RI-GAN: An End-to-End Network for Single Image Haze Removal

Akshay Dudhane, Harshjeet Singh Aulakh, Subrahmanyam Murala
Computer Vision and Pattern Recognition Lab
Indian Institute of Technology Ropar
2017eez0001@iitrpr.ac.in

Abstract

The presence of the haze or fog particles in the atmosphere causes visibility degradation in the captured scene. Most of the initial approaches anticipate the transmission map of the hazy scene, airlight component and make use of an atmospheric scattering model to reduce the effect of haze and to recover the haze-free scene. In spite of the remarkable progress of these approaches, they propagate cascaded error upstretched due to the employed priors. We embrace this observation and designed an end-to-end generative adversarial network (GAN) for single image haze removal. Proposed network bypasses the intermediate stages and directly recovers the haze-free scene. Generator architecture of the proposed network is designed using a novel residual inception (RI) module. Proposed RI module comprises of dense connections within the multi-scale convolution layers which allows it to learn the integrated flavors of the haze-related features. Discriminator of the proposed network is built using the dense residual module. Further, to preserve the edge and the structural details in the recovered haze-free scene, structural similarity index and edge loss along with the L1 loss are incorporated in the GAN loss. Experimental analysis has been carried out on NTIRE2019 dehazing challenge dataset, D-Hazy [1] and indoor SOTS [22] databases. Experiments on the publically available datasets show that the proposed method outperforms the existing methods for image de-hazing.

1. Introduction

Visibility in the outdoor scene drastically decreases due to the presence of fog in the atmosphere. This degrades the ability of humans or computer vision algorithms to perceive the scene information. Thus, in the presence of the haze or fog particles, a computer vision algorithm faces difficulty to achieve the desired output as in general they expect an input image without a quality degradation. Thus, the presence of the haze or fog particles in the atmosphere degrades the performance of computer vision algorithms such as object detection [9], moving object segmentation [25] etc. Therefore, to improve the performance of vision algorithms in the hazy environment, image de-hazing is a required pre-processing task.

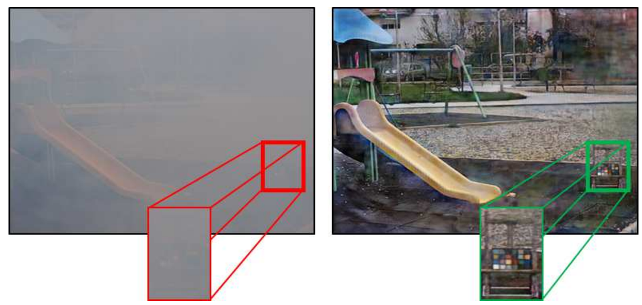

Figure 1. Haze-free scene recovered using the proposed method. The hazy scene is on the left side and recovered haze-free scene using proposed RI-GAN is shown on the right side. Marked regions show the color patch in hazy and haze-free scene.

Research in the field of image de-hazing is roughly divided into prior based methods [12, 35, 16, 3, 36, 43, 18, 13] and learning based methods [6, 26, 8]. Among these, prior based methods rely on the haze relevant priors and extract haze relevant features. These haze relevant features are further used to estimate the scene transmission map and atmospheric light followed by atmospheric scattering model [16] to recover the haze-free scene. Learning-based approaches anticipate these parameters using the trained deep network. In spite of the remarkable progress of these approaches, they propagate cascaded error upstretched due to the employed priors. To resolve this issue, we propose an end-to-end conditional generative adversarial network (cGAN) for single image haze removal. Figure 1 shows the outdoor hazy scene from NTIRE2019 validation set [4, 2] and haze-free scene recovered by proposed network. Proposed network is built using basic principles of residual and inception (RI) modules. Thus, named as RI-GAN. Proposed RI-GAN bypasses the estimation of intermediate feature maps and directly recovers the haze-free scene.

The key contributions of this work are listed below:

1. End-to-end conditional generative adversarial network named as RI-GAN is proposed for image de-hazing.

2. A novel generator network is proposed which is designed using a combination of both residual and inception module.

3. A novel discriminator network is designed using dense connections in the residual block.

4. A combination of structural similarity index (SSIM) loss and edge loss is incorporated along with the L1 loss to optimize the network parameters.

Rest of the paper is organized as Section 1 and 2 illustrate the introduction and literature survey on image de-hazing respectively. Section 3 presents the proposed method for image de-hazing. Section 4 depicts the training of the proposed RI-GAN. Further, the experimental results are discussed in Section 5. Finally, Section 6 concludes the proposed method for image de-hazing.

2. Literature Survey

Effect of the haze is directly proportional to the depth of an object from the camera device. To understand this non-linearity, various approaches have been proposed such as polarized filters [28, 30], use of multiple images of same scenery [7, 23], prior based hazy models [12, 35, 16, 3, 36, 43, 18, 13] etc. Initially, in the area of image de-hazing, Schechner *et al.* [28, 30] proposed the polarized filters. Their approach works with multiple images of the same scene but differs in polarization angle. This approach fails because of its multi-image dependancy. Nayer *et al.* [23] overcame the hardware complexity by correlating dissimilarity between multiple images of the same scene but captured in different weather. With this approach, it is unable to restore the haze-free scene immediately, if multiple images of the same scene are not available for different weather conditions. Cozman *et al.* [7] resolved multi-image dependency by utilizing 3D geometrical model which is based upon the depth information of the hazy scene.

In the last decade, image de-hazing has made remarkable progress due to the convincing assumptions regarding the haze spread or haze density. Tan *et al.* [35] proposed contrast enhancement of the hazy scene. They removed haze by maximizing the local contrast of the hazy image. However, this method fails and create blocking artifacts when there is a depth discontinuity in the hazy image. He *et al.* [16] proposed dark channel prior (DChP) to restore the visibility in the hazy scene. It comprises of dark pixels *i.e.* pixels which are having very low intensity among one of the color channels for a given hazy-free scene. This simple but effective assumption is used to estimate the haze density and atmospheric light to recover the haze-free scene. DChP fails in complicated edgy structures and undergoes the halo effect [6]. The efficiency of [16] depends upon the accurate estimation of the scene transmission map. To estimate the robust transmission map of the hazy scene, researchers follow post-processing techniques such as guided filtering [15], median filtering [18, 41] etc. Lai *et al.* [20] proposed two priors to estimate the optimal transmission map. They estimated locally consistent scene radiance and context-aware scene transmission and utilized atmospheric scattering model to recover the haze-free scene. Wang *et al.* [38] utilized multi-scale retinex algorithm to estimate the brightness components. Further, with the help of a physical model, they recovered the haze-free image. Zhu *et al.* [43] proposed a color attenuation prior (CAP) which considers a HSV color space to extract the haze-relevant features.

To avail the advantages of multiple haze priors, Tang *et al.* [36] proposed regression framework for image de-hazing. They have proposed the extraction of different haze relevant features using existing haze relevant priors and learned the integrated features to estimate the robust scene transmission map. This approach improved the accuracy in single image haze removal. However, it propagates the errors upstretched due to the employed priors. Thus, to minimize the cascading error, researchers make use of convolutional neural networks (CNN). Existing learning-based approaches [6, 26, 8, 9] estimate the scene transmission map using CNN. Further, a global airlight estimation followed by atmospheric scattering model restores the haze-free scene.

Methods discussed above share the same belief that in order to recover a haze-free scene, estimation of an accurate scene transmission map is essential. The atmospheric light is calculated separately and the clean image is recovered using the atmospheric scattering model. Although being intuitive and physically grounded, such a procedure does not directly measure or minimize the reconstruction distortions. As a result, it will undoubtedly give rise to the sub-optimal image restoration quality. The errors in each separate estimation step will accumulate and magnify the overall error. In this context, Li *et al.* [21] designed an end-to-end architecture known as AOD-Net for single image haze removal. They analyzed the internal relationship between the end-to-end de-hazing network and traditional atmospheric model. Further, Swami *et al*. [33] proposed an end-to-end network based on conditional GAN for image dehazing. Recently, researchers [11, 10, 24] make use of unpaired training approach for various computer vision applications. [11, 10] utilized unpaired training approach for image de-hazing whereas [24] found its use for moving object segmentation. In the next Section, we discussed the proposed method for single image haze removal.

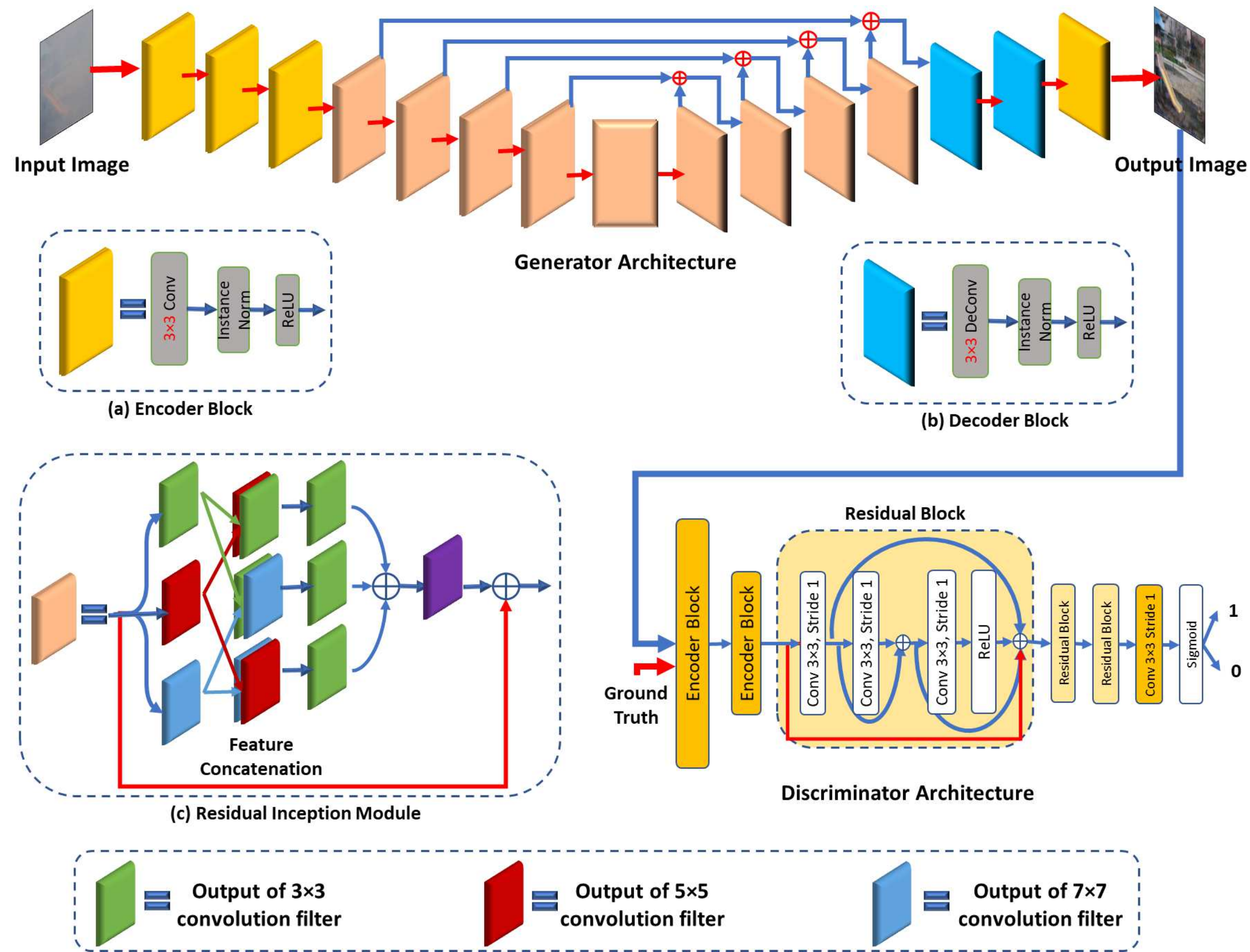

Figure 2. Architecture of the proposed generative adversarial de-hazing network. (a) Encoder block (b) Decoder block (c) Proposed Residual Inception module. $\bigoplus$ denotes the element-wise summation operation. *Best viewed in color*.

3. Proposed Method for Image Dehazing

As discussed in the previous Section, existing atmospheric model-based approaches propagate cascaded error upstretched due to the employed priors. Thus, in this paper, we propose an end-to-end generative adversarial network for single image haze removal. Proposed network incorporates advantages of both residual and inception modules. The motivation behind the use of residual and inception module is discussed in the next subsection.

3.1. Motivation

Training of deeper network is a crucial task as it undergoes vanishing gradients. Thus, training accuracy of such networks degrades with an increase in network depth [14, 32]. Another aspect of training of deeper CNN is the availability of training data. Learning millions of network parameters over comparatively small training set turns into network overfitting. In these contexts, He *et al.* [17] proposed the residual learning approach (ResNet) to optimize the deep networks irrespective of the network depth and a number of network parameters. They introduced the concept of identity mappings to overcome the vanishing gradient problem. The concept of identity mappings is coarsely analogous to the connections in the visual cortex (as it comprises of feed-forward connections). Thus, it outperforms in different popular vision benchmarks such as object detection, image classification, and image reconstruction.

Another important aspect of the design of a deep network involves the choice for filter size (*i.e.* 3×3 *or* 5×5 *etc.*). It becomes more difficult as one has to do this for every layer. There is no as such ground rule which could tell the best combination of filter sizes. Szegedy *et al.* [34] resolved this problem using Inception module. Inception module helps to tackle the network design difficulties by letting the network to decide the best route for itself. These aspects of convolution neural network motivate us to design a network using principles of both residual and inception modules for single image haze removal.

3.2. Proposed Generator Network

The proposed generator network architecture is divided into three parts namely: (1) Encoder block (2) Residual-Inception Module, and (3) Decoder block. Encoder/Decoder block consists of simple convolution/deconvolution layer followed by nonlinear activation function (ReLU). We use instance normalization [37] to normalize the network feature maps. Figure 2 (a) shows the encoder block. We design four encoder and two decoder blocks with filters having a spatial size of 3×3. Parameter details are discussed in Section 3.2.1.

Proposed Residual-Inception module comprises of three parallel convolution layers having a spatial size of 3×3, 5×5 and 7×7 similar to the concept of inception module proposed in [34]. Further, to integrate features learned by the respective convolution layer, we employed feature concatenation followed by convolution layer as shown in Figure 2 (c). We call this a dense concatenation, as feature maps of every layer are integrated with every another layer. Proposed feature integration approach differs from original inception module [34] as it integrates the feature maps in two stages *i.e.* dense concatenation followed by element-wise summation. Finally, 1×1 convolution layer is designed to match the feature dimension with the input of the RI module. Here, we make use of identity mapping [14] shown by the red line in Figure 2 (c) to add these learned features with the input features. It helps to avoid vanishing gradient problem and keeps alive the error gradients across the network.

Olaf *et al.* [27] proposed U-Net architecture for medical image segmentation. The major purpose behind skip connections in U-Net architecture was to share the low-level features learned at initial convolution layers to the deconvolution layers. This concatenation of feature maps helped them to generate the prominent edge information in the output image. Inspired from this, we incorporated skip connections in the proposed RI-GAN as shown in Figure 2 generator architecture. These skip connections help to share the learned features across the network which causes better convergence.

3.2.1 Parameter Details of the Proposed Generator Network

Let a 3×3 Convolution-InstanceNormalization-ReLu layer with `n` filters and stride '1' denoted as `conv3sd1-n`. Similarly, a 3×3 DeConvolution-InstanceNormalization-ReLu layer with `n` filters and upsampling factor '2' denoted as `deonv3up2-n`. In Residual Inception (RI) module, parallel convolution layers of filter-size 3×3, 5×5 and 7×7 having stride 1 and `n/2` filters. Further, convolution layer with stride 1 and filter-size 3×3 has `n` filters. With these details, let us denote RI module by `ParConvSd1-n`. Thus, proposed generator network is represented as: `conv3sd1-64, conv3sd2-256, conv3sd2-256, (ParConvSd1-256)×9, deonv3up2-64, deonv3up2-32, conv3sd1-3.`

3.3. Proposed Discriminator Network

Isola *et al.* [19] proposed a patchGAN to discriminate the generator's fake output from the real one. We utilized the approach of patchGAN in proposed discriminator network. It comprises of combination of encoder block followed by the proposed dense residual block (*shown in Figure 2*). Each encoder block down-samples the input feature maps by a factor of 2 and feeds to the next layer. We went up to the two down-samplings[1].

3.3.1 Parameter Details of the Proposed Discriminator Network

Let, a 3×3 Convolution-InstanceNormalization-ReLu layer with `n` filters and stride '1' denoted as `conv3sd1-n`. The residual block having 3×3 Convolution layers followed by ReLU Layer is denoted by `Res-n` (*shown in Figure 2*). Sigmoid layer is represented by `Sigm`. Thus, proposed discriminator network is represented as: `conv3sd2-64, conv3sd2-256, (Res-256)×2, conv2sd1-1, sigm.`

3.4. Network Loss Function

It is a prime requirement of any image restoration technique to recover the strructural details. Especially in single image haze removal, retaing structural details in recovered haze-free scene improves the scene visibility. Thus, it is required to aquaint the network learning about structural loss along with the L1 loss. Thus, we utilized the structural similarity index metrix (SSIM) as a loss function along with traditional L1 loss. Also, to generate the true edge information we considered the edge loss while training the proposed RI-GAN.

3.4.1 SSIM Loss

Let, x and y are the observed and output image respectively. Also, $G(x)$ represents output of the proposed generator for the input x. Thus, SSIM between $G(x)$ and y is given as follows:

$$SSIM = [l(G(x), y)]^{\alpha} \cdot [c(G(x), y)]^{\beta} \cdot [s(G(x), y)]^{\gamma} \tag{1}$$

[1] Number of encoding levels are experimentally finalized.

where, luminance (l), contrast (c), and structural terms (s) are the characteristics of an image having α, β and γ as the exponents respectively and given as,

$$l\left(G\left(x\right),y\right)=\frac{2\mu_{G(x)}\mu_y+C_1}{\mu_{G(x)}^2+\mu_y^2+C_1}$$

$$c\left(G\left(x\right),y\right)=\frac{2\sigma_{G(x)}\sigma_y+C_2}{\sigma_{G(x)}^2+\sigma_y^2+C_2}$$

$$s\left(G\left(x\right),y\right)=\frac{\sigma_{G(x)y}+C_3}{\sigma_{G(x)}\sigma_y+C_3}$$

If $\alpha=\beta=\gamma$ (the default exponents) and $C_3=\dfrac{C_2}{2}$ then Eq. 1 reduces to,

$$SSIM=\frac{\left(2\mu_{G(x)}\mu_y\right)\left(2\sigma_{G(x)y}+C_2\right)}{\left(\mu_{G(x)}^2+\mu_y^2+C_1\right)\left(\sigma_{G(x)}^2+\sigma_y^2+C_2\right)} \quad (2)$$

where, $\mu_{G(x)}$, μ_y, $\sigma_{G(x)}$, σ_y and $\sigma_{G(x)y}$ are the local means, standard deviations, and cross-covariance for images $G\left(x\right)$, y respectively. C_1 and C_2 are the small constants added to avoid the undefined values.
Hence, SSIM loss can be defined as,

$$\ell_{SSIM}\left(G\right)=1-SSIM\left(G\left(x\right),y\right) \quad (3)$$

3.4.2 Edge Loss

To compute the edge map for a given image, we consider sobel edge detector. Let, $E_{G(x)}$ and E_y represents edge maps of generated and reference (ground truth) scene, then edge loss is given as,

$$\ell_{Edge}\left(G\right)=\parallel E_{G(x)}-E_y\parallel_1 \quad (4)$$

Therefore, overall loss function is,

$$\begin{aligned}\mathcal{L}\left(G,\,D\right)=l_{cGAN}\left(G,\,D\right)+\lambda\cdot l_{SSIM}\left(G\right)\\ +l_{Edge}\left(G\right)+\lambda\cdot l_{L1}\left(G\right)\end{aligned} \quad (5)$$

where, $l_{cGAN}\left(G,\,D\right)$ is a conditional GAN loss [19], l_{L1} is traditional $L1$ loss, and λ is loss weightage[2]. Thus, overall objective of the proposed GAN is given as,

$$G^*=\arg\underset{G}{min}\,\underset{D}{max}\;\mathcal{L}\left(G,\,D\right) \quad (6)$$

In order to generate haze-free scene G^* is used during testing phase.

[2]Experimentally λ=10 is considered for the proposed network.

Table 1. Quantitative evaluation of proposed and existing methods for image de-hazing on a NTIRE2019 [4] challenge database. Note: *SSIM and PSNR - higher is better.*

Approach	**Validation**		**Test**	
	SSIM	**PSNR**	**SSIM**	**PSNR**
Baseline	0.39	10.79	-	-
DChP [16]	0.28	12.90	-	-
Pix2Pix [19]	0.35	13.72	-	-
Proposed Method	**0.47**	**16.31**	**0.54**	**16.47**

4. Training of the Proposed RI-GAN

Training dataset comprises of synthetic and real-world hazy images and their respective haze-free scenes. To generate synthetic images, we consider NYU depth [31] dataset. Indoor hazy images (100) are synthetically generated using procedure given in [8] with $\beta=0.8$, 1.6 and $airlight\left(A\right)=[1,\,1,\,1]$. These synthetically generated hazy images are used only for training and there is no overlap between these images and images used for testing. Real-world hazy and haze-free scenes are collected from the training set of outdoor NTIRE2018 dehazing challenge (35) [5] and NTIRE2019 dehazing challenge (45) [4]. Combinely, 100 synthetic hazy images generated from NYU depth database and 80 outdoor hazy scenes from NTIRE database along with their respective haze-free scenes are used to train the proposed RI-GAN. The remaining setting of the model is similar to the [19]. Proposed network is trained for 200 epochs on a computer having 4.20 GHz Intel Core i7 processor and NVIDIA GTX 1080 8GB GPU.

5. Experimental Results

In this Section, we carry both quantitative and qualitative evaluation to validate the proposed RI-GAN for image de-hazing. We consider structural similarity index (SSIM) [39], peak signal to noise ratio (PSNR) and color difference measure (CIEDE 2000) [29] for quantitative evaluation. We categorize the experiments into two parts: performance of the proposed RI-GAN on synthetic and real-world hazy scenes.

5.1. Performance on Synthetic Hazy Images

We utilized three databases (1) Validation and testing set of NTIRE2019 [4] dehazing challenge, (2) D-Hazy [1], and (3) SOTS [22] to validate the proposed RI-GAN for image de-hazing.

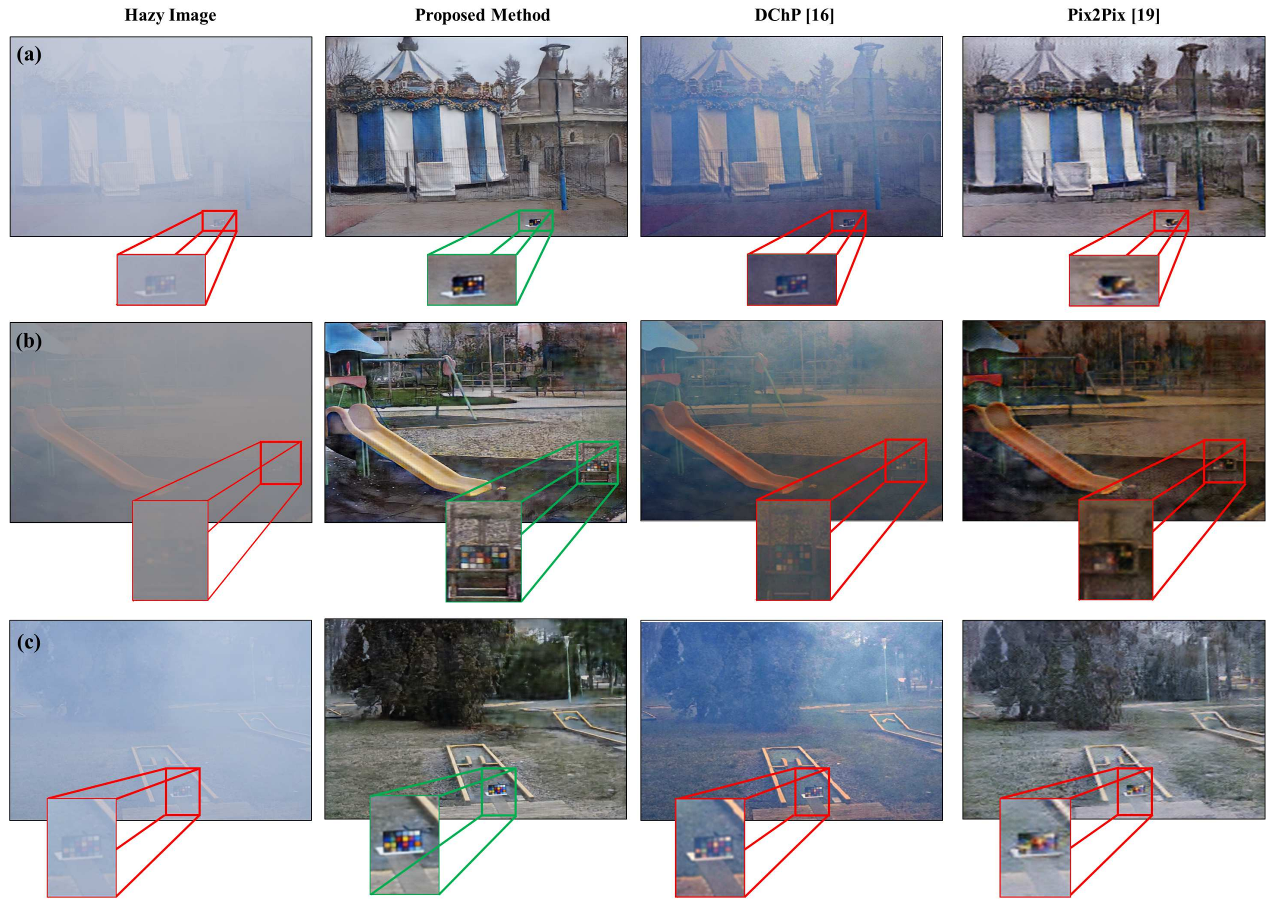

Figure 3. Visual result of proposed RI-GAN and existing methods [16, 19] on NTIRE2019 dehazing challenge database. Marked region denotes the color patch in hazy and haze-free scene. Note: *Please magnify the figure to understand the minor details.*

Table 2. Quantitative evaluation of the proposed and existing methods for image de-hazing on D-Hazy [1] database. Note: *SSIM and PSNR - higher is better and CIEDE2000 - lower is better.*

Approach	SSIM	PSNR	CIEDE2000
DehazeNet [6]	0.7270	13.4005	13.9048
C^2MSNet [8]	0.7201	13.6017	12.4800
MSCNN [26]	0.7231	12.8203	15.8048
DChP [16]	0.7060	12.5876	15.2499
CAP [43]	0.7231	13.1945	16.6783
DDN [40]	0.7726	15.5456	11.8414
AODNet [21]	0.7177	12.4120	16.6565
Pix2Pix [19]	0.7519	16.4320	11.1876
CycleDehaze [11]	0.6490	15.4130	15.0263
Proposed Method	**0.8179**	**18.8167**	**9.0730**

5.1.1 Quantitative Analysis

NTIRE2019 [4] dehazing challenge database consists set of five hazy scenes of spatial resolution 1200×1600 for both validation and testing phase. These images are characterized by dense haze. The dense haze has been produced using a professional haze/fog generator that imitates the real conditions of hazy scenes. Table 1 describes the result of proposed RI-GAN and existing state-of-the-art methods on the NTIRE2019 dehazing challenge database. From Table 1, we can clearly observe that the proposed RI-GAN outperforms prior-based methods for SSIM by a large margin (*almost 20*%). Also, we validate proposed RI-GAN against the conditional GAN approach known as Pix2Pix network [19]. To train the Pix2Pix, we follow the same training data and the training procedure by which proposed RI-GAN is trained. From Table 1, we can clearly observe that the proposed RI-GAN outperforms the existing Pix2Pix approach by a large margin. As ground truths of the testing set are not available, we have not evaluated the existing approaches on the testing set using SSIM and PSNR.

Figure 4. Comparison between proposed RI-GAN and existing methods [16, 6, 8, 21] on real-world hazy images for single image haze removal. Please magnify the figure to understand the minor details in images.

D-Hazy [1] is a standard dataset used to evaluate the performance of various algorithms for image de-hazing. It comprises of pair of 1,449 indoor hazy and respective haze-free scenes. We utilized the entire database *i.e.* 1,449 images for quantitative analysis of the proposed RI-GAN for image de-hazing. The performance of RI-GAN is compared with the existing state-of-the-art methods on D-Hazy [1] database as shown in Table 2. It is evident from Table 2 that the proposed RI-GAN outperforms the other existing methods by a large margin for single image de-hazing. Specifically, proposed RI-GAN increases SSIM by almost 9% as compared to the prior-based deep learning approaches [8, 6, 26] and increases by 5% as compared to end-to-end deep learning methods [40, 21, 42, 11] which shows the robustness of RI-GAN to recover the haze-free scene. Also, there is a significant improvement in the PSNR and CIEDE2000 of the proposed RI-GAN as compared with the existing state-of-the-art methods.

SOTS database [22] is generated from set of 50 images and their respective depth maps from NYU-depth database [31]. From each haze-free image and its depth map, 10 hazy images are generated with different values of β and airlight using atmospheric scattering model. Thus, even though there is an overlap of some scene in D-Hazy and SOTS database, different airlight and density of haze make a large difference between them. Thus, we evaluated the performance of proposed RI-GAN on SOTS database. We considered all 500 hazy images for the analysis. Table 3 depicts the result of the proposed and existing methods on SOTS database. We can observe that proposed RI-GAN outperforms state-of-the-art methods in terms of PSNR and appears very close to AODNet [21].

5.1.2 Qualitative Analysis

Figure 3 shows the sample hazy images from validation and test set of NTIRE2019 image dehazing challenge database and corresponding de-hazed images using the proposed method. From Figure 3, we can clearly observe that proposed RI-GAN adequately reduces the effect of dense haze

Table 3. Quantitative evaluation of proposed and existing methods for image de-hazing on SOTS [22] database. Note: *SSIM and PSNR - higher is better and CIEDE2000 - lower is better.*

Approach	SSIM	PSNR	CIEDE2000
DehazeNet [6]	0.8472	21.1412	6.2645
C^{2}MSNet [8]	0.8152	20.1186	8.306
MSCNN [26]	0.8102	17.5731	10.7991
DChP [16]	0.8179	16.6215	9.9419
CAP [43]	0.8364	19.0524	8.3102
DDN [40]	0.8242	19.3767	9.5527
AODNet [21]	0.8599	19.0868	8.2716
Pix2Pix [19]	0.8200	16.8440	9.8386
CycleDehaze [11]	0.6923	15.8593	14.0566
Proposed Method	**0.8500**	**19.8280**	**8.2993**

and recovers the haze-free scene. Due to robust training, proposed RI-GAN is able to generate the scene information even in dense haze region. Simplest visual evaluation for any method on NTIRE2019 challenge database is to observe the color patch in the recovered haze-free scene. The marked region in Figure 3 witnessed the recovery of the color patch in a haze-free scene without much color distortion. Thus, it can be concluded that the proposed RI-GAN preserves the color information irrespective of the dense haze. Combination of the proposed losses serves as an important role in preserving the structural information. We can clearly observe this from Figure 3 (a) as the minute net-like structure is also recovered in the resultant haze-free scene.

We compared the results of proposed RI-GAN with existing methods [16, 19]. Figure 3 witnessed the failure of prior-based method [16] to reduce the effect of dense haze and to recover the haze-free scene without color distortion. On the other hand, end-to-end approach [19] reduces the haze at satifactory level but fails to retain the structural details in the recovered haze-free scene. From visual analysis, we can conclude that, proposed RI-GAN recovers the haze-free scene at the same time preserves the structural details and color information.

5.2. Real World Hazy Images

Due to the unavailibility of pair of the real-world hazy and haze-free scenes, it is difficult to carry quantitative analysis of image de-hazing algorithms for real-world hazy scenes. Therefore, we carry only qualitative analysis for the real-world hazy scenes. Five frequently used real-world hazy scenes are utilized here for analysis. Result comparison of proposed and existing approaches on these images is shown in Figure 4. From Figure 4, we can clearly observe that the proposed RI-GAN generates the appropriate scene information at the same time preserves the structural details in recovered haze-free scene. We compare the results of existing prior-based hand-crafted and learning approaches [16, 6, 8] and end-to-end dehazing approach [21]. Qualitative analysis shows that proposed RI-GAN outperforms the other existing approaches and generates a visually pleasant haze-free scene.

6. Conclusion

In this work, we propose an end-to-end generative adversarial de-hazing network for single image haze removal. A novel generator network which is designed using residual and inception principles named as Residual Inception GAN (RI-GAN). Also, a novel discriminator network using dense residual module is proposed to discriminate between fake and real samples. To preserve the structural information in the recovered haze-free scene, we propose a combination of SSIM and edge loss while training the proposed RI-GAN. Performance of the proposed RI-GAN has been evaluated on four benchmark datasets namely: NTIRE2019 challenge dataset [4], D-Hazy [1], SOTS [22] and real-world hazy scenes. The qualitative analysis has been carried out by analyzing and comparing the results of proposed RI-GAN with existing state-of-the-art methods for image de-hazing. Experimental analysis shows that the proposed RI-GAN outperforms the other existing methods for image de-hazing. In the future, this work can be extended to analyze the effect of haze on the performance of different algorithms for high-level computer vision task such as object detection, human action recognition, and person re-identification. Also, the architecture of the proposed residual inception module can be extended for other computer vision applications such as single image depth estimation, semantic segmentation.

References

[1] Cosmin Ancuti, Codruta O Ancuti, and Christophe De Vleeschouwer. D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 2226–2230. IEEE, 2016. 1, 5, 6, 7, 8

[2] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte et al. Ntire 2019 challenge on image dehazing: Methods and results. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019. 1

[3] Codruta O Ancuti, Cosmin Ancuti, Chris Hermans, and Philippe Bekaert. A fast semi-inverse approach to detect and remove the haze from a single image. In *Asian Conference on Computer Vision*, pages 501–514. Springer, 2010. 1, 2

[4] Codruta O. Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. In *arXiv:1904.02904*, 2019. 1, 5, 6, 8

[5] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and 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*, pages 754–762, 2018. 5

[6] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 1, 2, 6, 7, 8

[7] F. Cozman and E. Krotkov. Depth from scattering. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 801–806, Jun 1997. 2

[8] Akshay Dudhane and Subrahmanyam Murala. Cˆ 2msnet: A novel approach for single image haze removal. In *Applications of Computer Vision (WACV), 2018 IEEE Winter Conference on*, pages 1397–1404. IEEE, 2018. 1, 2, 5, 6, 7, 8

[9] Akshay Dudhane and Subrahmanyam Murala. Cardinal color fusion network for single image haze removal. *Machine Vision and Applications*, 30(2):231–242, 2019. 1, 2

[10] Akshay Dudhane and Subrahmanyam Murala. Cdnet: Single image de-hazing using unpaired adversarial training. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1147–1155. IEEE, 2019. 2

[11] Deniz Engin, Anil Genc, and Hazim Kemal Ekenel. Cycle-dehaze: Enhanced cyclegan for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 825–833, 2018. 2, 6, 7, 8

[12] Raanan Fattal. Single image dehazing. *ACM Transactions on Graphics (TOG)*, 27(3):72, 2008. 1, 2

[13] Kristofor B Gibson, Dung T Vo, and Truong Q Nguyen. An investigation of dehazing effects on image and video coding. *IEEE Transactions on Image Processing*, 21(2):662–673, 2012. 1, 2

[14] Kaiming He and Jian Sun. Convolutional neural networks at constrained time cost. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5353–5360, 2015. 3, 4

[15] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. In *European Conference on Computer Vision*, pages 1–14. Springer, 2010. 2

[16] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011. 1, 2, 5, 6, 7, 8

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[18] Shih-Chia Huang, Bo-Hao Chen, and Wei-Jheng Wang. Visibility restoration of single hazy images captured in real-world weather conditions. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(10):1814–1824, 2014. 1, 2

[19] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976. IEEE, 2017. 4, 5, 6, 8

[20] Y. Lai, Y. Chen, C. Chiou, and C. Hsu. Single-image dehazing via optimal transmission map under scene priors. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(1):1–14, Jan 2015. 2

[21] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4770–4778, 2017. 2, 6, 7, 8

[22] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 1, 5, 7, 8

[23] Shree K Nayar and Srinivasa G Narasimhan. Vision in bad weather. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 820–827. IEEE, 1999. 2

[24] Prashant Patil and Subrahmanyam Murala. Fggan: A cascaded unpaired learning for background estimation and foreground segmentation. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1770–1778. IEEE, 2019. 2

[25] Prashant W Patil and Subrahmanyam Murala. Msfgnet: A novel compact end-to-end deep network for moving object detection. *IEEE Transactions on Intelligent Transportation Systems*, 2018. 1

[26] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision*, pages 154–169. Springer, 2016. 1, 2, 6, 7, 8

[27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 4

[28] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. Instant dehazing of images using polarization. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001. 2

[29] Gaurav Sharma, Wencheng Wu, and Edul N Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 30(1):21–30, 2005. 5

[30] S. Shwartz, E. Namer, and Y. Y. Schechner. Blind haze separation. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1984–1991, 2006. 2

[31] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *European Conference on Computer Vision*, pages 746–760. Springer, 2012. 5, 7

[32] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. Highway networks. *arXiv preprint arXiv:1505.00387*, 2015. 3

[33] Kunal Swami and Saikat Kumar Das. Candy: Conditional adversarial networks based end-to-end system for single image haze removal. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3061–3067. IEEE, 2018. 2

[34] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015. 3, 4

[35] Robby T Tan. Visibility in bad weather from a single image. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 1, 2

[36] Ketan Tang, Jianchao Yang, and Jue Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2995–3000, 2014. 1, 2

[37] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 4

[38] J. Wang, K. Lu, J. Xue, N. He, and L. Shao. Single image dehazing based on the physical model and msrcr algorithm. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2018. 2

[39] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[40] Xitong Yang, Zheng Xu, and Jiebo Luo. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *In Thirty third-second AAAI conference on Artificial Intelligence (AAAI-18)*, 2018. 6, 7, 8

[41] Jing Yu, Chuangbai Xiao, and Dapeng Li. Physics-based fast single image fog removal. In *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, pages 1048–1052. IEEE, 2010. 2

[42] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251. IEEE, 2017. 7

[43] Qingsong Zhu, Jiaming Mai, and Ling Shao. Single image dehazing using color attenuation prior. In *25th British Machine Vision Conference, BMVC 2014*, 2014. 1, 2, 6, 8