Multi-scale Single Image Dehazing using Perceptual Pyramid Deep Network

He Zhang Vishwanath Sindagi Vishal M. Patel
Department of Electrical and Computer Engineering
Rutgers University, Piscataway, NJ 08854
{he.zhang92, vishwanath.sindagi, vishal.m.patel}@rutgers.edu

Abstract

Haze adversely degrades quality of an image thereby affecting its aesthetic appeal and visibility in outdoor scenes. Single image dehazing is particularly challenging due to its ill-posed nature. Most existing work, including the recent convolutional neural network (CNN) based methods, rely on the classical mathematical formulation where the hazy image is modeled as the superposition of attenuated scene radiance and the atmospheric light. In this work, we explore CNNs to directly learn a non-linear function between hazy images and the corresponding clear images. We present a multi-scale image dehazing method using Perceptual Pyramid Deep Network based on the recently popular dense blocks and residual blocks. The proposed method involves an encoder-decoder structure with a pyramid pooling module in the decoder to incorporate contextual information of the scene while decoding. The network is learned by minimizing the mean squared error and perceptual losses. Multi-scale patches are used during training and inference process to further improve the performance. Experiments on the recently released NTIRE2018-Dehazing dataset demonstrates the superior performance of the proposed method over recent state-of-the-art approaches. Additionally, the proposed method is ranked among top-3 methods in terms of quantitative performance in the recently conducted NTIRE2018-Dehazing challenge. Code can be found at https://github.com/hezhangsprinter/NTIRE-2018-Dehazing-Challenge

1. Introduction

Haze is a common atmospheric phenomenon where the presence of floating matter in the air such as dust, smoke and water particles absorb or scatter the light reflected by objects in the scene, thus causing serious degradation of image quality. In addition to adversely affecting the aesthetic appeal of the image, these degradations introduce severe challenges to computer vision-based systems such as autonomous navigation and driving, where accuracy is of critical importance. Hence, dehazing is an important problem and is being actively addressed by the research community.

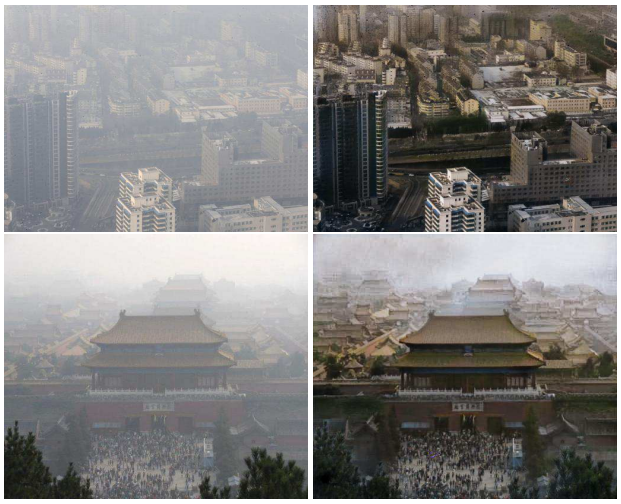
Numerous methods have been proposed in the past and most of these methods, including the recent convolutional neural network (CNN) based approaches, rely on the classical mathematical formulation where the observed hazy image is modeled as a combination of attenuated scene radiance and atmospheric light [3, 9, 14, 19] as described by the following equation:

$$I(x) = J(x)t(x) + A(x)(1 - t(x)), \quad (1)$$

where I is the observed hazy image, J is the true scene radiance, A is the global atmospheric light indicating the intensity of the ambient light, t is the transmission map and x is the pixel location. Transmission map is the distance-dependent factor that affects the fraction of light that reaches the camera sensor. When the atmospheric light A is homogeneous, the transmission map can be expressed as $t(x) = e^{-\beta d(x)}$, where β represents the attenuation coefficient of the atmosphere and d is the scene depth. Most existing single image dehazing methods attempt to recover the clear image or scene radiance J based on the observed hazy image I via estimation of the transmission map t.

Image dehazing is a difficult problem due to it ill-posed nature. It can be observed from Equation 1 that multiple solutions can be found for a single input hazy image. While some approaches tackle this problem by using multiple images [39, 35] or scene depth [27], others add constraints into the optimization framework by including prior information. Other challenges include (i) dependency of haze transmission on unknown depth that varies at different locations, and (ii) inconsistency of haze concentration in different regions of the scene that results in non-uniform dehazing.

Several approaches have been proposed to systematically tackle one or more of the above issues. Initial approaches involved traditional image processing techniques such as histogram-based [25], contrast-based [42] and saturation-based [13] methods to perform dehazing. Methods such as [39, 35] improved over the existing approaches by employing multiple images. Schechner *et al*. used multiple im-



(a) (b)

Figure 1. Sample dehazing results using the proposed method. (a) Input (b) Dehazed output.

ages which are taken with different degrees of polarization. Similarly, Narasimhan *et al*. [35] employed multiple images of the same scene, however under different weather conditions.

Further progress was made based on the physical model by using better assumptions and priors [3, 9, 14, 19]. For instance, Tan *et al*. [43] maximized local contrast of an image using Markov Random Field (MRF) by assuming that local contrast of a clear image is higher as compared to that of a hazy image. He *et al*. [19] proposed dark channel prior method which is based on the assumption that most local image patches in outdoor haze free images consist of some pixels that have very low intensity values. While these methods used local patch-based priors, Berman *et al*. [6] proposed a non-local prior that relies on the observation that colors of a haze-free image are well approximated by a few hundred distinct colors, that form tight clusters in RGB space.

Most recently, several CNN-based approaches [8, 37, 53] have been proposed that demonstrate more robustness as compared to the traditional non-learning based methods. CNN-based methods attempt to learn a non-linear mapping between input image and the corresponding transmission map while following the atmospheric scattering model described by Equation 1. Cai *et al*. [8] presented DehazeNet which is a trainable end-to-end system for medium transmission estimation and consists of specially designed CNN layers that embody existing assumptions/priors. Similarly, Ren *et al*. [2] learned a non-linear mapping using a multi-scale deep neural network. Li *et al*. [30] reformulated the atmospheric scattering model and presented a light weight CNN to produce clear images from hazy images.

As it can be observed from the above discussion, most existing methods rely heavily on the atmospheric scattering model to first estimate the transmission map, followed by the calculation of clear image using Equation 1. Considering this observation, we explore the use of CNNs for directly learning a non-linear mapping between hazy and clear images, which is contrary to existing approaches that learn a mapping between hazy image and the transmission map. This idea is largely motivated by other similar complex computer vision tasks such as visible face synthesis from sparse sample [47, 10], saliency detection [54], deblurring [57], de-raining [16, 52, 51], crowd density estimation [40, 41] *etc*., where CNNs have been successfully used to directly learn a non-linear mapping between input and output. In this attempt, we present a multi-scale image dehazing method using Perceptual Pyramid Deep Network based on the recently popular dense blocks [22] and residual blocks [21]. The proposed method involves an encoder-decoder structure, where the encoder is constructed using dense blocks and the decoder is based on a set of residual and dense blocks followed by a pyramid pooling module [58] to incorporate contextual information. In addition to mean squared loss, perceptual loss based on VGG-16 is used to learn the network weights. To further improve the performance, multi-scale patches are used during training and inference process. Figure 1 shows sample results from the proposed method. Experiments are conducted on two synthetic datasets and a real world dataset to demonstrate the superior performance of the proposed method.

2. Related work

In this section, we review some related work on single image dehazing, starting from the traditional approaches to the most recent CNN-based approaches.

As discussed earlier, some of the initial work on dehazing involved the use of classical image enhancement techniques [25, 42, 13] such as histogram processing, contrast and saturation-based processing to improve the visual appeal of hazy images. Most methods follow the physical atmospheric scattering model and attempt to recover the scene radiance. In order to address the ill-posed nature of the problem, researchers make different assumptions and use appropriate priors. Methods such as [39, 35] employed multiple images to improve the performance. Among the other early approaches, Tan *et al*. [43] proposed to maximize the per-patch contrast based on the observation that haze or fog reduces the contrast of the color images. Kratz and Nishino [28] proposed a factorial MRF model to estimate the albedo and depths field. Fattal *et al*. [14] proposed a physically grounded method by estimating albedo of the scene. He et al. in [19] proposed a dark-channel model to estimate the transmission map which is based on the observation that in case of haze-free image patches, at least one color channel has some pixels with very low intensi-

ties. In the haze image, the intensity of these dark pixels in that channel is mainly contributed by the airlight and hence, these dark pixels can directly provide accurate estimation of the transmission map. Combining a haze imaging model and a soft matting interpolation method, the authors recover a high-quality haze-free image.

To improve the computational efficiency of the dark channel prior-based method, standard median filtering [17], median of median filter [45] and guided image filter [18] are used to replace the time-consuming soft matting [19]. Tang *et al*. [44] combined different types of haze-relevant features with Random Forests to estimate the transmission. Zhu *et al*. [59] estimated the scene depth of a hazy image under color attenuation prior using a linear model whose parameters are learned with a supervised method. Recently, Fattal *et al*. [15] proposed a color-line method based on the observation that small image patches typically exhibit a one-dimensional distribution in the RGB color space. While most of these approaches were based on local patch priors, Berman et al. [6] introduced a non-local prior based on the assumption that colors of a haze-free image are well approximated by a few hundred distinct colors, that form tight clusters in the RGB space.

While these methods relied on hand-crafted representations, the success of CNNs in various computer vision tasks motivated researchers to explore their ability to directly learn a non-linear mapping between input hazy image and its corresponding transmission map. Cai et al. [8] proposed an end-to-end CNN network for estimating the transmission map given an input hazy image. Ren et al. [36] proposed a multi-scale deep neural network to learn the mapping between hazy images and their corresponding transmission maps. The authors first employed a coarse-scale network to predict a holistic transmission map, followed by a refinement stage where a fine-scale network is used to obtain a more detailed transmission map. Dudhane and Murala [12] addressed the issue of color distortion in the earlier CNN-based work by presenting a multi-stage CNN. In the first stage, their network fuses color information present in hazy images and generates multi-channel depth maps, where as the second stage estimates the scene transmission map using a multi channel multi scale CNN.

Since these methods consider only the transmission map in their CNN frameworks, they are limited in their abilities to perform end-to-end dehazing. More recent methods [30, 50, 53] address this issue by considering the dehazing task in addition to transmission map estimation in their frameworks. Li *et al*. [30] designed a light-weight CNN by including the atmospheric scattering model into the network. By doing so, these methods minimize the reconstruction errors thereby improving the quality of dehazing. More recently, He and Patel [50] proposed an end-to-end dehazing method called Densely Connected Pyramid Dehazing Network (DCPDN), which can jointly learn the transmission map, atmospheric light and dehazing all together. The end-to-end learning is achieved by directly embedding the atmospheric scattering model into the network, thereby ensuring that the proposed method strictly follows the physics driven scattering model for dehazing. For training their transmission map estimation network, they use additional edge-preserving loss to preserve sharp edges and avoid halo artifacts. Simultaneously, several benchmark datasets for both synthetic and real-world hazy images for dehazing problems are introduced to the community [56, 31, 1, 46, 38].

3. Proposed method

Figure 2 illustrates the overview of the proposed CNN-based multi-scale single image dehazing framework. Inspired by the success of encoder-decoder architectures in various tasks such as image denoising [55, 48], segmentation [33] and other image-to-image translation [23], our proposed network consists of an encoder, that takes in input hazy image and maps it to a latent space (intermediate feature maps), and a decoder that maps the latent space to the corresponding clear haze-free image. We carefully design the architecture of the encoder and decoder with appropriate type and set of convolution blocks. Our work is closely related to that of [50] with a few important differences such as (i) unlike their method where they involve transmission map estimation as an intermediate step, we aim to directly learn a non-linear mapping between hazy image and its corresponding clear image, (ii) the network architectures are different, and (iii) in our case, we use the perceptual loss function in addition to the standard L2 loss to train the network, which results in substantial improvements in the quality of the dehazed images. In the following subsections, we present the details of the proposed network architecture, loss functions and the training methodology.

3.1. Network architecture

As discussed earlier, we aim to directly learn a mapping between the input hazy image and its corresponding clear image using an encoder-decoder type network.

Encoder. The encoder is constructed using dense blocks from Densely Connected Convolutional Networks [22] and a residual block as shown in Figure 2. Huang *et al*., based on the observation that CNNs can be significantly deeper and can be trained efficiently if they contain shorter connections between layers close to the input and those close to the output, introduced densely connected networks, where each layer is connected to every other layer in the feed-forward fashion. In contrast to earlier approaches, the feature maps of all preceding layers are used as inputs to a particular layer. By employing these dense connections, the au-

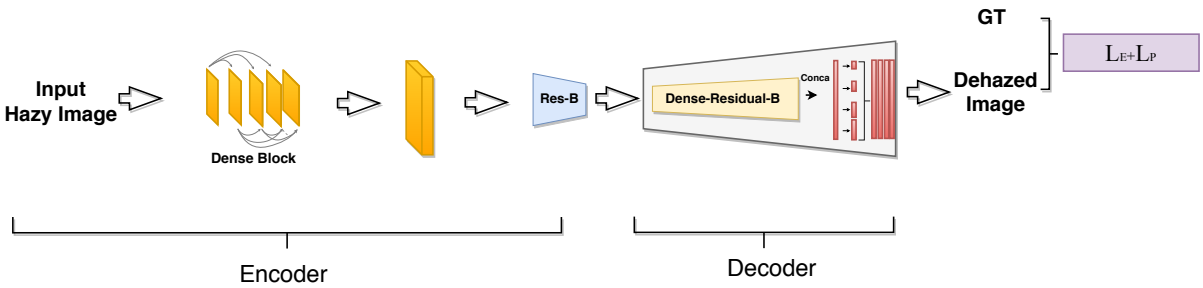
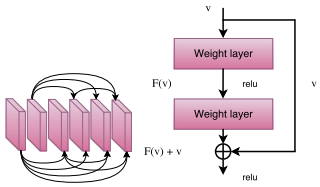


Figure 2. Overview of the proposed Multi-scale Single Image Dehazing using Perceptual Pyramid Deep Network.

thors are able to address the issue of vanishing gradients and strengthen feature propagation while substantially reducing the number of parameters in the network. Due to these convincing advantages, we construct the encoder using dense blocks. The dense-net blocks have a similar structure to that of Dense-net 121 network [22], where the first dense-block contains 12 densely-connected layers, second block contains 16 densely-connected layers and the third block contains 24 densely-connected layers. The weights for each stream are initialized from the pre-trained Dense-net 121 network. Each block consists of a set of layers where each layer receives feature maps from all earlier layers as input as shown in Figure 3(a). This type of connectivity ensures maximum information flow to occur during the forward and backward pass thus making the training process much easier especially when using deeper networks.



(a) (b)

Figure 3. (a) Illustration of dense connections in a dense-block with 6 layers. (b) Residual block.

Decoder. Similar to the encoder, we carefully design the decoder structure with a set of residual and dense blocks. The residual blocks are mainly inspired by residual learning in ResNet [21], where network layers are intelligently reformulated to learn residual functions with reference to the layer inputs, instead of learning unreferenced functions. This reformulation eases the training process, especially in case of deeper networks. The residual block (illustrated in Figure 3(b)), which is the building block in ResNet, is defined as:

$$u = F(v, W_i) + v, \tag{2}$$

where v and u are input and output features of a particular layer and $F(x, W_i)$ is the residual function that has to be learned.

The basic building block of the proposed decoder, which we call as dense-residual block, consists of a two-layer dense block and an upsampling transition block, followed by two residual blocks. Such a configuration allows us to efficiently combine the advantages offered by these two types of blocks thereby enabling high quality reconstruction of dehazed images. Note that, the dense-block along with the upsampling transition block behaves as a refinement function to recover the high-level details lost during the encoding process, thereby resulting in better quality results. The decoder consists of a set of five dense-residual blocks followed by a pyramid pooling module. Similar to [58], where context at various levels in the image is fused for scene parsing, the key idea is to include hierarchical global prior, containing information at different scales and different sub-regions. Four pyramid scales of bin sizes 1×1, 2×2, 4×4 and 8×8 are used. The pooled features undergo dimensionality reduction along the depth via 1×1 convolutions. These pooled and reduced features are upsampled using bilinear interpolation and concatenated to the input features. Finally, these feature maps are combined using 1×1 convolutions to produce the dehazed output.

3.2. Loss function

It has been demonstrated in many earlier work [23, 49, 32] that, it is very important to choose an appropriate loss function especially while training a CNN-based reconstruction network. Traditional methods that use L_2 error have known to produce blurry output and several recent work have attempted to address this issue by using additional loss functions [24, 11]. Inspired by these methods, the weights of the proposed network are learned by minimizing the combination of L_E reconstruction error and perceptual loss (L_P) function as shown below:

$$L = L_E + \lambda_P L_P, \tag{3}$$

where L_E is the standard L_2 loss function and is defined as:

$$L_E = \frac{1}{CWH} \sum_{c=1}^{C} \sum_{w=1}^{W} \sum_{h=1}^{H} \|G(I(c,w,h), \Theta) - I_t(c,w,h)\|_2. \tag{4}$$

Here, I is a C-channel input hazy image with a resolution of $W \times H$, I_t is the corresponding target clear image, G represents the network function and Θ are network parameters. The perceptual loss function L_P is defined using high-level features extracted from a pre-trained convolutional network. The aim is to minimize the perceptual difference between the reconstructed image and the ground truth image. In this work, L_P is based on VGG-16 architecture and is defined as follows:

$$L_P = \frac{1}{C_v W_v H_v} \sum_{c=1}^{C_v} \sum_{w=1}^{W_v} \sum_{h=1}^{H_v} \|\phi_V(G(I,\Theta)) - \phi_V(I_t)\|_2,$$

where ϕ_V is a non-linear transformation that produces a $C_v \times W_v \times H_v$-dimensional feature map. In our work, we use features at layer relu3_1 in the VGG-16 model.

4. Datasets and training details

In this section, we describe the dataset used for training along with other details about training and inference methodology.

4.1. Dataset

For training the proposed network, we use the NTIRE2018-Dehazing challenge dataset[1]. This is one of the most recent datasets introduced to benchmark the current state-of-the-art image dehazing techniques and promote further research in the field. This dataset consists of a wide variety of images categorized into two subsets: indoor set and outdoor set.

Indoor: The NTIRE-Dehazing Indoor dataset [4] consists of 25 training images, 5 validation images, and 5 test images. All images are with very large image sizes (approximately 3000×3000). As the test dataset ground truth is not released, we report and compare the performances on the validation set.

Outdoor: The NTIRE-Dehazing Outdoor dataset [5] consists of 35 training images, 5 validation images, and 5 test images. All images are with very large image sizes (approximately 3000×3000). The haze has been produced using a professional haze/fog generator that imitates the real conditions of hazy scenes. As the test dataset ground truth is not released, we report and compare the performances on the validation set.

4.2. Training

The resolution of images in the NTIRE2018-Dehazing challenge dataset is very large and due to memory considerations, it is infeasible to use the entire image for training. A potential solution is to downsample the images to smaller resolution and use them for training/inference. However, this might result in loss of crucial high-level details thereby affecting the test performance. Hence, we address this issue by following a patch-based training strategy, where the whole image is divided into different smaller-sized patches. Then the network is optimized using the cropped pairs (input and ground truth). Although the memory issue is addressed using the cropping strategy, the patch-based learning reduces receptive field of the network due to which the global context information is lost. To overcome this, we employ a multi-scale cropping strategy, where we crop patches of different sizes (512×512, 1024×1024, 1024×2048, 2048×2048 and original resolution). These patches are then resized to a resolution of 640×640 before training.

During training, we use ADAM [26] as the optimization algorithm with learning rate of 2×10^{-3} for both generator with batch size of 1. All the training samples are resized to 640×640 [2]. We trained the network for 400000 iterations. We choose $\lambda_{E,l_2} = 1$, $\lambda_{E,g} = 0.5$ for the losses.

4.3. Multi-scale ensemble inference

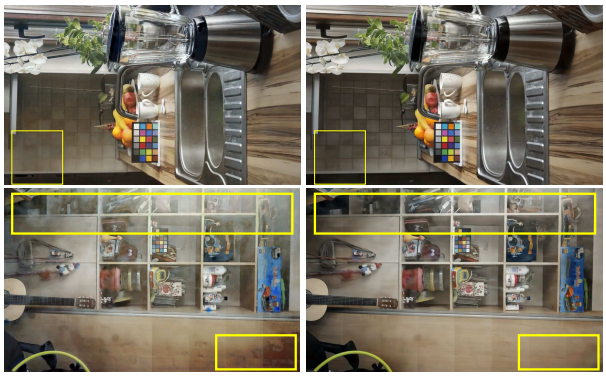
In order to maximize the potential performance of our model, we employ the multi-scale ensemble strategy similar to the one used to improve performance in object detection systems [34], where a multi image pyramid is used during the inference process and detection results are then combined using non-maximum suppression. Similarly, we use multi-scale inference as described below.

Indoor. For indoor dataset, we leverage a two-scale strategy for testing. Basically, we created two sets of overlapping patches from the test image. For the first set, the size of the overlapping patch is chosen such that width is larger than the height (2048×1024). For the second set, the size of the overlapping patch is chosen such that the width is equal to the height (2048×2048). Patches in both the sets are forwarded through the network to obtain the dehazed results (patches). The output patches in each set are then merged appropriately to form the entire output image. The output images from both sets are then combined using a simple ensembling method, where the final output is computed as the mean of the two output images from the two sets.

Outdoor. For outdoor dataset, we follow a slightly different strategy that involves two scales. We created two sets of overlapping patches from the test image. For the first set, the size of the overlapping patch is chosen such that width is greater than the height (3072×1536). The second set consists of a single image that is obtained by downsampling the input test image to a resolution of 1024×1024. The patches in the first set are forwarded through the network and the resulting patches are merged into a single image and resized to the original resolution. The image in the second set is forwarded through the network and the result is upsampled

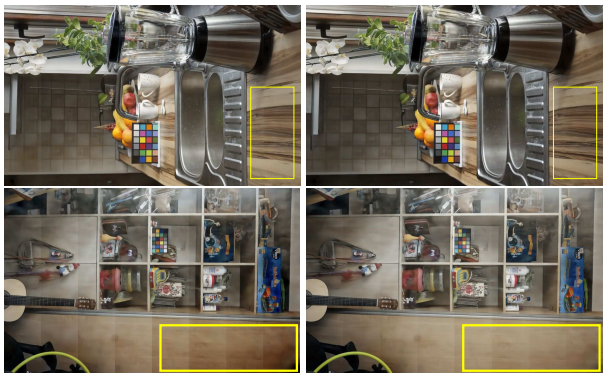
[1] http://www.vision.ee.ethz.ch/en/ntire18/

[2] This is the largest image size that can be fitted in Titan X.



(a) (b)

Figure 4. Ablation study on loss function. (a) Without perceptual loss (only L_E is used). (b) With perceptual loss.



(a) (b)

Figure 5. Ablation study on the type of inference used. (a) Single-scale. (b) Multi-scale.

Table 1. Ablation study: Quantitative results for different configurations of the proposed network.

	S-L_E	S-L_E-L_P	M-L_E-L_P
PSNR (dB)	21.38	21.45	22.53
SSIM	0.8467	0.8576	0.8705

to the original resolution. In addition, we upsample the results from the second set to original resolution. The output images from both sets are then combined using a simple ensembling method, where the final output is computed as the mean of the two output images from the two sets.

5. Experiments and results

In this section, we first present the results of ablation studies conducted to understand the effects of different components in the proposed method. This is followed by a detailed comparison of results of the proposed method with several recent approaches [20, 60, 36, 6, 7, 29] on both synthetic and real datasets.

Table 2. Quantitative results on the NTIRE2018-Dehazing challenge indoor and outdoor datasets.

	Indoor		Outdoor	
Method	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Input	13.8	0.7302	13.56	0.5907
He *et al*. [20]	14.43	0.7516	16.78	0.6532
Zhu *et al*. [60]	12.24	0.6065	16.08	0.5965
Ren *et al*. [36]	15.22	0.7545	17.56	0.6495
Berman *et al*. [6, 7]	14.12	0.6537	15.98	0.5849
Li *et al*. [29]	13.98	0.7323	15.03	0.5385
Ours	**22.53**	**0.8705**	**24.24**	**0.7205**

5.1. Ablation Study

In order to study the effect of different components in the proposed method such as perceptual loss and multi-scale inference, we conduct experiments to perform a detailed ablation study. Following five configurations of the proposed method are trained and evaluated on the NTIRE2018-Dehazing challenge indoor dataset: (i) S-L_E: Single scale inference with the proposed netowrk optimized using only L_E loss, (ii) S-L_E-L_P: Single scale inference with the proposed netowrk optimized using L_E and L_P loss, and (iii) M-L_E-L_P: Multi-scale inference with the proposed network optimized using L_E and L_P loss.

The quantitative results of these configurations are shown in Table 1. It can be observed that the addition of perceptual loss results in improved performance. Similarly, the use of multi-scale ensemble-based inference results in additional improvements. Similar observations about the qualitative performance can be made from Figure 4 and Figure 5.

5.2. Comparison with State-of-the-art Methods

In this section, we demonstrate the effectiveness of the proposed approach by conducting various experiments on two synthetic datasets (NTIRE2018-Dehazing challenge indoor and outdoor) and a real-world dataset. All the results are compared with five state-of-the-art methods: He *et al.* (CVPR'09) [20], Zhu *et al* (TIP'15) [60], Ren *et al.* [36] (ECCV'16), Berman *et al.* [6, 7] (CVPR'16 and ICCP'17) and Li *et al.* [29] (ICCV'17).

Evaluation on real dataset. The proposed method is evaluated and compared against recent approaches on many real-world images that are downloaded from the Internet published by earlier work. Figure 6 shows results for sample real images. It can be observed that some of the methods such as [19] and [36] are unable to completely remove haze, while the other methods ([60, 6]) tend to darken some regions or result in color distortion. In contrast, our method is able to remove the haze completely, in most cases, while generating realistic colors.

Evaluation on synthetic dataset. The proposed network

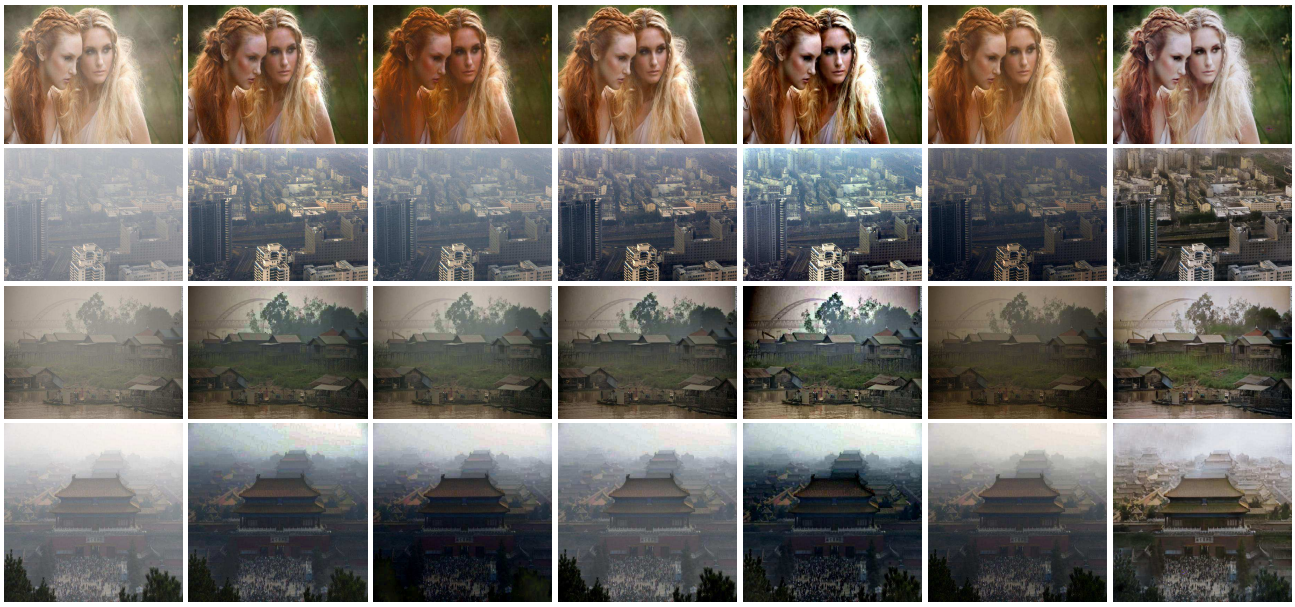


Figure 6. Qualitative comparison of results on real world images.

Figure 7. Qualitative comparison of results on the NTIRE2018-Dehazing indoor dataset.

is evaluated on two synthetic datasets **Indoor** and **Outdoor**. Since the datasets are synthesized, the ground truth images for validation set are available, enabling us to evaluate the performance qualitatively and quantitatively. Table 2 shows the quantitative performance of the proposed method against several recent methods on indoor and outdoor dataset respectively. It can be observed that the proposed method outperforms other approaches by significant margin.

Figures 7 and 8 illustrate dehazing results of the proposed method compared with recent approaches on indoor and outdoor validation sets respectively. It can be observed that even though previous methods are able to remove haze from the input image, they tend to either over-dehaze or under-dehaze the image making the result either darker or hazier in the result. In contrast, it can be observed from our results that the proposed approach preserve sharp contours with less color distortion and are more visually closer to the ground-truth.

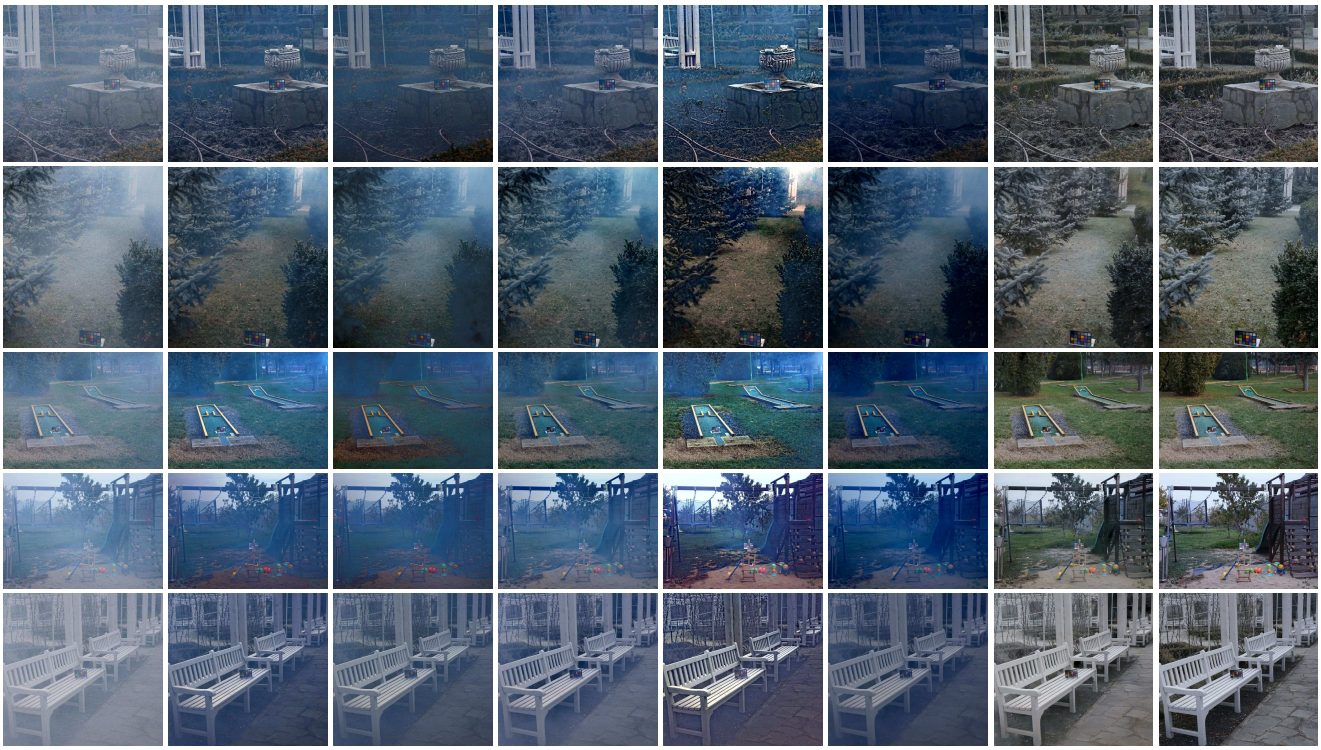


Figure 8. Qualitative comparison of results on the NTIRE2018-Dehazing outdoor dataset.

5.3. NTIRE-2018 Dehazing Challenge

This work was initially proposed for the purpose of participating in the NTIRE2018 Dehazing Challenge. The challenge consists of two different tracks: **Indoor** and **Outdoor**. Qualitative performance results (SSIM and PSNR) on the test dataset for top-five ranked methods, as provided by the organizers, are presented in Table 3. As it can be observed from the table, our proposed method achieves best performance in the indoor dataset and comparable performance in the outdoor dataset.

Table 3. PSNR and SSIM results for NTIRE-2018 Dehazing Challenge.

	Indoor		Outdoor	
	PSNR	*SSIM*	*PSNR*	*SSIM*
Ours	**24.973**	**0.881**	24.029	0.775
Method 1	22.866	0.857	**24.598**	**0.777**
Method 2	22.909	0.864	23.877	0.775
Method 3	20.911	0.751	23.180	0.705
Method 4	20.354	0.829	24.232	0.687

6. Conclusion

We presented an end-to-end single image dehazing network that efficiently learns a non-linear mapping between hazy images and corresponding clear images. In contrast to the existing methods, our method attempts to directly recover the dehazed image instead of first estimating the transmission map. The proposed method involves an encoder-decoder structure with a pyramid pooling module in the decoder to incorporate context information while decoding. The network is learned by minimizing the standard means squared error and perceptual loss. Multi-scale patches are used during training and inference process to further improve the performance. Experiments are conducted on the recently released NTIRE2018-Dehazing dataset and the results are compared against several recent methods. Furthermore, ablation studies are conducted to understand the effects of different components of the proposed method.

Acknowledgement

This work was supported by an ARO grant W911NF-16-1-0126.

References

[1] C. Ancuti, C. O. Ancuti, and C. 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. 3

[2] C. O. Ancuti and C. Ancuti. Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing*, 22(8):3271–3282, 2013. 2

[3] C. O. Ancuti, C. Ancuti, C. Hermans, and P. 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] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer. I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images. *ArXiv e-prints*, Apr. 2018. 5

[5] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer. O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. *ArXiv e-prints*, Apr. 2018. 5

[6] D. Berman, S. Avidan, et al. Non-local image dehazing. In *CVPR*, pages 1674–1682, 2016. 2, 3, 6, 7, 8

[7] D. Berman, T. Treibitz, and S. Avidan. Air-light estimation using haze-lines. In *Computational Photography (ICCP), 2017 IEEE International Conference on*, pages 1–9. IEEE, 2017. 6

[8] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 25(11):5187–5198, 2016. 2, 3

[9] F. Cozman and E. Krotkov. Depth from scattering. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 801–806. IEEE, 1997. 1, 2

[10] X. Di, V. A. Sindagi, and V. M. Patel. Gp-gan: Gender preserving gan for synthesizing faces from landmarks. *arXiv preprint arXiv:1710.00962*, 2017. 2

[11] A. Dosovitskiy and T. Brox. Generating images with perceptual similarity metrics based on deep networks. In *Advances in Neural Information Processing Systems*, pages 658–666, 2016. 4

[12] A. Dudhane and S. Murala. C2msnet: A novel approach for single image haze removal. *arXiv preprint arXiv:1801.08406*, 2018. 3

[13] R. Eschbach and B. W. Kolpatzik. Image-dependent color saturation correction in a natural scene pictorial image, Sept. 12 1995. US Patent 5,450,217. 1, 2

[14] R. Fattal. Single image dehazing. *ACM transactions on graphics (TOG)*, 27(3):72, 2008. 1, 2

[15] R. Fattal. Dehazing using color-lines. volume 34, New York, NY, USA, 2014. ACM. 3

[16] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017. 2

[17] K. B. Gibson, D. T. Vo, and T. Q. Nguyen. An investigation of dehazing effects on image and video coding. *IEEE transactions on image processing*, 21(2):662–673, 2012. 3

[18] K. He, J. Sun, and X. Tang. Guided image filtering. In *European conference on computer vision*, pages 1–14. Springer, 2010. 3

[19] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2011. 1, 2, 3, 6

[20] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Trans. on PAMI*, 33(12):2341–2353, 2011. 6, 7, 8

[21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2, 4

[22] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, volume 1, page 3, 2017. 2, 3, 4

[23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint*, 2017. 3, 4

[24] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016. 4

[25] T. K. Kim, J. K. Paik, and B. S. Kang. Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering. *IEEE Transactions on Consumer Electronics*, 44(1):82–87, 1998. 1, 2

[26] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[27] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, and D. Lischinski. *Deep photo: Model-based photograph enhancement and viewing*, volume 27. ACM, 2008. 1

[28] L. Kratz and K. Nishino. Factorizing scene albedo and depth from a single foggy image. In *ICCV*, pages 1701–1708. IEEE, 2009. 2

[29] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. An all-in-one network for dehazing and beyond. *ICCV*, 2017. 6, 7, 8

[30] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4770–4778, 2017. 2, 3

[31] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. RESIDE: A Benchmark for Single Image Dehazing. *ArXiv e-prints*, Dec. 2017. 3

[32] C. Li and M. Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *ECCV*, pages 702–716, 2016. 4

[33] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015. 3

[34] M. Najibi, P. Samangouei, R. Chellappa, and L. Davis. Ssh: Single stage headless face detector. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4875–4884, 2017. 5

[35] S. G. Narasimhan and S. K. Nayar. Contrast restoration of weather degraded images. *IEEE transactions on pattern analysis and machine intelligence*, 25(6):713–724, 2003. 1, 2

[36] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, pages 154–169. Springer, 2016. 3, 6, 7, 8

[37] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang. Gated fusion network for single image dehazing. *arXiv preprint arXiv:1804.00213*, 2018. 2

[38] C. Sakaridis, D. Dai, and L. Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 2018. 3

[39] Y. Y. Schechner, S. G. Narasimhan, and S. 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. 1, 2

[40] V. A. Sindagi and V. M. Patel. Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting. In *Advanced Video and Signal Based Surveillance (AVSS), 2017 14th IEEE International Conference on*, pages 1–6. IEEE, 2017. 2

[41] V. A. Sindagi and V. M. Patel. Generating high-quality crowd density maps using contextual pyramid cnns. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1879–1888, Oct 2017. 2

[42] J. A. Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on image processing*, 9(5):889–896, 2000. 1, 2

[43] R. 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. 2

[44] K. Tang, J. Yang, and J. Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *CVPR*, pages 2995–3000, 2014. 3

[45] J.-P. Tarel and N. Hautiere. Fast visibility restoration from a single color or gray level image. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2201–2208. IEEE, 2009. 3

[46] J.-P. Tarel, N. Hautiere, L. Caraffa, A. Cord, H. Halmaoui, and D. Gruyer. Vision enhancement in homogeneous and heterogeneous fog. *IEEE Intelligent Transportation Systems Magazine*, 4(2):6–20, 2012. 3

[47] L. Wang, V. A. Sindagi, and V. M. Patel. High-quality facial photo-sketch synthesis using multi-adversarial networks. In *FG*, 2018. 2

[48] P. Wang, H. Zhang, and V. M. Patel. Sar image despeckling using a convolutional neural network. *IEEE Signal Processing Letters*, 24(12):1763–1767, 2017. 3

[49] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. *arXiv preprint arXiv:1711.11585*, 2017. 4

[50] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. *arXiv preprint arXiv:1803.08396*, 2018. 3

[51] H. Zhang and V. M. Patel. Density-aware single image deraining using a multi-stream dense network. *arXiv preprint arXiv:1802.07412*, 2018. 2

[52] H. Zhang, V. Sindagi, and V. M. Patel. Image de-raining using a conditional generative adversarial network. *arXiv preprint arXiv:1701.05957*, 2017. 2

[53] H. Zhang, V. Sindagi, and V. M. Patel. Joint transmission map estimation and dehazing using deep networks. *arXiv preprint arXiv:1708.00581*, 2017. 2, 3

[54] J. Zhang, Z. Lin, J. Brandt, X. Shen, and S. Sclaroff. Top-down neural attention by excitation backprop. In *European Conference on Computer Vision*, pages 543–559. Springer, 2016. 2

[55] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 3

[56] Y. Zhang, L. Ding, and G. Sharma. Hazerd: an outdoor scene dataset and benchmark for single image dehazing. In *Proc. IEEE Intl. Conf. Image Proc.*, pages 3205–3209, 2017. 3

[57] J. Zhang13, J. Pan, W.-S. Lai, R. W. Lau, and M.-H. Yang. Learning fully convolutional networks for iterative non-blind deconvolution. 2

[58] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2881–2890, 2017. 2, 4

[59] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015. 3

[60] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015. 6, 7, 8