J. Vis. Commun. Image R. 74 (2021) 103014

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Single image dehazing using improved cycleGAN☆

B.S.N.V. Chaitanya [a], Snehasis Mukherjee [b,*]

[a] Indian Institute of Information Technology, Sri City, India
[b] Shiv Nadar University, Greater Noida, India

ARTICLE INFO

Keywords:
CycleGAN
Cyclic consistency loss
AOD-NET
Single image dehazing
SSIM loss

ABSTRACT

Haze is an aggregation of very fine, widely dispersed, solid and/or liquid particles suspended in the atmosphere. In this paper, we propose an end-to-end network for single image dehazing, which enhances the CycleGAN model by introducing a transformer architecture within the generator, which is specific for haze removal. The proposed model is trained in an unpaired fashion with clear and hazy images altogether and does not require pairs of hazy and corresponding ground-truth clear images. Furthermore, the proposed model does not depend on estimating the parameters of the atmospheric scattering model. Rather, it uses a K-estimation module as the generator's transformer for complete end-to-end modeling. The feature transformer introduced in the proposed generator model transforms the encoded features into desired feature space and then feeds them into the CycleGAN decoder to create a clear image. In the proposed model we further modified the cycle consistency loss to include the SSIM loss along with pixel-wise mean loss to produce a new loss function specific for the reconstruction task, which enhances the performance of the proposed model. The model performs well even on the high-resolution images provided in the NTIRE 2019 challenge dataset for single image dehazing. Further, we perform experiments on NYU-Depth and reside beta datasets. Results of our experiments show the efficacy of the proposed approach compared to the state-of-the-art in removing the haze from the input image.

1. Introduction

Haze is an atmospheric phenomenon where dust, smoke and other dry objects obscure the perspective of the sky and provide an atmosphere of opalescent appearance, subdued colors. In general, outdoor photography often suffers from haze due to bad weather conditions which results in low-quality images. Haze and fog dramatically diminish the visibility of indoor pictures as well, where contrast gets reduced and the original color of the objects get affected. Significant haze in the images or videos can affect the performance of other analysis tasks, such as object identification/tracking, event recognition and many more. Image dehazing especially when no reference image is available is gaining the interest of the computer vision and graphics researcher community as this can act as an efficient pre-processing step to improve the other tasks. A few sample hazy images and the corresponding haze-free images from the NYU depth dataset [1] and NTRIE 2019 dataset [2] are shown in Fig. 1.

The approaches for single image dehazing available in the literature can be categorized into two major classes: information-based and learning-based approaches. Traditional approaches for image dehazing are information-based, where several image features (priors) are used. A few example of such priors include color attenuation [3], dark channel prior (DCP) [4] and color line [5]. These priors are applied to recover the transmission map of the corresponding haze-free image and obtain the haze-free image after smoothening the obtained map. Most of the recent haze removal techniques are learning-based. The learning-based haze removal techniques either aim to obtain the haze-free transmission map for the input image by training a CNN architecture with image patches [6] or propose an end-to-end framework for obtaining the haze-free image directly from the input hazy image [7]. The CNN-based methods focus mainly on estimating the transmission map and/or atmospheric light to recover smooth images through the atmospheric dispersion model. However, CNN based techniques require a huge number of hazy images and the corresponding haze-free images to train, which is difficult to get and computationally costly. Recently Generative Adversarial Networks (GANs) are being used to dehaze images [8–10].

GANs are deep neural networks that generate synthetic data (images) based on certain input information (noise/image), where the images of the input and the out domains may be different. The GAN based methods for single image dehazing require a hazy input image and the ground truth haze-free image as input (as a paired information) [9,10]. However, using a cycleGAN network eliminates the

☆ This paper has been recommended for acceptance by Dr Zicheng Liu.

* Corresponding author.
E-mail addresses: viswachaitanya.b16@iiits.in (B.S.N.V. Chaitanya), snehasis.mukherjee@snu.edu.in (S. Mukherjee).

https://doi.org/10.1016/j.jvcir.2020.103014
Received 8 April 2020; Received in revised form 27 November 2020; Accepted 30 December 2020
Available online 1 January 2021

Fig. 1. Two sample hazy images (left side) and the corresponding haze-free images (right side) are shown. First row shows a sample image from NYU dataset and the second row shows a sample image from NTIRE 2019 challenge dataset.

need for paired information [8]. The cycleGAN network proposes an end-to-end model for haze removal based on cycleGAN's architecture, introducing perceptual loss into cycle consistency loss as the loss function [8]. The cycleGAN was the first attempt to provide a cycleGAN architecture for haze removal and shown significant improvements in performance compared to the state-of-the-art. Motivated by the success of cycleGAN, the proposed model aims to build an end-to-end dehazing network by introducing a lightweight CNN architecture into the generator in the cycleGAN. We use the AOD-Net architecture as the light-weight CNN for our generator models transformer [7] because of its simplicity to be easily integrated with the proposed GAN architecture and its state-of-the-art performance in image dehazing. In other words, the AOD-Net is used in the proposed approach for two reasons. First, the simple and shallow structure of the transformer architecture of AOD-Net makes it easily deployable as a generator architecture in a GAN network. Second, AOD-Net has shown its efficacy in single image dehazing, providing the state-of-the-art results. Since the similarity measure between the ground truth and restored images has a proven record of reducing the error of the proposed model drastically [11], we introduce SSIM loss as a loss function in our GAN architecture. We propose a new loss function considering SSIM loss along with the cycle consistency loss.

The paper has two major contributions.

- We propose a simple, shallow and efficient end-to-end cyclic GAN architecture for single image dehazing.
- We introduce a loss function specific to the image dehazing problem, to enhance the performance of the proposed model.
- We perform an extensive experiments on the potential loss functions, suitable for image dehazing. So that the proposed loss function helps to provide a dehazed image close to the ground truth but at the same time, emphasize the perceptual factor, rather than blindly believing on the ground truth.

Next we discuss the related works that have been done in single image dehazing.

2. Related works

Single image dehazing is gaining significant attention of researchers in computational imaging and vision [4,6,7,12]. Schechner et al. proposed an instant polarization based model for image dehazing based on the intuition that, the air-light scattered by the suspended particles are partly polarized [13]. Similar approach was made in [14], where the Markov random fields was applied to optimize the cost function efficiently through multiple methods like graph-cuts and belief propagation for single image dehazing. Fattal, in [5] developed a refined model of clear image formation that involves a single picture, which in relation to the transmission function accounts for the surface shading. Unlike [14], Fattal [5] tried to distinguish uncorrelated fields, namely the shading of objects and the attenuation of particles. Ancuti et al. [15] applied a fusion-based approach that requires two tailored versions of the hazy picture as inputs which are weighted by particular maps to deliver precise haze-free outcomes.

The problem of image dehazing is formulated based on the classical scattering model of the atmosphere:

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

where $I(x)$ is the hazy image and $J(x)$ is the radiance of the scene ('clear image') to be recovered. Here A refers to global atmospheric light, $t(x)$ is the transmission matrix and β is the scattering coefficient of the medium and d(x) denotes pixel-wise Depth Map.

$$t(x) = e^{-\beta d(x)}, \tag{2}$$

$$J(x) = \frac{1}{t(x)} I(x) - A\frac{1}{t(x)} + A. \tag{3}$$

Various methods e.g. [3,4,16–18] have been proposed for image dehazing using some additional information other than depth map of the image. Xie et al. [16] proposed an image dehazing algorithm using dark channel prior and Multi-Scale Retinex. In [3], a color attenuation was estimated before removing haze. In [3], the depth information of a hazy image is reconstructed using a linear model for generating the color attenuation prior and learning the parameters of this linear model by supervised model. Using this reconstructed depth information, transmission is estimated and scene radiance is recovered by means of classical scattering model of the atmosphere. In [4], the dark channel prior (DCP) is used for haze removal. DCP is one of the most effective image priors for haze removal, which is a type of outdoor Haze-free image statistics. The idea of DCP is based on an observation that there are pixels with very small intensities in at least one channel of color in most of the local patches of haze free outdoor images. Colores et al. [19] applied DCP and Radon transform for haze removal. However, in cases of indoor images and images containing sky region, the DCP based approaches fail [20].

Various visual priors other than DCP have been proposed to remove haze from images [5,21,22]. Fattal [5] proposed the color line model for haze removal. Recently non-local total variation (NLTV) regularization has been suggested in [21], concentrating on finding the excessive information and depth structure in images that reduce noises and artifacts after dehazing. Kumar et al. proposed a color correction mechanism to generate the clear (smoothed) transmission map of the image, from which, the haze free image is produced by atmospheric scattering model [22]. Borkar et al. [12] proposed a combination of the DCP and color line priors to obtain a haze-free transmission map. Recently, effort has been made to clear the artifacts observed in the images after haze removal, by Nearest Neighbor regularization on the haze-free transmission map of the image [20]. Further in [20], adaptive filters has been proposed to handle sky regions in the outdoor images. A similar effort was made in [23] by applying an adaptive filter to smoothen the atmospheric light at the sky region of the image.

After introduction of deep learning techniques, several methods have been proposed for haze removal using deep networks [6,17,18]. In [17], Liu et al. used both domain knowledge and haze distribution in the from of prior and data respectively to estimate scene radiance. In [18], Yang et al. proposed a Disentangled end-to-end Dehazing Network to generate realistic images without haze using only unpaired supervision and multi-scale adversarial training. In [6], Cai et al. suggested a trainable end-to-end scheme called DehazeNet for the

assessment of medium transmission which is designed based on the atmospheric scattering model.

Several attempts have been made to employ CNNs to recover $t(x)$ from $I(x)$ (Eq. (1)) [7,24–27]. Then $J(x)$ is estimated from (3) with A (global atmospheric light) estimated using some empirical methods. The CNN-based techniques focus on estimating either the transmission map or the atmospheric light from the given image, from which after applying a smoothening operation, clean images are recovered by atmospheric scattering model shown in (3). In [24] an effective CNN is proposed for image dehazing using three architectural variants to investigate the dependence of dehazed image quality on parameter count and model the design. In [28] a cardinal color fusion network is suggested to remove single image haze. The first two versions of [24] use a single encoder–decoder convolutional extractor and the third uses an encoder–decoder pair to estimate the atmospheric coefficient and transmission map. Every version of the above model finish with a network of pyramid image refinement to make the final dehazed image. A two stage network is used in [28] to dehaze image using color fusion network. The first phase fuses color data in hazy images and produces multi-channel depth maps, whereas in the second phase, scene transmission map is estimated from the previous dark channels using multi-channel multi-scale CNN to restore the scene. On the other hand, the Ranking-CNN proposes to increase the CNN structure so that attributes of hazy images in the statistics and structures can be captured simultaneously. In [25], a U-Net like encoder–decoder profound network was suggested via progressive feature fusions to directly learn extremely nonlinear transformation function from observed hazy picture to the ground-truth. In [26], the CNN is used as a quality benchmark, proposing a technique comparing different output patches with the initial hazy version and then selecting the best one to unravel a clear image patch. Recently, AOD-NET [7] (All-in-One Dehazing Network) proposed a full end-to-end dehazing CNN model based on reformulation (1), generating $J(x)$ straight from $I(x)$ without any other intermediate steps like transmission map or atmospheric light.

The architecture of AOD-Net consists of two components [7]: a K-estimation module with five convolutional layers all together to estimate $K(x)$ by $I(x)$ and a clear image generation module to estimate $J(x)$ by $K(x)$ and $I(x)$ using the following equations:

$$J(x) = K(x)I(x) - K(x), \tag{4}$$

$$K(x) = \frac{\frac{1}{t(x)}(I(x) - A) + A}{I(x) - 1}. \tag{5}$$

The end-to-end learning procedure makes AOD-NET simple and easy to fit in any other architectures. Densely Connected Pyramid Dehazing Network (DCPDN) [29] and Gated Context Aggregation Network [30] are few other techniques for image dehazing. In [30], image is dehazed using an end-to-end gated context aggregation network instead of using traditional low-level image priors as the restore constraints, whereas in [29], transmission map and atmospheric light are learnt collectively using an end-to-end network. In [31], multi-scale convolution networks (MSCNN) was proposed using prior-atmospheric illumination.

Although significant efforts have been made for dehazing, the validation of the dehazing processes remains a less-visited problem, due to the lack of sufficient number of real hazy pairs and the corresponding haze-free images. Several datasets have been created to resolve this issue, such as RESIDE (REalistic Single-Image DEhazing) [32], composed of both synthetic and real-world hazy pictures; I-HAZE [33], comprising of 35 picture pairs of hazy and matching hazy-free (ground-truth) indoor pictures; O -HAZE [34], comprising of 45 distinct outdoor scenes representing the same visual material collected in haze-free pictures; Dense-Haze [35], a dataset characterized by thick and homogeneous hazy scenes, including 33 pairs of true hazy and matching hazy-free pictures of different outdoor scenes; a dataset of 1449 RGBD pictures is implemented in [36], capturing 464 different indoor scenes with comprehensive annotations.

The Generative adversarial Networks (GANs) are deep architectures gaining attention of researchers of different areas related to machine learning [37] having potential to learn to imitate any data distribution. Efforts have been made to apply GANs for single image dehazing [38, 39]. Zhang et al. proposed a multi-task technique in [38], consisting of three modules which are map assessment through GANs, hazy feature extraction and picture dehazing. All modules were trained together using a loss function combining the perception loss and Euclidean loss of pixels. Enhanced Pix2pix Dehazing Network (EPDN), produces a haze-free picture without depending on the physical scattering model [40]. The EPDN architecture is integrated in a GAN, followed by a well-designed enhancer. Efforts have been made to apply conditional GAN (cGAN) [41] for haze removal. In [9], the clear image is estimated by a trainable, end-to-end cGAN architecture. In [10], a GAN is trained using a gradient penalty to implement the Lipschitz restriction, to know how the probabilities of the clear images are distributed based upon haze affect images.

As combined information from different domains are difficult to locate in most cases, cycleGAN's provides unsupervised capacity for training. CycleGAN was implemented using Cycle-Consistent Adversarial Networks in [42] for Unpaired Image-to-Image Translation. This is helpful in applications such as image dehazing where paired images are difficult to get, during training. Disentangled Dehazing Network (DDN) was implemented in [8] to estimate scene radiance, map transmission, and global light atmosphere through the combined use of three generators. These techniques involve parameters estimation of the atmospheric scattering model in the training stage which is different from the proposed technique.

Motivated by the success of cycleGAN for haze removal, we propose an improved cycleGAN architecture with an encoder–decoder (ED) architecture as generator. We fit the AOD-NET architecture [7] within the ED architecture in order to generate a clear image using an end-to-end system. Next we discuss the proposed method in detail.

3. Proposed method

We propose an improved CycleGAN architecture for end-to-end dehazing of image, with two major modifications. First, we substitute the CycleGAN feature transformer portion with the AOD-NET architecture. Second, we apply a loss function specific to the dehazing problem. In this section we start with an overview of the AOD-Net architecture, followed by a description of the proposed cycleGAN architecture. Finally we discuss the proposed loss function.

3.1. AOD-NET architecture

We use AOD-NET [7] as a feature transformer in the generator of the CycleGAN architecture [42]. The AOD-NET comprises of a module of estimation K which has 5 layers of convolution to estimate $K(x)$ and many element wise multiplication layers and addition layers to retrieve the clear picture. The K-estimation module comprises of five layers of convolution, which has a "concat1" layer combining the characteristics from the "conv1" and "conv2" layers. Likewise, "concat2" layer brings together the characteristics from "conv2" and "conv3" layers. Similarly, "concat3" combines the characteristics from "conv1", "conv2", "conv3" and "conv4" layers. The need for using K-estimate module is for complete end-to-end modeling for restoring the clean image.

In the proposed method we make a slight modification on the traditional AOD-NET architecture in order to make the proposed model symmetric so that it can be fit into the symmetric architecture of the generator model in the CycleGAN. The modifications are only related to the feature dimensions so as to comply with the input–output feature sizes of the CycleGAN generator. The modified AOD-NET architecture is shown in Fig. 2.

The major reason for using AOD-NET is that it can be seamlessly integrated with other profound models (such as the generator of CycleGAN), forming a pipeline that conducts high-level functions on hazy

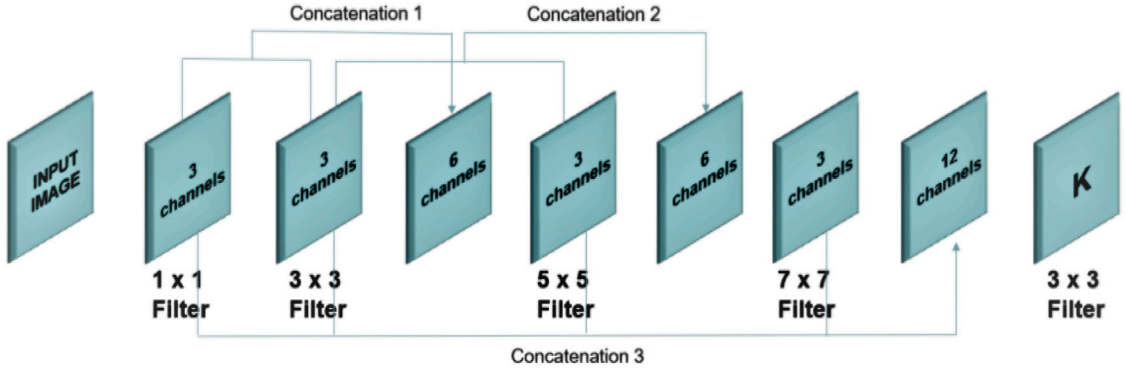

Fig. 2. A diagram of the modified AOD-NET architecture.

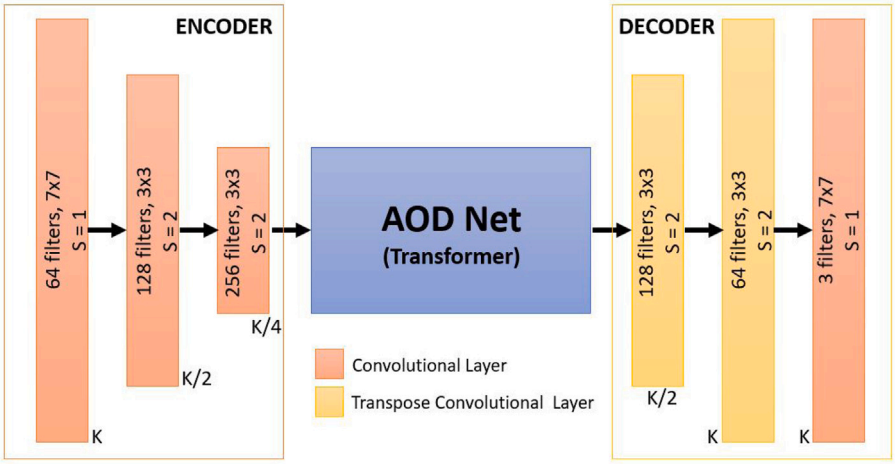

Fig. 3. Architecture of the proposed Generator model of the CycleGAN proposed in this paper.

pictures with an implicit method of dehazing. Moreover, the AOD-NET architecture [7] shows a proven record of ability of extracting image features related to haze. AOD-NET is effective in generalizing well in dehazing an outdoor image.

3.2. Proposed CycleGAN architecture

The proposed CycleGAN architecture follows [8] due to the impressive results produced by the architecture in single image dehazing. However, we include a feature transformer following the AOD-NET architecture between the encoder and the decoder architectures in the generator model of [8]. The generator of the proposed network consists of an encoder, a transformer and a decoder. The encoder is used to obtain the latent features from the hazy image. A feature transformer is used to transform the latent feature of the given image obtained from the encoder, into a different feature space ideally to obtain the latent features of the corresponding haze-free image. Finally a decoder is used to generate the clear image from the transformed features obtained from the transformer part of the generator.

Fig. 3 shows the generator architecture used in the proposed CycleGAN model. We use 3 convolution layers for the encoder part and 2 de-convolution layers for the decoder part. The weights of the conv layers in the AOD-Net are initialized with the pretrained weights, which were obtained by training the AOD-Net using the hyper-parameters setup as in [7].

We use the AOD-NET architecture as a feature transformer inside the generator to transform the features obtained from the encoder part of the proposed CycleGAN architecture, to a desired feature space and feed into the decoder of the generator architecture of the CycleGAN, to generate a proposal for the dehazed image.

The proposed dehazing architecture, as shown in Fig. 4, consists of two generators A and B, and two discriminators DA and DB. The CycleDehaze architecture in [8] benefits from the use of cycle-consistency losses in addition to the adversarial losses in favor of clearing/adding the haze. As a result, the architecture is compelled to maintain input image textural data and produce distinctive haze-free outputs. Cyclic consistency loss enables us to train a model that does not specifically need paired clear and hazy instances of images. However, we add the SSIM loss to the combined loss function in order to enhance the similarity between the ground truth and the haze-free images. The proposed loss function is discussed next.

3.3. Proposed loss function

The objective function (loss function) of the proposed CycleGAN architecture is based upon the loss function proposed in CycleDehaze [8], which is a combination of adversarial loss and cycle consistency loss. The adversarial loss, (i.e., least squares loss) for the generator A, $AL(A, D_y, X)$, is given as follows.

$$AL(A, D_y, X) = (\frac{1}{m}) \sum_{i=0}^{m} [1 - D_y(A(x_i))]^2, \tag{6}$$

where $A(x_i)$ is the generated image from generator A and D_y distinguishes ground truth y from $A(x_i)$ where x_i is an image from domain X. Similarly, the adversarial loss for the generator B, $AL(B, D_x, Y)$, is given as follows.

$$AL(B, D_x, Y) = (\frac{1}{m}) \sum_{i=0}^{m} [1 - D_x(B(y_i))]^2, \tag{7}$$

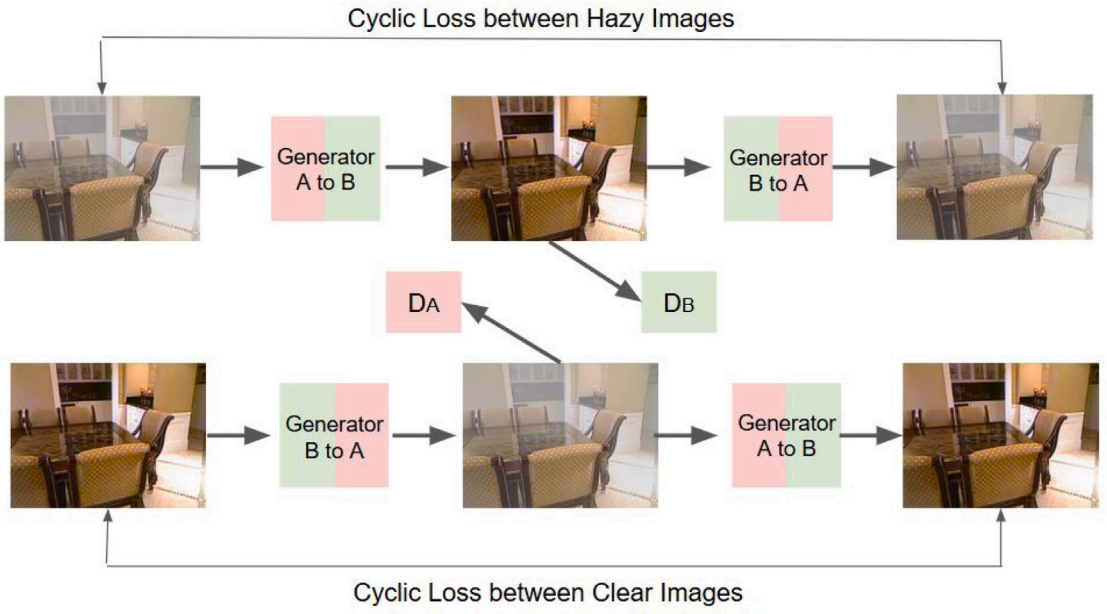

Fig. 4. Overall architecture of the proposed CycleGAN architecture for single image dehazing.

where $B(y_i)$ is the generated image from generator B and D_x distinguishes ground truth x from $B(y_i)$ where y_i is an image from domain Y. Hence, the overall adversarial loss (AL) is given as,

$$AL = AL(A, D_y, X) + AL(B, D_x, Y). \tag{8}$$

The Cycle Consistency loss, which is the pixel-wise mean error (PME) uses an assumption to demonstrate how close we are to returning to the original model when we convert a sample from Domain X to Y by using the function A on an image and then map it back to X by using the inverse function B. In the same way, the loss caused when the samples are translated from Y to X and again back to Y is estimated by PME as follows, which is the pixel-wise error between original and recreated images.

$$PME(A, B, X, Y) = (\frac{1}{m}) \sum_{i=0}^{m} ([B(A(x_i)) - x_i] + [A(B(y_i)) - y_i]), \tag{9}$$

where m is the batch size. Hence, the loss function used in [8] becomes,

$$L = AL + \lambda * PME, \tag{10}$$

where λ controls the impact of loss of cyclic consistency. Hence, the update for the generators A and B are as follows:

$$A *, B *= arg \min_{A,B} \max_{D_x, D_y} L(A, B, D_x, D_y). \tag{11}$$

We attempt to optimize this loss function by using the generators A or B respectively to try to produce clear/hazy images.

Along with the above shown cyclic consistency loss function, an additional SSIM loss function was also considered in the proposed work, to improve the similarity index of the generated clear images. This SSIM loss along with the pixel-wise loss helps in better recovery of objects which were in clutter or a bit darker background. The SSIM loss is introduced to make the generated images more visually appealing. The SSIM loss function is defined as follows:

$$SSIM(A, B, X, Y) = (\frac{1}{m}) \sum_{i=0}^{m} (1 - ssim(B(A(x_i)), x_i) + (1 - ssim(A(B(y_i)), y_i))), \tag{12}$$

where the $ssim(I_1, I_2)$ between two images I_1 and I_2 is calculated by using a Gaussian filter with standard deviation σ. Means and standard deviations at each pixel of both the images are calculated using this filter and finally the ssim at a pixel i is computed by using the following equation

$$ssim(I_1, I_2, i) = \frac{2\mu_{I_1}\mu_{I_2} + C1}{\mu_{I_1}^2 + \mu_{I_2}^2 + C1} \frac{2\sigma_{I_1}\sigma_{I_2} + C1}{\sigma_{I_1}^2 + \sigma_{I_2}^2 + C1}, \tag{13}$$

where μ is the mean and σ is the standard deviation of the images at pixel i over the 11 × 11 Gaussian filter. The constants C1 and C2 are used to prevent the loss value from going to ∞. The considered values of C1 and C2 are 0.0001 and 0.0004 respectively and a value of 10 was used for λ to control the effect of cyclic consistency loss and the adversarial loss.

Finally, the loss function for the proposed model can be written as follows

$$L_{final} = AL + \lambda((1 - \lambda_1)PME + \lambda_1 SSIM). \tag{14}$$

PME in the above equation represents the pixel-wise mean error and now the second part in the above equation is the newly considered cyclic consistency loss. Here λ_1 is used to balance the overall effect of the pixel-wise mean loss and the SSIM loss in cyclic consistency loss. After doing some experiments by using different values of λ_1, the best value was found out to be 0.84, adding more weight-age to the ssim loss and as a result, it puts more emphasis on learning weights which can restore the minute structures from the hazy images more predominantly. Next, we describe the implementation details for the proposed method.

3.4. Implementation details

We train the proposed model using the TensorFlow framework on a NVIDIA TITAN X GPU. Each of the experimental models was trained for 100 epochs with a batch size of 1, which took an average of three days to complete training for each considered model. Adam Optimizer with a learning rate of 0.0001 was used for training. GAN training was unstable when it was used to produce images of large sizes, which constrained us to set the picture sizes to a fixed resolution of 256 × 256.

The proposed generator incorporates the advantages of the AOD-NET architecture and uses it to transform the features from one space to another and to produce high visual quality clear images. We randomly cropped the images to 256 × 256 size during training and while testing,

the images were downsampled to the size of 256 × 256. The model was trained using a batch size of 1 so that more robust characteristics can be captured from each image by the generator in a much better way.

Our motivation was to replace the original cycle GAN's transformer with the model proposed by the authors of AOD-Net, which is a method for single image dehazing. The hyperparameters of AOD-Net were already trained over the popular image dehazing dataset. We experimented with the proposed method after the fine tuning, but the results were not improved. Hence we did not consider tuning the hyperparameters in AOD-Net after combining with the GAN's generator. Next, we show our model's quantitative comparison with the state of the art models and qualitative results on a few well-known hazy images from the real world.

4. Experiments and results

In this section, we provide the results of various experiments carried out on the proposed model. We apply the PSNR and SSIM [43] metrics for quantitative measurement of the proposed model compared to the state-of-the-art. First, we provide a brief description of all the datasets used in this study to validate the proposed method. Next, we discuss the results obtained by the proposed method on different datasets, followed by the result of applying the proposed method on some real hazy images for qualitative comparison with the state-of-the-art.

4.1. Datasets and experiments

Following the recent deep learning-based methods for single image dehazing, we use the NYU-Depth V2 dataset [36] and the RESIDE-β dataset [32] for conducting our experiments. The NYU-Depth V2 dataset contains up of 1449 pairs of scenes, each comprising of a clear scene and the corresponding hazy image after applying synthetic haze on it.

The reside-β dataset [32] consists of two parts, namely Outdoor Training Set (OTS) and Real-world Task Driven Testing Set (RTTS). The OTS part contains around 70,000 hazy images obtained from 2061 real-world outdoor clear images by applying haze of different intensities. A smaller version of this dataset is also available which is called as Reside-Standard which has Synthetic Objective Testing Set (SOTS) containing 500 hazy images. This set was used to get the quantitative performance of all the competing methods.

The above two datasets are first used to train the proposed model without SSIM loss function and then were used to train the model with SSIM loss function separately, and the results are compared. Next, we present the results obtained by the proposed method compared to the state-of-the-art, when applied to the two datasets.

4.2. Results on NYU-depth dataset

We experiment with the proposed modified CycleGAN model with the two different loss functions: first without including the SSIM loss and second, including the SSIM loss. The results are shown in Table 1 compared to the state-of-the-art. The average PSNR and SSIM values are taken as the measures for analyzing the performance of the competing methods.

Table 1 shows that, the proposed haze-specific generator network (i.e., inclusion of the AOD-NET architecture inside the generator network) has improved the performance compared to the general CycleGAN network. Also, the use of perceptual loss and cycle consistency loss helped improve the performance over general CycleGAN architecture (CycleDehaze). Moreover, inclusion of the SSIM loss have enhanced the performance of the proposed model, both in terms of PSNR and SSIM measures. This demonstrates that using AOD-NET as a transformer and also using ssim loss function along with adversarial and cyclic consistency loss functions have actually helped us achieve better outcomes than just using the normal CycleGAN.

Table 1
Average PSNR and SSIM values obtained by the state of the art models and the proposed models when applied on the NYU depth V2 dataset [36].

Model	PSNR	SSIM
Cycle-GAN [42]	13.38	0.59
Cycle-Dehaze [8]	15.41	0.66
DehazeNet [6]	18.96	0.775
MSCNN [31]	19.11	0.8295
Deep DCP [44]	19.25	0.832
AOD-NET [7]	19.69	0.8478
GMAN [45]	20.53	0.8081
Proposed model without ssim loss	22.84	0.9114
Proposed model with ssim loss	21.31	0.8846
*Proposed model (with PSNR loss)	6.92	0.33
*Proposed model (with PSNR and SSIM loss)	6.65	0.29

Table 2
Average PSNR and SSIM values obtained by the state of the art models and the proposed models when applied on the Reside-β dataset [32].

Model	PSNR	SSIM
DehazeNet [6]	21.14	0.8472
MSCNN [31]	17.57	0.8102
AOD-NET [7]	19.06	0.8504
Proposed model (with normal loss function)	21.78	0.7988
Proposed model (with SSIM loss included)	20.05	0.83070

Table 3
The PSNR and SSIM values obtained by the proposed model for different values of λ_1 to penalize PME and SSIM loss, in the proposed loss function, when applied on the NYU depth V2 dataset.

λ_1	PSNR	SSIM
0.25	22.00	0.8186
0.40	22.26	0.8217
0.50	22.49	0.8271
0.75	22.42	0.8330
0.80	22.56	0.8331
0.84	22.77	0.8301
0.90	22.01	0.8175

4.3. Results on reside-β dataset

The next experiment conducted was to train the proposed modified CycleGAN models on Reside-β dataset. This section contains the quantitative results obtained by the competing methods, when reside-β dataset was used for training and testing. The average PSNR and SSIM values of the proposed models when tested on Reside-β dataset compared to the other image dehazing models are shown in Table 2.

Table 2 depicts that, the GAN based model has reduced the performance of the dehazing model in general, typically when applied on outdoor images (as most of the images in the Reside-β dataset are outdoor images). Another point can be noted from Table 2. The PSNR values obtained from the GAN based models shows better results compared to the other deep network models, whereas, the SSIM values for the GAN based models reduces compared to the other deep networks. However, the use of SSIM loss in the proposed model still improved the performance and made it comparable with the state-of-the-art.

Table 1 shows the results of the proposed method compared to the state-of-the-art, when applied on the NYU depth v2 dataset, which consists mainly of indoor images with complicated texture. Whereas, Table 2 shows the results related to the Reside-beta dataset, which mostly consists of outdoor images with varying depth and light. The proposed method, due to the SSIM loss, works better for outdoor images where color variation is more. The SSIM loss helps the result image to come closer to the ground truth RGB values. Hence, works better for Reside-β dataset. However, the proposed method shows comparable results on indoor images as well. The PSNR metric emphasizes on the texture information, especially for indoor images. Whereas, SSIM

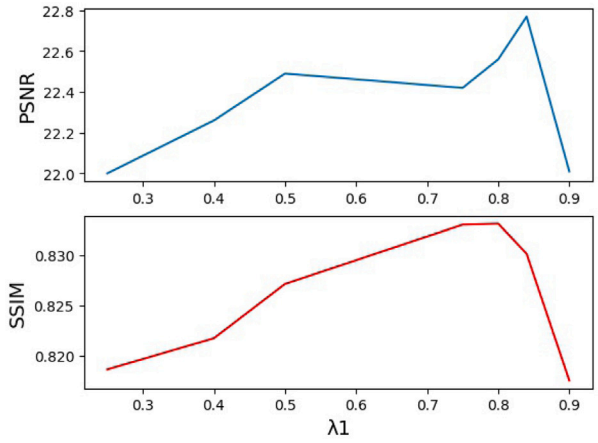

Fig. 5. Graphical representation of the varying trends in PSNR and SSIM results of the reconstructed images as the values of λ_1 is varied.

Fig. 6. Results of the proposed method when tested on two sample images taken from the NYU depth V2 dataset. (a) the hazy images and (b) the clear image dehazed by the proposed method. The proposed model, due to the inclusion of SSIM loss, provides more closer results to the ground truth, by preserving the minute texture information (especially in the second image).

emphasizes on color information. Since Table 2 is based on the Reside-β dataset, where outdoors scenes are available, the proposed SSIM loss based method provides a higher SSIM value, but a lower PSNR value.

4.4. Experiments on the λ_1 value

We perform further experiments on the proposed model to find a suitable value for λ_1 in Eq. (14); the hyper-parameter used for penalizing PME and SSIM loss functions in the cyclic consistency loss considered in this paper. For determining the best suitable value for λ_1, each of the models with respective to a λ_1 value, were trained for 50 epochs each. The efficacy of the proposed model in terms of PSNR and SSIM values obtained on the NYU Depth V2 dataset, for different values of λ_1 are shown in Table 3. A graphical depiction of the same can be seen in Fig. 5.

Table 3 and Fig. 5 depict that, the efficacy of the proposed method achieves its best for a λ_1 value within the range of 0.75 to 0.84.

Fig. 7. Results of the proposed method when tested on two sample images taken from the Reside-β dataset. (a) the hazy images and (b) the clear image dehazed by the proposed method. The proposed model, due to the inclusion of SSIM loss, provides more closer results to the ground truth, by preserving the minute texture information (houses in the first image and background trees in the second image).

Fig. 8. Results of the proposed method when tested on few popular real world hazy images. (a) hazy image (b) generated clear image by the proposed model without ssim loss (c) generated clear image by the proposed model with ssim loss.

This means, the proposed SSIM loss plays more significant role as an objective function, compared to the PME.

4.5. Additional experiments with loss functions

We further experiment with different potential loss functions to make sure that, the proposed model provides result images close to

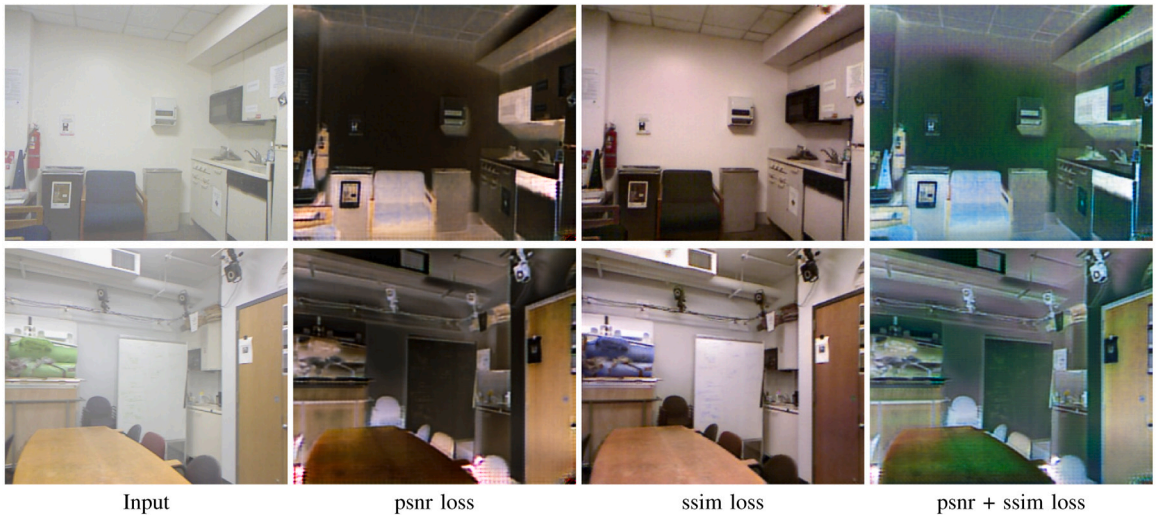

Fig. 9. Qualitative results comparing the proposed models performance when different loss function were considered.

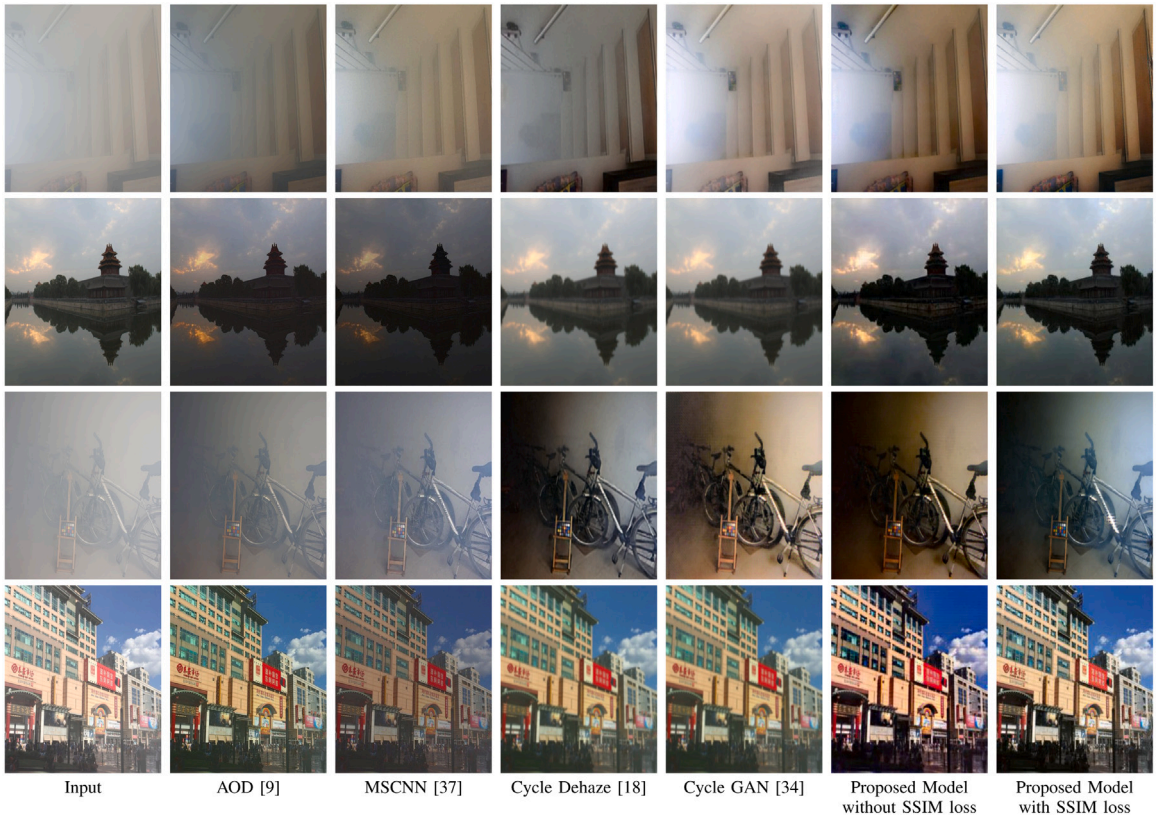

Fig. 10. Qualitative results comparing the proposed method with the state-of-the-arts.

the ground truth, without loosing its perceptual quality by blindly relying on the ground truth. To achieve this goal, we train the proposed model with two more different combinations of loss functions. First, the SSIM loss is replaced by inverse of PSNR values of the reconstructed images, in the proposed loss function. The second experiment was to consider both the SSIM loss and inverse of PSNR, along with the PME in the cyclic consistency loss term of the generator's training loss. The quantitative results for the two additional experiments are shown in Table 1 with the asterisk (*) mark. The results are comparatively inferior than the proposed loss function. The qualitative results for the considered variations on loss functions are shown in Fig. 9, which again establish the efficacy of the proposed loss function.

4.6. Qualitative results

In order to experiment on the robustness of the proposed model, we apply the proposed image dehazing model on two kinds of images: indoor hazy images containing minute texture information (NYU Depth V2 dataset) and outdoor images containing textured background (Reside-β dataset). By analyzing the performance of the proposed method on these two datasets, we observed that the proposed method is robust for both indoor and outdoor hazy images and is able to preserve the minute texture information present in the images while dehazing them.

Fig. 6 demonstrates the qualitative results obtained by the proposed model when tested on two sample images of the NYU Depth V2 dataset. We can observe, especially in the second image of Fig. 6, that the proposed method is able to preserve the minute texture information present in the image, while dehazing it. Fig. 7 shows the qualitative results obtained from the proposed method when tested on two sample images of the Reside-β dataset. Clearly, the proposed model is able to preserve the minute information in both the images of Fig. 7 (houses in the first image and background trees in the second image).

Fig. 8 shows a qualitative comparison between the results of the proposed model with SSIM loss and without SSIM loss. We can observe in Fig. 8 that, the use of SSIM loss to the proposed model help in preserving the minute texture information of the image, and make it more closure to realistic haze-free images. Moreover, the proposed SSIM loss can provide much better color to the objects in the image (e.g., trees in the first image). It can be observed from Fig. 8(b) and (c) that the proposed model has clearly distinguished between what is fog and what is scattered sunlight and also removed only the portion of fog that is noticeable to us while effectively regenerating the sunlight in the picture. As AOD-NET was able to generalize well in distinguishing fog and white light, embedding the same AOD-Net architecture in the generator of the proposed model has boosted the ordinary cycleGAN architecture to successfully generalize well in distinguishing between fog and white light.

In addition to being able to remove haze, the proposed model also focuses on adding haze to pictures due to the cyclic mechanism inherent in the architecture, which can produce much realistic haze-free image. Therefore, the technique learns what is haze in the specified picture irrespective of the issue of dehazing the picture. Since the thickening of haze in some parts of the pictures is very high, in such cases, our model cannot estimate the actual color of the ground truth image.

Few attempts have been made to address the above problem of different haze levels in the image. The most recent of such methods, Xiao et al. [46] have proposed a haze layer-based dehazing algorithm which involves predicting a residual image as an intermediate result which is used to get the final dehazed image. The algorithm was able to get over the problem of not being able to estimate the color in cases of dense haze by being able to predict the different layers of haze that might be present in the input image. But due to the usage of an intermediate layer, the algorithm proposed in [46] may result in loss of textural information in low light parts of the image, which was also one of the problems with AOD-Net. The proposed model has gotten over this issue by using the improved loss function.

We believe that PSNR and SSIM cannot be considered as very good metrics for measuring the effectiveness of a dehazing algorithm, as they both can provide a measure of dispersion of the result image from the ground truth image. However, none of them can measure the perceptual quality of the result image. To overcome this aspect, the proposed loss function combines the benefit of the closeness measure of the result image from ground truth, as well as, the perceptual quality of the result image, which is measured by the perceptual loss term. The qualitative results shown in Fig. 8 of the revised manuscript, on real life images, where paired ground truth images are not available. We can observe that, the proposed loss function could preserve the original color of the image, even better compared to the loss function without SSIM loss.

Finally, we show the qualitative outcomes of the proposed models which were trained on NYU depth V2 dataset and compared the results with the state-of-the-art methods including AOD-Net [7], MSCNN [31], Cycle-Dehaze [8] and CycleGAN [42] in Fig. 10. Clearly, the proposed model along with the SSIM loss, outperforms the state-of-the-art techniques preserving the color and texture of the images.

5. Conclusions

We have proposed an end-to-end cycleGAN network for single image dehazing to produce haze free images from the input hazy images without estimating atmospheric scattering model parameters. In addition, our network offers unpaired training of hazy and ground-based pictures. We have enhanced the standard CycleGAN network by incorporating a haze-specific transformer network into the GAN generator to maintain the visual quality of the produced transparent pictures. Second, we proposed a loss function specific to the problem of single image dehazing, by adding weighted SSIM loss to the traditional loss functions for the dehazing problem. Our experiments show that, inclusion of the haze-specific transformer network in the generator of the CycleGAN architecture enhance the performance of dehazing the image. Moreover, introduction of SSIM loss to the loss function of the network boost the performance of the proposed model preserving the minute textures of the image. Observing the ability of SSIM loss to help in producing images closer to the ground truth, in future the SSIM loss can also be used in some other applications such as image de-noising, de-raining or image to image translation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors wish to thank the NVIDIA company, United States for providing a TITAN X GPU card as a research grant, which is used for conducting the experiments reported in this paper.

References

[1] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, Rob Fergus, Indoor segmentation and support inference from rgbd images, in: European Conference on Computer Vision, Springer, 2012.
[2] Cosmin Ancuti, Codruta O. Ancuti, Radu Timofte, Ntire 2018 challenge on image dehazing: Methods and results, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.
[3] Qingsong Zhu, Jiaming Mai, Ling Shao, A fast single image haze removal algorithm using color attenuation prior, IEEE Trans. Image Process. 24 (11) (2015) 3522–3533.
[4] Kaiming He, Jian Sun, Xiaoou Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2010) 2341–2353.
[5] Raanan Fattal, Dehazing using color-lines, ACM Trans. Graph. (TOG) 34 (1) (2014).
[6] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, Dacheng Tao, Dehazenet: An end-to-end system for single image haze removal, IEEE Trans. Image Process. 25 (11) (2016) 5187–5198.
[7] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, Dan Feng, An all-in-one network for dehazing and beyond, 2017, arXiv preprint arXiv:1707.06543.
[8] Deniz Engin, Anil Genç, Hazim Kemal Ekenel, Cycle-dehaze: Enhanced cycle-gan for single image dehazing, in: Computer Vision and Pattern Recognition Workshops, 2018.
[9] Runde Li, Jinshan Pan, Zechao Li, Jinhui Tang, Single image dehazing via conditional generative adversarial network, in: CVPR, 2018.
[10] Joshua Peter Ebenezer, Bijaylaxmi Das, Sudipta Mukhopadhyay, Single image haze removal using conditional wasserstein generative adversarial networks, 2019, arXiv preprint arXiv:1903.00395.
[11] Yeyao Chen, Mei Yu, Gangyi Jiang, Zongju Peng, Fen Chen, End-to-end single image enhancement based on a dual network cascade model, J. Vis. Commun. Image Represent. 61 (2019) 284–295.
[12] Kushal Borkar, Snehasis Mukherjee, Single image dehazing based on generic regularity, 2018, arxive: 1808.08610.
[13] Yoav Y. Schechner, Srinivasa G. Narasimhan, Shree K. Nayar, Instant dehazing of images using polarization, in: CVPR (1), 2001.
[14] Robby T. Tan, Visibility in bad weather from a single image, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008.
[15] Codruta Orniana Ancuti, Cosmin Ancuti, Philippe Bekaert, Effective single image dehazing by fusion, in: International Conference on Image Processing, IEEE, 2010.

[16] Bin Xie, Fan Guo, Zixing Cai, Improved single image dehazing using dark channel prior and multi-scale retinex, in: 2010 International Conference on Intelligent System Design and Engineering Application, Vol. 1, IEEE, 2010.
[17] Risheng Liu, Xin Fan, Minjun Hou, Zhiying Jiang, Zhongxuan Luo, Lei Zhang, Learning aggregated transmission propagation networks for haze removal and beyond, IEEE Trans. Neural Netw. Learn. Syst. 99 (2018) 1–14.
[18] Xitong Yang, Zheng Xu, Jiebo Luo, Towards perceptual image dehazing by physics-based disentanglement and adversarial training, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[19] Sebastián Salazar Colores, Eduardo Ulises Moya-Sánchez, Juan-Manuel Ramos-Arreguín, Eduardo Cabal-Yépez, Statistical multidirectional line dark channel for single-image dehazing, IET Image Process. 13 (14) (2019) 2877–2887.
[20] Kushal Borkar, Snehasis Mukherjee, Single image dehazing by approximating and eliminating the additional airlight component, Neurocomputing (2020) http://dx.doi.org/10.1016/j.neucom.2020.03.027.
[21] Qi Liu, Xinbo Gao, Lihuo He, Wen Lu, Single image dehazing with depth-aware non-local total variation regularization, IEEE Trans. Image Process. 27 (10) (2018) 5178–5191.
[22] Himanshu Kumar, Sumana Gupta, K.S. Venkatesh, Realtime dehazing using colour uniformity principle, IET Image Process. 13 (11) (2019) 1931–1939.
[23] Yin Gao, Yijing Su, Qiming Li, Jun Li, Single fog image restoration with multi-focus image fusion, J. Vis. Commun. Image Represent. 55 (2018) 586–595.
[24] Peter Morales, Tzofi Klinghoffer, Seung Jae Lee, Feature forwarding for efficient single image dehazing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[25] Kangfu Mei, Aiwen Jiang, Juncheng Li, Mingwen Wang, Progressive feature fusion network for realistic image dehazing, in: Asian Conference on Computer Vision, 2018.
[26] Sanchayan Santra, Ranjan Mondal, Bhabatosh Chanda, Learning a patch quality comparator for single image dehazing, IEEE Trans. Image Process. 27 (9) (2018) 4598–4607.
[27] Yafei Song, Jia Li, Xiaogang Wang, Xiaowu Chen, Single image dehazing using ranking convolutional neural network, IEEE Trans. Multimed. 20 (6) (2017) 1548–1560.
[28] Akshay Dudhane, Subrahmanyam Murala, C^{2}MSNet: A novel approach for single image haze removal, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018.
[29] He Zhang, Vishal M. Patel, Densely connected pyramid dehazing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[30] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, Gang Hua, Gated context aggregation network for image dehazing and deraining, in: Winter Conference on Applications of Computer Vision (WACV), IEEE, 2019.
[31] Anna Wang, Wenhui Wang, Jinglu Liu, Nanhui Gu, Aipnet: Image-to-image single image dehazing with atmospheric illumination prior, IEEE Trans. Image Process. 28 (1) (2018) 381–393.
[32] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, Zhangyang Wang, Benchmarking single-image dehazing and beyond, IEEE Trans. Image Process. 28 (1) (2019) 492–505.
[33] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, Christophe De Vleeschouwer, I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images, in: International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, Cham, 2018.
[34] Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, Christophe De Vleeschouwer, O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.
[35] Codruta O. Ancuti, Cosmin Ancuti, Mateu Sbert, Radu Timofte, Dense haze: A benchmark for image dehazing with dense-haze and haze-free images, 2019, arXiv preprint arXiv:1904.02904.
[36] N. Silberman, D. Hoiem, P. Kohli, R. Fergus, Indoor segmentation and support inference from rgbd images, in: European Conference on Computer Vision, 2012, pp. 746–760.
[37] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. (2014).
[38] He Zhang, Vishwanath Sindagi, Vishal M. Patel, Joint transmission map estimation and dehazing using deep networks, 2017, arXiv preprint arXiv:1708.00581.
[39] Kunal Swami, Saikat Kumar Das, Candy: Conditional adversarial networks based fully end-to-end system for single image haze removal, 2018, arXiv preprint arXiv:1801.02892.
[40] Yanyun Qu, Yizi Chen, Jingying Huang, Yuan Xie, Enhanced Pix2pix dehazing network, in: Computer Vision and Pattern Recognition, 2019.
[41] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, Image-to-image translation with conditional adversarial networks, in: Computer Vision and Pattern Recognition, 2017.
[42] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: International Conference on Computer Vision, 2017.
[43] Alain Hore, Djemel Ziou, Image quality metrics: PSNR vs. SSIM, in: 2010 20th International Conference on Pattern Recognition, IEEE, 2010.
[44] Alona Golts, Daniel Freedman, Michael Elad, Unsupervised single image dehazing using dark channel prior loss, IEEE Trans. Image Process. 29 (2019) 2692–2701.
[45] Zheng Liu, et al., Generic model-agnostic convolutional neural network for single image dehazing, 2018, arXiv preprint arXiv:1810.02862.
[46] Jinsheng Xiao, et al., Single image dehazing based on learning of haze layers, Neurocomputing (2020).