which are explained in detail below.

### 2.1 The Transmission Module

The transmission module $\mathcal{T}$ is a fully convolutional recurrent architecture which takes a hazy image as input and estimates its transmission map. The network is initialized with `VGG19` [29] convolution layers pretrained on `ImageNet` [16] Dataset. The initial `VGG19` [29] convolution layers is proven to be rich feature extractors [26]. The initial 8 convolution layers of `VGG19` [29] are excluded from



(a) Schematic of the proposed method

(b) Transmission Module $\mathcal{T}$

(c) Discriminator Net $\mathcal{D}$

(d) Generator Net $\mathcal{G}$

**Fig. 2.** An overview of the proposed `DHSGAN`. The first module $\mathcal{T}$ estimates the transmission map from hazy image, which in turns concatenated with hazy image and passed to Generator Net $\mathcal{G}$. The generator net $\mathcal{G}$, trained under GAN framework, directly synthesize the realistic clean image without using inverse atmospheric model (2) or any post processing step.

the training. The `VGG19` [29] architecture is followed by three `Inception` [30] modules that simultaneously process the input feature map at multiple scales. We then use two `ConvLSTM` [24] layers capable of exploiting the temporal correlation of input frames in case of a video stream. The detailed architecture of transmission module is illustrated in figure 2(b).

### 2.2 The GAN Module

The GAN [7] module consists of two CNN architectures: Generative Model $\mathcal{G}$, and Discriminative Model $\mathcal{D}$. We follow the architecture guidelines proposed by Super-Resolution network SRGAN [17]. The transmission map estimated by the transmission module is concatenated with image and passed to $\mathcal{G}$. Our ultimate goal is to learn a generating function $\mathcal{G}$ which recovers the dehazed/clear images given a pair of transmission map and hazed image.

The core module $\mathcal{G}$ as illustrated in figure 2(d), contains 16 identical residual blocks [8, 12] with $3 \times 3$ kernels. Similar to[8, 17], we use `Batch-Normalization` [13] followed by `ParametricReLU` [11] activation function in each residual block. Each convolution layer in $\mathcal{G}$ is used with zero padding to get the dehazed image with same spatial dimension as input hazy image.

The discriminator model $\mathcal{D}$ is employed to discriminate between real image $I_r$ and dehazed images $\hat{I}_r$. Formally, the models $\mathcal{G}$ and $\mathcal{D}$ with respective learning parameters $\theta_G$ and $\theta_D$ are alternatively trained to solve the following adversarial min-max function [7]:

$$\begin{aligned}\min_{\theta_G}\max_{\theta_D} \quad & \mathbb{E}_{I\sim P_{\text{data}(I_r)}}[\log \mathcal{D}_{\theta_D}(I)] \\ & + \mathbb{E}_{I\sim P_{\text{data}(I_h)}}[\log(1-\mathcal{D}_{\theta_D}(\mathcal{G}_{\theta_G}(\mathcal{T}(I),I))].\end{aligned} \tag{4}$$

The game-theoretic problem allows one to learn a function $\mathcal{G}$ with the aim of fooling the network $\mathcal{D}$ that is trained to discriminate between real and generated dehazed image. This approach enables the generator $\mathcal{G}$ to synthesize realistic haze-free images by learning the distribution of pixel intensity from real data.

We use the same discriminator architecture purposed in [17]. The network consists of eight convolutional layers with `Batch-Normalization` [13] and `Leaky ReLU` activation as displayed in figure 2(c). At the end, two fully connected layers followed by a `Sigmoid` layer is used to classify the sample input as either real or dehazed image.

### 2.3 Loss Function

**Transmission Module.** The Mean Squared Error (MSE) has been commonly used to train transmission estimation model in previous works [3, 27]. We observe that if we assume that the atmospheric light $A$ is known and the transmission map estimated by the network is given by $t_e$ then the dehazed image $\hat{I}_r$ can be obtained from the atmospheric scattering model:

$$\hat{I}_r = \frac{I_h - A}{t_e} + A = I_r \frac{t_r}{t_e} + A\left(1 - \frac{t_r}{t_e}\right). \tag{5}$$

Therefore, the error in the pixel intensity values of the dehazed image is given by

$$error = \left|\hat{I}_r - I_r\right| = \left|(A - I_r)\left(1 - \frac{t_r}{t_e}\right)\right| \tag{6}$$

Based on this observation, we propose a **normalized MSE** loss function for learning the network parameter $\theta_T$ of transmission module $\mathcal{T}$ , which we define as:

$$\mathcal{L}^{N}_{\theta_T} = \frac{1}{WH} \sum_{u=1}^{W} \sum_{v=1}^{H} \left(1 - \frac{\mathcal{T}_{\theta_T}(I_h)_{u,v}}{t_r(u,v)}\right)^2 \tag{7}$$

Here, $\mathcal{T}_{\theta_T}(I_h)$ represent the output transmission of $\mathcal{T}$ module, with network parameter $\theta_T$, for a hazy image $I_h$ with $W \times H$ spatial dimension.

It ensures that the penalty for an actual and estimated transmission pair $(t_r, t_e)$ remains the same as for pair $(\alpha t_r, \alpha t_e)$, where $\alpha$ is any constant value. Since the `normalized MSE` can be unstable when $t_r$ is near zero, we clip the transmission value in training samples to 0.01.

**GAN Module.** The generative model $\mathcal{G}$ is trained on `MSE` loss and **VGG-loss** [17, 35] along with the adversarial loss [7] in stage-wise learning. The standard mean squared error (MSE) for hazed image $I_h$ and real image $I_r$ is defined as:

$$\mathcal{L}^{MSE}_{\theta_G} = \frac{1}{WH} \sum_{u=1}^{W} \sum_{v=1}^{H} \left(I_r(u,v) - \mathcal{G}_{\theta_G}(\mathcal{T}(I_h), I_h)_{u,v}\right)^2. \tag{8}$$

The `VGG` Loss function is formulated as euclidean distance between feature map extracted from intermediate layers of `VGG` network. `VGG` loss function allows perceptually more convincing image generation and successfully used in super-resolution [17], style transfer [35], and image enhancement [33, 34]. It is defined as:

$$\mathcal{L}_{\theta_G}^{VGG_{i,j}} = \frac{1}{2WH} \sum_{u=1}^{W} \sum_{v=1}^{H} \Big( \mathcal{V}_{i,j}(I_r)_{u,v} - \mathcal{V}_{i,j}(\mathcal{G}_{\theta_G}(\mathcal{T}(I_h), I_h))_{u,v} \Big)^2, \tag{9}$$

Where $\mathcal{V}_{i,j}$ represent the feature map extracted after $j^{th}$ convolution activation and before $i^{th}$ max-pooling of `VGG` network. In our experiments, we used $VGG19_{54}$ for training $\mathcal{G}$. The $\mathcal{L}_{\theta_G}^{VGG_{i,j}}$ loss is multiplied with a rescaling factor of 0.0061 to make it comparable with $\mathcal{L}_{\theta_G}^{MSE}$. In addition to the losses described above, the generator $\mathcal{G}$ is also optimized by the following adversarial loss [7] over training samples:

$$\mathcal{L}_{\theta_G}^{Gadv} = -\log \mathcal{D}(G_{\theta_G}(\mathcal{T}(I_h), I_h)). \tag{10}$$

Here for an input image, $\mathcal{D}(.)$ represent the classification probability, calculated by discriminator $\mathcal{D}$, of the image being from real image distribution. Similarly, the adversarial loss for training Discriminator $D$ is defined as:

$$\mathcal{L}_{\theta_D}^{Dadv} = -\left(\log \mathcal{D}_{\theta_D}(I_r) + \log(1 - \mathcal{D}_{\theta_D}(\mathcal{G}(\mathcal{T}(I_h), I_h))\right). \tag{11}$$

### 2.4 Dataset

We use image-depth pairs from publicly available datasets `NYU-v2` [28] (indoor), `SceneNet` [23] (indoor), `RESIDE` [19] (indoor-outdoor) and `KITTI` [6] (outdoor) to synthesize training samples {Hazy/Clean/Transmission Map} based on (1). We generated two training sets: **TrainA** and **TrainSeq**. `TrainA` contains approximately 100,000 training samples synthesized, using (1), from the above four datasets by randomly sampling $\beta$ from [0.6, 1.8] and $A$ from [0.2, 1.0]. We used a wide range of atmospheric light $A$ for training dataset as compared to previous methods. `TrainSeq` is synthesized to fine-tune recurrent transmission module $\mathcal{T}$. It is generated from the sequential datasets `SceneNet` [23] and `KITTI` [6] and contains 3403 image sequences each containing 4 images. Our training set contains a much larger number of samples and more challenging cases due to higher variance in the atmospheric light component as opposed to previous methods.

For testing, we use `RESIDE-SOTS` [19] indoor and outdoor test sets. As `RESIDE` [19] includes images with atmospheric light between [0.7, 1.0] only, we also generate the test set with atmospheric light between [0.25, 0.7] using provided image-depth pairs in the `RESIDE-SOTS`. `RESIDE-SOTS` indoor test set contain images generated using last 50 images from `NYU-v2`. We include additional 49 images from `NYU-v2` (which are not used in training) for test data generation with $A \in [0.25, 0.7]$. We denote the overall dataset as **TestAL**, which includes `RESIDE-SOTS` ($A \in [0.7, 1.0]$) test set along with self generated test set for lower $A \in [0.25, 0.7]$. To further demonstrate the generalization capability of our method we use `ICL-NUIM` [9] room and office sequences to generate 10 test sequences, denoted as **TestSeq**, having $A$ from 0.25 to 1.0 with an interval of

0.08. Each of the 10 sequences in TestSeq contains 99 images with same $A$ and attenuation coefficient $\beta$. For both test sets, the attenuation coefficient $\beta$ is randomly sampled from [0.6, 1.8].

### 2.5 Learning

The transmission module is trained separately from the GAN module. Initially, the transmission network $\mathcal{T}$ is optimized using TrainA dataset for 20 epochs with $10^{-5}$ learning rate, followed by fine-tuning using the TrainSeq dataset until it converges. We optimize this network $\mathcal{T}$ using the Adam [15] solver with $\beta_1 = 0.9, \beta_2 = 0.999$ as hyper-parameters.

Following the guidelines summarized in [17], we train the GAN module using TrainA dataset in three stages. At first step, only the generator $\mathcal{G}$ is trained using only $\mathcal{L}_{\theta_G}^{MSE}$ (8) with a learning rate of $10^{-4}$. In the second step, both $\mathcal{G}$ and $\mathcal{D}$ are alternatively trained to jointly minimize overall generator loss $\mathcal{L}_{\theta_G}^{G}$ (12) and discriminator loss $\mathcal{L}_{\theta_D}^{Dadv}$ (11) to solve the adversarial min-max problem (4). The overall generator loss function $\mathcal{L}_{\theta_D}^{Dadv}$ is calculated as:

$$\mathcal{L}_{\theta_G}^{G} = \mathcal{L}_{\theta_G}^{VGG_{5,4}} + 10^{-3}\mathcal{L}_{\theta_G}^{Gadv}. \tag{12}$$

For the first two steps of training, we use the ground truth transmission map along with the corresponding synthesized hazy image and real image samples from the TrainA dataset. In the final step, we replace the ground-truth transmission with the output of the trained transmission network $\mathcal{T}$ to further fine-tune $\mathcal{G}$ and $\mathcal{D}$ with a reduced learning rate of $10^{-5}$. Similar to the transmission module, the GAN framework is also optimized using Adam [15] with same hyper-parameters.

**Table 1.** Quantitative comparison using PSNR/SSIM between haze-free images and generated dehazed images on TestAL test set. DCPDN [33] performance on indoor dataset is not computed because the indoor partition of TestAL includes images from NYU-v2 [28] that overlap with the DCPDN [33] training set, this is primarily because DCPDN is trained on randomly selected images from NYU-v2 [28]. Below, we also include performance of two sub-versions of DHSGAN: InvDHSGAN, and DHSGANv0.5. For fair comparison with [33], DHSGANv0.5 is trained on images with atmospheric light between [0.5, 1.0]. To conduct further ablation study, similarly to [33], we add inverse atmospheric model (2) layer after DHSGAN, keeping the existing architecture untouched. This configuration is termed as InvDHSGAN and follows the same training procedure as DHSGAN with same training data.

| | NLD [1, 2] | DehazeNet [3] | AOD-Net [18] | DCPDN [33] | InvDHSGAN | DHSGANv0.5 | DHSGAN |
|---|---|---|---|---|---|---|---|
| Indoor $A \in [0.25, 0.5]$ | 17.4539 / 0.6960 | 15.7652 / 0.6362 | 14.2226 / 0.5312 | NA | 17.9265 / 0.7660 | 17.8997 / 0.8109 | **19.6296** / **0.8356** |
| Indoor $A \in [0.5, 0.7]$ | 18.0956 / 0.7211 | 16.6766 / 0.7101 | 15.4835 / 0.6587 | NA | 18.3253 / 0.7785 | 20.4201 / 0.8523 | **20.9710** / **0.8545** |
| Reside-In [19] $A \in [0.7, 1.0]$ | 18.9228 / 0.7489 | 22.9989 / **0.8756** | 21.1155 / 0.8504 | NA | 21.3629 / 0.7666 | **23.3469**/ 0.8388 | 21.7236 / 0.8120 |
| Outdoor $A \in [0.25, 0.5]$ | 19.9369 / 0.8140 | 19.6741 / 0.8029 | 15.8383 / 0.6275 | **23.1771** / 0.8106 | 21.9664 / 0.8707 | 22.6735 / **0.8781** | 22.7472 / 0.8772 |
| Outdoor $A \in [0.5, 0.7]$ | 21.0152 / 0.8313 | 21.1485 / 0.8618 | 17.4319 / 0.7885 | 24.2113 / 0.8601 | 22.5285 / 0.8488 | **24.4375** / **0.8821** | 23.6601 / 0.8722 |
| Reside-Out [19]$A \in [0.8, 1.0]$ | 19.9338 / 0.8016 | 27.2963 / 0.8886 | 24.3560 / 0.9080 | 22.6375 / 0.8447 | 25.9092 / 0.8868 | **27.7660** / 0.9009 | 26.6127 / **0.9105** |

Images Count ⇒ Indoor: 4 × 99, Outdoor: 2 × 492, RESIDE-In: 500, and RESIDE-Out: 500

**Table 2.** Quantitative Comparison using SSIM on TestSeq Indoor Dataset. Here, we also display a subversion of DHSGAN , referred as DHSGANv0, which forgets the state of ConvLSTM [24] after each video frames. It shows the performance improvement by using temporal correlation with DHSGAN.

| $A$ | CAP [36] | NLD [1, 2] | DehazeNet [3] | AOD-Net [18] | DCPDN [33] | DHSGANv0 | DHSGAN |
|---|---|---|---|---|---|---|---|
| Transmission | 0.8419 | **0.8787** | 0.8603 | NA | 0.8587 | 0.8473 | 0.8512 |
| Image | 0.7269 | 0.6467 | 0.7851 | 0.7436 | 0.8304 | 0.8476 | **0.8501** |

## 3 Experimental Results

In this section, we discuss the experiments performed with our model and compare them with the existing state-of-the-art methods to evaluate the robustness of the proposed method.

### 3.1 Quantitative Evaluation on Synthetic Dataset

In this section, we quantitatively compare the performance of our method with five recent state of the art methods: CAP [36], NLD [1,2], DehazeNet [3], AOD-Net [18], and DCPDN [33]. The availability of ground truth images in synthetic dataset enables us to evaluate our method using Peak Signal-to-Noise Ratio (PSNR) and another well-known metric Structural Similarity Index (SSIM) [32], which is proven to be coherent with human perception.

In table 1, we compare the mean SSIM and mean PSNR between haze-free images and dehazed images estimated from various methods on the synthesized TestAL dataset. The algorithms CAP [36] and Dehazenet [3], which use the algorithm proposed in DCP [10] for atmospheric light $A$ estimation, perform better for the higher $A$ in comparison to low $A$, especially, for indoor images. The AOD-Net [18] also follows the same trend as it was trained on samples generated using NYU-v2 [28] images keeping $A$ between $[0.6, 1]$ only. The algorithm NLD $[2, 1]$ which uses haze-line prior to estimate $A$ seems unaffected by different $A$ range but displays low accuracy in both indoor and outdoor dataset. The most recent work DCPDN [33] is trained on 4000 images having $A \in [0.5, 1]$ which was generated using 1000 randomly selected images from NYU-v2 [28]. This makes DCPDN [33] training set overlapping with TestAL indoor test set, which consists of images generated from last 99 images from NYU-v2 [28]. However, in case of outdoor dataset DCPDN [33] shows good generalization capability for all $A$ as it is also trained on a relatively wide range of $A \in [0.5, 1]$ . It should be noted that DCPDN [33] recovers the clean image using the inverse atmospheric model (2) embedded within the neural network architecture. DCPDN [33] estimates the transmission map and atmospheric light, separately and then uses (2) to generate the dehazed output. In contrast the proposed DHSGAN generates the output image directly from the generator network, which means it learns the inverse atmospheric model and therefore it is not dependent on separate estimation of atmospheric light. We observe that because our method is purely based on the generation network and is trained on a wide range of atmospheric light it works

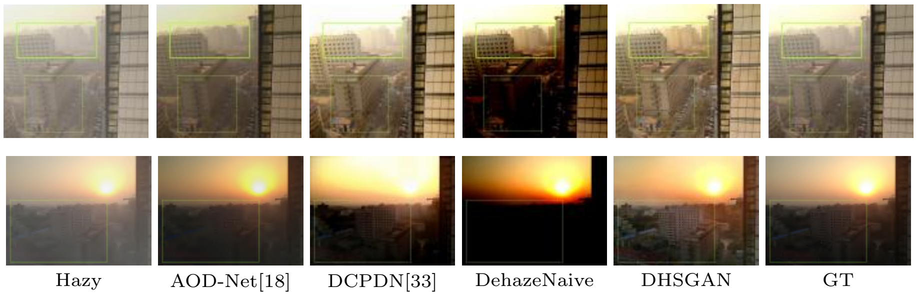


**Fig. 3.** Qualitative comparison with recent dehazing methods on synthetic dataset TestAL. We notice that our model is generating images which are cleaner than provided ground truth. It shows that quantitative comparison with the ground truth doesn't fully explain the capacity of DHSGAN. Further, to demonstrate that generator function $\mathcal{G}$ is not just approximating inverse atmospheric scattering model (2), in column DehazeNaive we display the recovered image with (2) by using $\mathcal{T}$ output and known $A$.

well in both indoor and outdoor partition of TestAL. As an ablation study, similarly to [33], we add inverse atmospheric model (2) layer after DHSGAN, which uses $\mathcal{T}$ module output and generator $\mathcal{G}$ output (now behaving as $A$ estimation module) in (2) to restore the dehazed image. This modification is termed as InvDHSGAN. This configuration limits the generator $\mathcal{G}$ capability to just atmospheric light estimation. It is evident from the results shown in table 1 that removing the inverse atmospheric model (2) layer makes model more accurate and generalized. Further, to show generalization capability of our method, in table 2 we display the performance of our method on TestSeq dataset that is generated using video sequence from ICL-NUIM [9] dataset. In this experiment, we deploy a sub-version of our model, termed as DHSGANv0, in which we consider image sequence as independent images, i.e. we forget the state of ConvLSTM in $\mathcal{T}$ module after each image. We observe that temporal correlations give a some boost to accuracy as observed by difference between DHSGANv0 and DHSGAN SSIM accuracy.

### 3.2 Qualitative Evaluation on Real and Synthetic Datasets

In this section, we present a visual comparison of our network with other methods by dehazing hazed images. Figure 3 shows the comparison of the proposed method with the state-of-the-art techniques on synthetic set TestAL, the output of the proposed method looks visually better or at par with the existing methods. In some cases the proposed method recovers the image details so well that it looks better than the ground truth image. This shows that our method is actually learning the clean image colour patterns also as opposed to just learning the inverse Atmospheric model (2). In Figure 4 we show the results on a standard foggy image dataset, these images contain challenging fog density from outdoor. To further show the robustness of our model on different haze conditions, we also collected indoor smoke images containing high smoke density as displayed

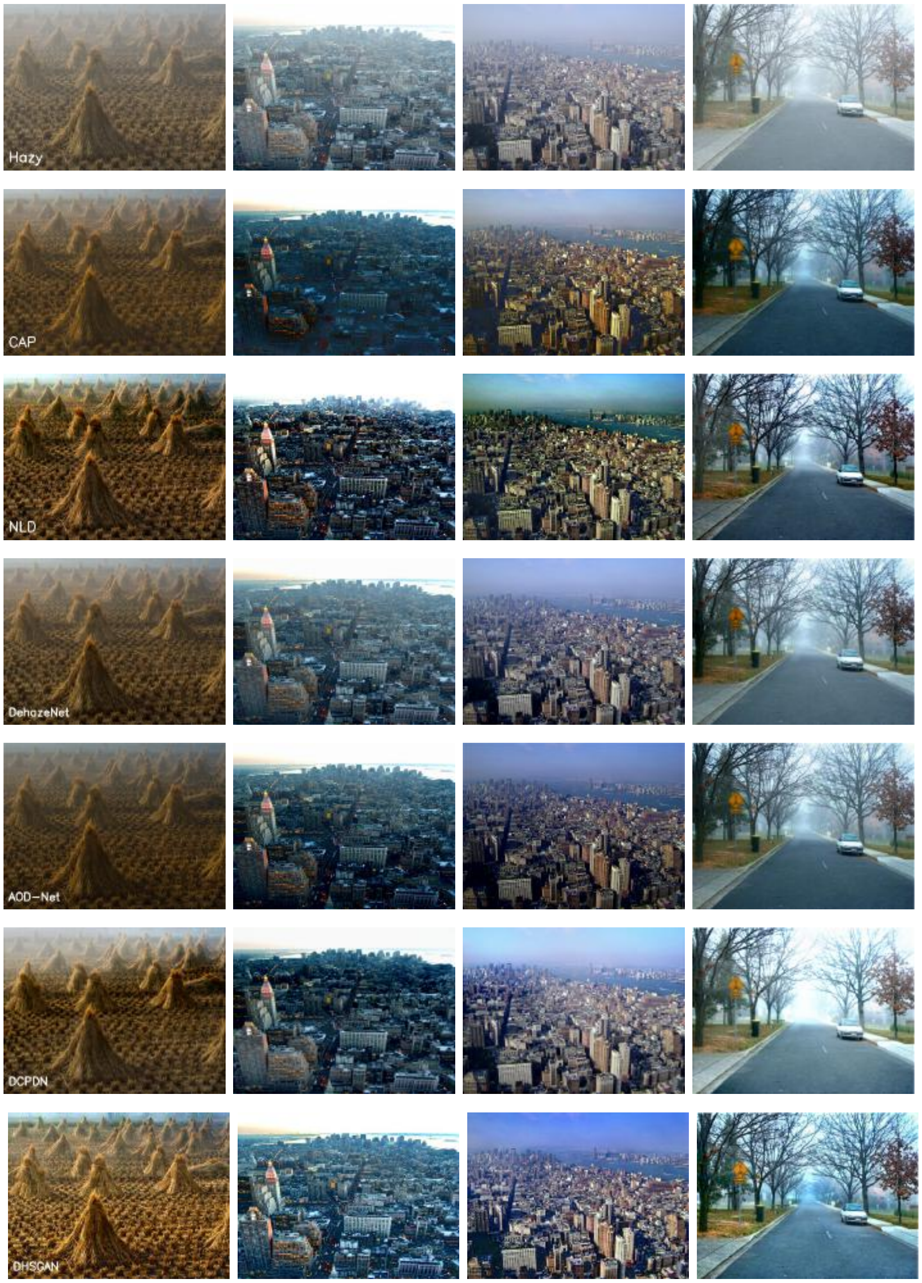


**Fig. 4.** Qualitative comparison of `DHSGAN` with recent state-of-the-art methods on real fog images released by previous authors. From top to bottom: the hazy image, CAP[36], NLD[1, 2], DehazeNet[3], AOD-Net[18], DCPDN [33], and DHSGAN.

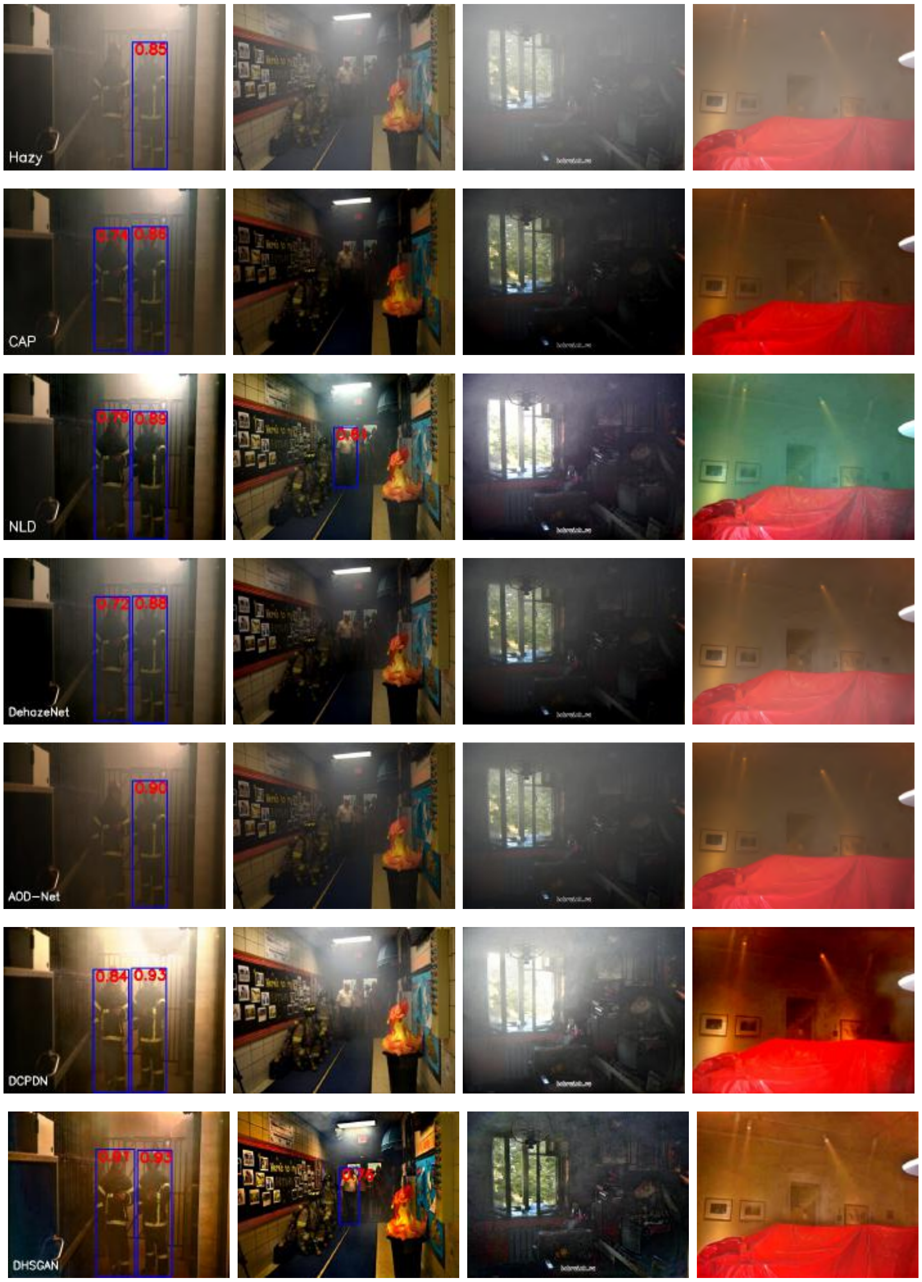
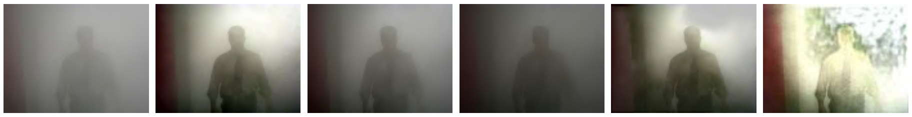
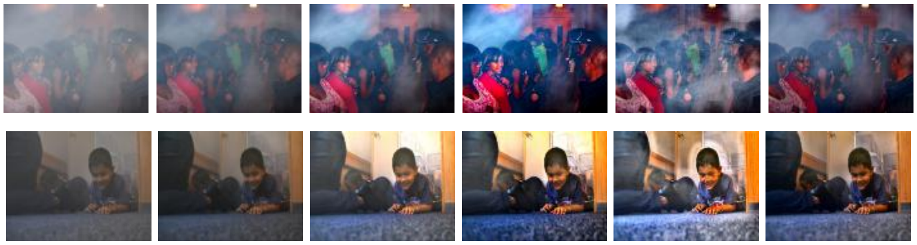


**Fig. 5.** Qualitative comparison of DHSGAN with recent methods on real smoke images. From top to bottom: the hazy image, CAP[36], NLD[1, 2], DehazeNet[3], AOD-Net[18], DCPDN [33], and DHSGAN. Further to show that DHSGAN is able to preserve the information as opposed to just visual improvement, we employ SSD [21] as person detector. The confidence of object detection model is printed over the detected bounding box.



Hazy NLD[1, 2] DehazeNet[3] AOD-Net[18] DCPDN[33] DHSGAN

**Fig. 6.** Qualitative comparison of `DHSGAN` with recent methods on extreme haze condition.



(a) Hazy (b) $\beta = 1.0$ (c) $\beta = 2.0$ (d) $\beta = 3.0$ (e) $\mathcal{T}$-$\mathcal{G}$ MSE (f) DHSGAN

**Fig. 7.** Effect of different transmission on generator $\mathcal{G}$. In column (b), (c), and (d) we generate scene depth using CAP [36] and use different attenuation coefficient $\beta$ to generate transmission maps. It can be seen that low transmission (high $\beta$) tells generator $\mathcal{G}$ that haze density is high and thus more color shifting occurs. It implies that $\beta$ can be configured to exploit trade-off between information retrieval and realness. In last two columns, (e) and (f), we show the performance of generator $\mathcal{G}$ with transmission estimated by $\mathcal{T}$ module. In column (e), the generator $\mathcal{G}$ is trained outside GAN framework using only MSE-Loss (8).

in figure 5. The smoke in case of fire incidents present challenges to other vision tasks and it's not frequently investigated in recent year as an independent problem.

Figure 4 shows that we are able to clean fog from images without losing any information. As our method is purely network generation based, we are able to provide similar enhancement to the pixels at depth as opposed to other methods. Our method learns what should be the colour compositions of the road, buildings, sky, trees etc., that makes our method resilient to colour shifting and blurring in case of normal fog conditions. In figure 5, we present the robustness of our method on a more challenging image degradation due to smoke. It can be seen that we are able to recover richer image content than other dehazing methods, which also holds true in extreme situations (figure 6). Although, in case of high haze density the information is preserved, some artificial colour appears in the output image (figures 5 and 6).

In figure 7, we investigate the effect of transmission map on the GAN Module of our model. For this experiment, we estimate the depth of real images using algorithm proposed in [36] and use several attenuation coefficient $\beta$ value to generate different transmission maps along with the one by $\mathcal{T}$ module. It can be seen that with increasing $\beta$ the artificial colour becomes dominant but images get cleaner. Further in Figure 7, we also display the result generated by $\mathcal{G}$ network

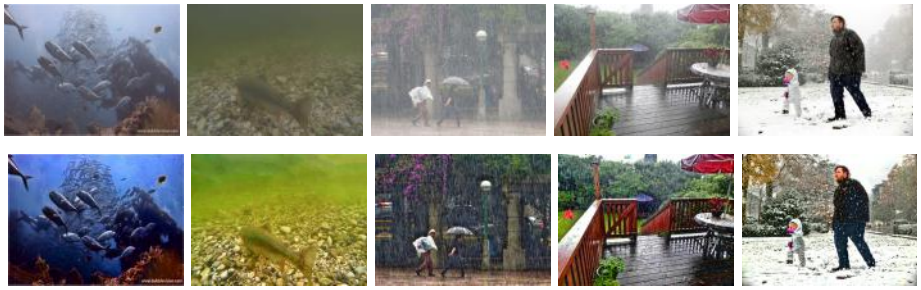


**Fig. 8.** Performance of DHSGAN on other reduced vision conditions: Underwater [27], Rain [34], and Snow-fall [34].

which is trained on only MSE loss (8) to demonstrate the realness introduced by VGG Loss (9) and GAN [7] framework in our method. Finally, in figure 8 we present the performance of our method on other haze conditions such as underwater, raining and snow. It shows that, in future, a single model can be developed to tackle a wide range of image degrading scenarios.

## 4 Conclusion

In this paper, we have presented a novel end-to-end convolutional network that can directly generate realistic dehazed images in a variety of haze conditions including fog, rain, underwater and smoke. We train our network under GAN framework with a wide range of atmospheric light conditions, which enables the network to learn the distribution of real haze-free images. We do not use the widely used inverse atmospheric model for recovering haze-free images and instead let the generator network learn the dehazing function directly from the training data resulting into a more robust solution. We performed several experiments to demonstrate the superior performance of our method as compared to the state-of-the-art on commonly used real and synthetic datasets. We also showed that the proposed method can be generalized to different types of hazy images including the more challenging smoky images from indoor environments. We show that the proposed method works well even for smoky images with low atmospheric light thereby proving the robustness and the generalization capabilities of the network.

## References

1. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1674–1682 (2016)
2. Berman, D., Treibitz, T., Avidan, S.: Air-light estimation using haze-lines. In: 2017 IEEE International Conference on Computational Photography. pp. 1–9 (2017)

3. 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)
4. Chen, C., Do, M.N., Wang, J.: Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In: 14th European Conference on Computer Vision. pp. 576–591. Springer (2016)
5. Fattal, R.: Single image dehazing. ACM transactions on graphics (TOG) **27**(3), 72 (2008)
6. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361 (2012)
7. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems 27. pp. 2672–2680 (2014)
8. Gross, S., Wilber, M.: Training and investigating residual nets. Facebook AI Research, CA.[Online]. http://torch. ch/blog/2016/02/04/resnets. html (2016)
9. Handa, A., Whelan, T., McDonald, J., Davison, A.J.: A benchmark for rgb-d visual odometry, 3d reconstruction and slam. In: Robotics and Automation (ICRA), 2014 IEEE International Conference on. pp. 1524–1531 (2014)
10. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(12), 2341–2353 (2011)
11. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. pp. 1026–1034 (2015)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
13. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. international conference on machine learning pp. 448–456 (2015)
14. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. international conference on learning representations (2018)
15. Kingma, D.P., Ba, J.L.: Adam: A method for stochastic optimization. international conference on learning representations (2015)
16. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
17. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A.P., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 105–114 (2017)
18. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 4780–4788 (2017)
19. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Reside: A benchmark for single image dehazing. arXiv preprint arXiv:1712.04143 (2017)
20. Li, Y., You, S., Brown, M.S., Tan, R.T.: Haze visibility enhancement: A survey and quantitative benchmarking. Computer Vision and Image Understanding **165**, 1–16 (2017)

21. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. european conference on computer vision pp. 21–37 (2016)
22. McCartney, E.J., Hall, F.F.: Optics of the atmosphere: Scattering by molecules and particles. Physics Today **30**(5), 76–77 (1977)
23. McCormac, J., Handa, A., Leutenegger, S., Davison, A.J.: Scenenet rgb-d: Can 5m synthetic images beat generic imagenet pre-training on indoor segmentation? In: 2017 IEEE International Conference on Computer Vision. pp. 2697–2706 (2017)
24. Patraucean, V., Handa, A., Cipolla, R.: Spatio-temporal video autoencoder with differentiable memory. arXiv preprint arXiv:1511.06309 (2015)
25. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster r-cnn: towards real-time object detection with region proposal networks. In: NIPS'15 Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. vol. 2015, pp. 91–99 (2015)
26. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 806–813 (2014)
27. Shin, Y.S., Cho, Y., Pandey, G., Kim, A.: Estimation of ambient light and transmission map with common convolutional architecture. In: OCEANS 2016 MTS/IEEE Monterey. pp. 1–7 (Sept 2016). https://doi.org/10.1109/OCEANS.2016.7761342
28. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: ECCV'12 Proceedings of the 12th European conference on Computer Vision - Volume Part V. pp. 746–760 (2012)
29. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. international conference on learning representations (2015)
30. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9 (2015)
31. Tan, R.T.: Visibility in bad weather from a single image. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8 (June 2008)
32. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
33. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. computer vision and pattern recognition pp. 3194–3203 (2018)
34. Zhang, H., Sindagi, V., Patel, V.M.: Image de-raining using a conditional generative adversarial network. arXiv preprint arXiv:1701.05957 (2017)
35. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2242–2251 (2017)
36. 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)