Fast Deep Multi-patch Hierarchical Network for Nonhomogeneous Image Dehazing

Sourya Dipta Das*
Jadavpur University
Kolkata, India
dipta.juetce@gmail.com

Saikat Dutta*
IIT Madras
Chennai, India
cs18s016@smail.iitm.ac.in

Abstract

Recently, CNN based end-to-end deep learning methods achieve superiority in Image Dehazing but they tend to fail drastically in Non-homogeneous dehazing. Apart from that, existing popular Multi-scale approaches are runtime intensive and memory inefficient. In this context, we proposed a fast Deep Multi-patch Hierarchical Network to restore Non-homogeneous hazed images by aggregating features from multiple image patches from different spatial sections of the hazed image with fewer number of network parameters. Our proposed method is quite robust for different environments with various density of the haze or fog in the scene and very lightweight as the total size of the model is around 21.7 MB. It also provides faster runtime compared to current multi-scale methods with an average runtime of 0.0145s to process 1200×1600 HD quality image. Finally, we show the superiority of this network on Dense Haze Removal to other state-of-the-art models.

1. Introduction

Outdoor images are often deteriorated due to the extreme weather, such as fog and haze, which influences visibility issues in the scene because of the degradation of color, contrast and textures for different distant objects, selective attenuation of the light spectrum. Restoring such hazed images has become an important problem in many computer vision applications like visual surveillance, remote sensing, and Autonomous transportation etc. Most of early methods proposed for image dehazing are based on the classic atmospheric scattering model which is shown as the following equation. 1.

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

where, x represents pixel locations, $I(x)$ is the observed hazy image, $J(x)$ is the dehazed image, $t(x)$ is called

---
*Equal contribution.

medium transmission function and A is the global atmospheric light. Recently, Deep learning based methods have shown remarkable improvements though those methods suffer from degradation of colour, texture in image, halo artifacts, haze residuals and distortions. In our problem statement, Non-homogeneous haze in the scene can be seen in the real world situation where different spatial domains of the image can be affected by different levels of haze. The degradation level also vary a lot for objects at different scene depth due to non-uniform haze distribution in the image. Few example images of such Non-homogeneous haze are shown in figure 4. Dehazing model should put more effort to handle non-uniform haze and different degradation between different scene depth jointly. Multi-scale and scale-recurrent models can be a viable solution in this type of problem because of its coarse-to-fine learning scheme by hierarchical integration of features from different spatial scale of the image. This type of methods is inefficient because of high runtime and large model size due to a lot of convolution and Deconvolution layers. Apart from that, increasing depth of layers at fine scale levels may not always improve the perceptual quality of the output dehazed image. On the contrary, main goal of our model is to aggregate features multiple image patches from different spatial sections of the image for better performance. The parameters of our encoder and decoder are very less due to residual links in our model which helps in fast dehazing inference. The main intuition behind our idea is to make the lower level network portion focus on local information by extracting local features from the finer grid to produce residual information for the upper level part of the network to get more global information from both finer and coarser grid which is achieved by concatenating convolutional features.

2. Related Work

Most early work of image dehazing methods is developed on atmosphere scattering model as it's physical model. In that respect, previous works on image dehazing can be

segregated into two classes which are traditional image prior-based methods and end to end deep learning based methods. Traditional image prior based methods relies on hand-crafted statistics from the images to leverage extra mathematical constraints to compensate for the information lost during reconstruction. On contrary, deep learning based methods learn the direct relationship between haze and haze-free image by utilizing multistage, attention mechanisms etc. Here, we discussed some recent deep learning based methods with state-of-the-art results.

Zhang et al.[23] proposed a dehazing network with edge-preserving densely connected encoder-decoder architecture that jointly learns the dehazed image, transmission map and atmosphere light all together based on the scattering model for dehazing. In their encoder-decoder architecture, they use a multilevel pyramid pooling module and to improve their results further, joint-discriminator based on GAN is used to incorporate the correlation between estimated transmission map and dehazed image. Deng et al.[8] presents a multi-model fusion network to combine multiple models in its different levels of layers and enhance the overall performance of image dehazing. They generate the multi-model attention integrated feature from various CNN features at different levels and fed it to their fusion model to predict dehazed image for an atmospheric scattering model and four haze-layer separation models altogether. After that, they fused the corresponding results together to generate the final dehazed image. Qin et al.[17] proposed a novel Feature Attention module which fuses Channel Attention with Pixel Attention while considering different weighted information of different channel-wise features and uneven haze distribution on different pixels of the hazed image. For Outdoor hazy images, their work proves superiority though it didn't work well in case of dense dehazing. Liu et al.[14] proposed a grid network with attention-based multi-scale estimation which overcomes the bottleneck problems found in general multi-scale approach. Apart from that, their method also consists of pre-processing and post-processing modules. The pre-processing module used in this method is trainable to get more relevant features from diversified pre-processed image inputs and it outperforms the other hand picked classical pre-process techniques. The post-processing module is finally used on intermediate dehazed image to get more finer dehazed image. Their study shows how their method works quite independently and does not take any advantage from atmosphere scattering model for image dehazing.

Unlike other multi-stage methods, Li et al.[13] used a level aware progressive deep network to learn different levels of haze from its different stages of the network by different supervision. Their network tends to progressively learn gradually more intense haze from image by focusing on a specific part of image with a certain haze level. They have also devised a adaptive hierarchical integration tech-

nique by cooperating with the it's memory network component and domain information of dehazing to emphasize the well-reconstructed parts of the image in it's each stage of the network. Liu et al.[15] suggests a method to learn a haze relevant image priors by using a iteration algorithm with deep CNNs. They achieve this by using gradient descent method to optimize a variational model with image fidelity terms and proper regularization. this method indeed a great combination of properties from classical deep learning based method and physical hazed image formation model. Sharma et al.[19] explored the application of Laplacians of Gaussian (LoG) of the images to reattain the edge and intensity variation information. They optimize their end-to-end deep model by per-pixel difference between Laplacians of Gaussians of the dehazed and ground truth images. they additionally do adversarial training with a perceptual loss to enhance their results. Apart from other physical scattering model based methods, GAN , multiscale or multistage deep networks, Image dehazing can also be posed as image to image translation problem. Qu et al.(2019)[18] proposed their solution as an enhanced Pix2Pix Model which is widely used in image style transfer, image to image translation etc. problems. Their method consists of a GAN with a Enhancer modules to support the dehazing process to get more detailed, vivid image with less artifacts. Their work also proved superiority over other methods in the aspect of the perceptual quality of the dehazed images.

3. Proposed Method

We use a Multi-patch and a Multi-scale network for Non-homogeneous Image Dehazing. In this section, we describe these two architectures in detail.

3.1. Multi-patch Architecture:

We use Deep Multi-patch Hierarchical Network(DMPHN). DMPHN is originally used for Single Image Deblurring[22]. We use (1-2-4) variant of DMPHN in this paper. For the sake of completeness, we will discuss the architecture in the following.

DMPHN is a multi-level architecture. There is an encoder-decoder pair in each level. Each level works on different number of patches. In DMPHN(1-2-4), the number of patches used is 1,2 and 4 from top to bottom levels respectively. The top-most level (level-1) considers only one patch per image. In the next level(level-2), the image is divided into two patches vertically. In the bottom-most level(level-3) the patches from previous level are further divided horizontally, resulting in total 4 patches.

Let us consider an input hazy image I^H. We denote j-th patch in i-th level as $I^H_{i,j}$. In level-1, I^H is not divided into any patches. In level-2, I^H is divided vertically into $I^H_{2,1}$ and $I^H_{2,2}$. In level-3, $I^H_{2,1}$ and $I^H_{2,2}$ are divided horizontally
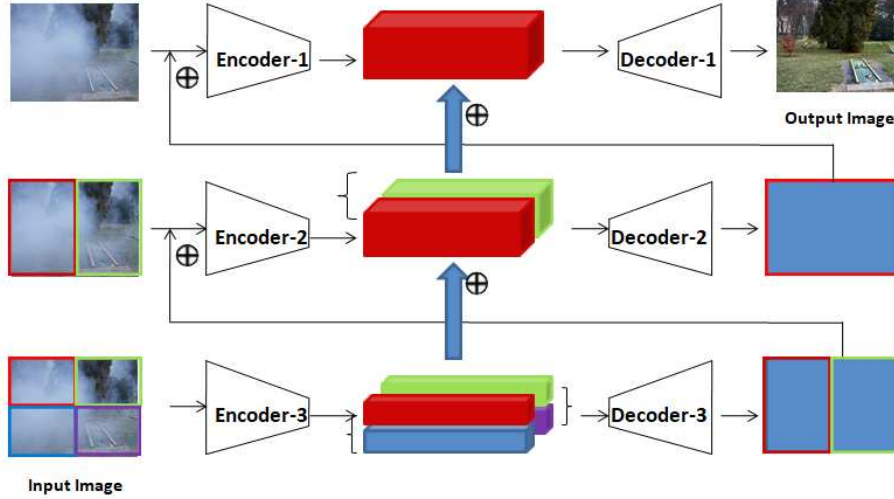
Figure 1: Architecture diagram of Deep Multi-Patch Hierarchical Network. '{' denotes spatial concatenation and $\oplus$ denotes residual addition.

to create 4 patches, $I_{3,1}^H, I_{3,2}^H, I_{3,3}^H$ and $I_{3,4}^H$. Encoders and Decoders at i-th level is denoted as Enc_i and Dec_i respectively.

The information flow in DMPHN is bottom-up. Patches in the lowest level are fed to encoder Enc_3 to generate corresponding feature maps.

$$F_{3,j} = Enc_i(I_{3,j}^H), \forall j \in [1,4] \qquad (2)$$

We concatenate spatially adjacent feature maps to obtain a new feature representation.

$$P_{3,j} = [F_{3,2j-1}, F_{3,2j}], \forall j \in [1,2] \qquad (3)$$

where [.,.] stands for concatenation.

The new concatenated features are passed through decoder Dec_3.

$$Q_{3,j} = Dec_3(P_{3,j}), \forall j \in [1,2] \qquad (4)$$

The decoder output is added with patches in the next level and fed to encoder.

$$F_{2,j} = Enc_2(I_{2,j}^H + Q_{3,j}), \forall j \in [1,2] \qquad (5)$$

The encoder outputs are added with respective decoder inputs from previous level. Then the resulting feature maps are spatially concatenated.

$$F_{2,j}^* = F_{2,j} + P_{3,j}, \forall j \in [1,2] \qquad (6)$$
$$P_2 = [F_{2,1}^*, F_{2,2}^*] \qquad (7)$$

P_2 is then fed to Dec_2 to generate residual feature maps for level-2.

$$Q_2 = Dec_2(P_2) \qquad (8)$$

Decoder output at level-2 is added to input image and passed through Enc_1. Encoder output F_1 is added with decoder output at level-2, Q_2.

$$F_1 = Enc_1(I^H + Q_2) \qquad (9)$$

F_1 is added with P_2 and fed to Dec_1 to produce the final dehazed output $\hat{I}$.

$$P_1 = F_1 + P_2 \qquad (10)$$
$$\hat{I} = Dec_1(P_1) \qquad (11)$$

3.2. Multi-scale Architecture:

We also experiment with a multi-scale architecture. We name this architecture Deep Multi-scale Hierarchical Network(DMSHN). The details of the architecture are described as follows.

Input hazy image I^H is downsampled by factor of 2 and 4 to create an image pyramid. We call these downsampled images $I_{0.5}^H$ and $I_{0.25}^H$ respectively. The architecture consists of 3 levels where each level has a pair of encoder and decoder. Encoder and decoder at level i is denoted as Enc_i and Dec_i respectively.

At the lowest level $I_{0.25}^H$ is fed to encoder Enc_3 to obtain feature map F_3 and is further passed through decoder Dec_3 to feature representation P_3.

$$F_3 = Enc_3(I_{0.25}^H) \qquad (12)$$
$$P_3 = Dec_3(F_3) \qquad (13)$$

P_3 is upscaled by factor of 2 and added to $I_{0.5}^H$ and passed through encoder Enc_2 to generate F_2^*. Encoder output from
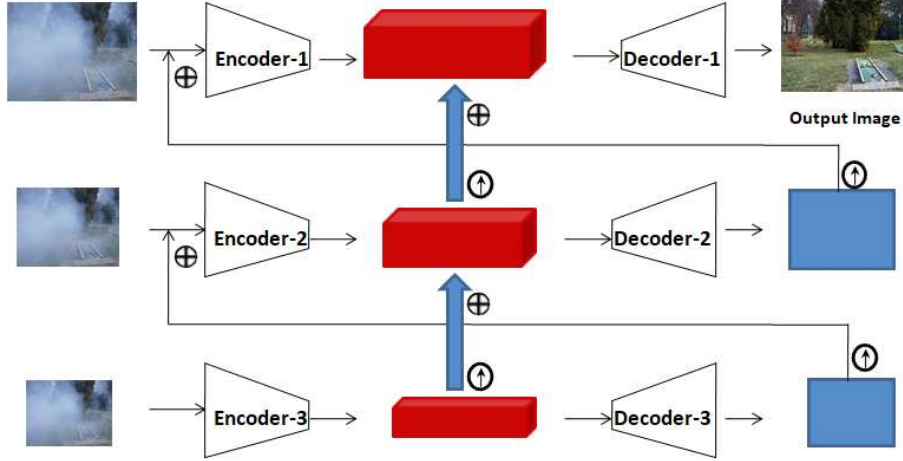
Figure 2: Architecture diagram of Deep Multi-Scale Hierarchical Network. $\oplus$ denotes Upsampling by factor of 2 and $\oplus$ denotes residual addition.

previous level is upscaled and added to intermediate feature map F_2^* and fed to the decoder Dec_2.

$$F_2^* = Enc_2(I_{0.5}^H + up(P_3)) \qquad (14)$$
$$F_2 = F_2^* + up(F_3) \qquad (15)$$
$$P_2 = Dec_2(F_2) \qquad (16)$$

where $up(.)$ denotes Upsampling operation by a factor of 2. Residual feature map P_2 from level-2 is added to the input hazy image and fed to encoder Enc_1. Encoder output is added with upscaled F_2 and passed through decoder to synthesize the dehazed output $\hat{I}$.

$$F_1^* = Enc_2(I^H + up(P_2)) \qquad (17)$$
$$F_1 = F_1^* + up(F_2) \qquad (18)$$
$$\hat{I} = Dec_1(F_1) \qquad (19)$$

3.3. Encoder and Decoder Architecture:

We use the same encoder and decoder architecture at all levels of DMPHN and DMSHN. The encoder consists of 15 convolutional layers, 6 residual connections and 6 ReLU units. The layers in the decoder and encoder are similar except that 2 convolutional layers are replaced by deconvolutional layers to generate dehazed images as output.

4. Experiments

4.1. Dataset Description:

We used NH-HAZE dataset[3] provided for NTIRE 2020 Nonhomogeneous Image Dehazing challenge in our experiments. This dataset contains a total of 55 hazy and clear image pairs, divided into trainset of 45 image pairs, validation set of 5 image pairs and test set of 5 image pairs. Val-

idation and test ground truth images are not publicly available at this moment. Resolution of images in this dataset is 1200×1600. This dataset contains hazed and hazefree images of various outdoor scenes. A few hazefree and hazed image pairs from this dataset is shown in Figure-4.

4.2. Training data preparation:

Due to the small amount of available data, we divide each image into 100 non-overlapping patches. Thus we obtain a training set of 4500 image-pairs of resolution 120×160. No data augmentation techniques were used.

4.3. Loss functions:

We use a linear combination of the following loss functions as our optimization objective.

Reconstruction loss: Reconstruction loss helps the network to generate dehazed frames close to the ground truth. Our reconstruction loss is a weighted sum of MAE or L_1 loss and MSE or L_2 loss. The reconstruction loss is given by,

$$L_r = \lambda_1 L_1 + \lambda_2 L_2 \qquad (20)$$

where $L_1 = \left\| \hat{I} - I \right\|_1$ and $L_2 = \left\| \hat{I} - I \right\|_2$

Perceptual loss: L_2 distance between features extracted from conv4_3 layer of VGGNet[20] of predicted and ground truth images are used as Perceptual loss[11]. Perceptual loss is given by,

$$L_p = \left\| \phi(\hat{I}) - \phi(I) \right\|_2 \qquad (21)$$

TV loss: We use Total Variation(TV) loss[11] makes predictions smooth. TV loss is given by,

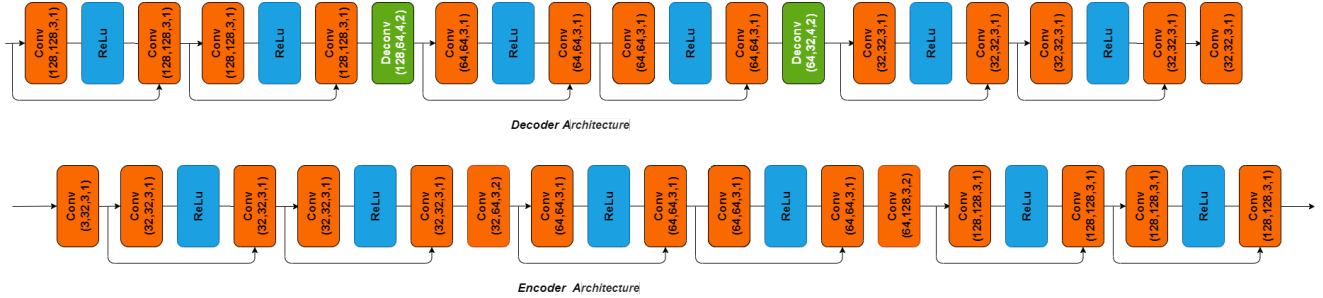$$L_{tv} = \left\| \nabla_x \hat{I} \right\|_2 + \left\| \nabla_y \hat{I} \right\|_2 \qquad (22)$$

Figure 3: Encoder and Decoder architecture. Within brackets written values are Input Channel, Output Channel,Kernel and Stride respectively.

Figure 4: A snapshot of Training Dataset. Top row contains hazy images and bottom row contains corresponding ground truth images.

Our final loss function is given by,

$$L = \lambda_r L_r + \lambda_p L_p + \lambda_{tv} L_{tv} \qquad (23)$$

In our experiments we choose $\lambda_r = 1$, $\lambda_p = 6e - 3$, $\lambda_{tv} = 2e - 8$. λ_1 and λ_2 is chosen to be 0.6 and 0.4 respectively.

4.4. Training details:

We developed our models using Pytorch[16] on a system with AMD Ryzen 1600X CPU and NVIDIA GTX 1080 GPU. We use Adam optimizer[12] to train our networks with values of β_1 and β_2 0.9 and 0.99 respectively. We use batchsize of 8. Initial learning rate is set to be 1e-4 which is gradually decreased to 5e-5. We train our models until convergence.

4.5. Testing details:

We test our models' performance on the given full resolution images of validation data. Please note that, our models are fully convolutional, hence difference between train and test image size should not matter.

4.6. Results:

4.6.1 Quantitative and Qualitative Results:

As ground truth for validation set is not publicly available, we submit our validation results to Codalab server. We compare performance of our models with three state-of-the-art dehazing models namely AtJ-DH[10], 123-CEDH[9] and FFA-Net[17]. The quantitative results on Validation set are given in Table-1. DMPHN is performing better than the rest of the models. It can be observed that our Multi-patch network is performing better than our Multi-scale network in terms of both PSNR and SSIM. At lower levels of DMPHN, the network works on patch level, so the network learns local features compared to global features learnt by DMSHN, which explains the performance gain in DMPHN.

Apart from decent dehazing results, it is to be noted that both DMPHN and DMSHN are lightweight and efficient models. Checkpoints of both the networks take 21.7 MB on disk. GPU processing times for DMPHN and DMPSN make them suitable for real-time applications.

| (a) Input Image | (b) AtJ-DH | (c) 123-CEDH | (d) FFA-Net | (e) Ours (DMPHN) | (f) Ours (DMSHN) |

Figure 5: Qualitative results on NH-HAZE[3] Validation dataset.

	PSNR	SSIM	Runtime(s)
AtJ-DH[10]	15.94	0.5662	0.0775
123-CEDH[9]	14.59	0.5488	0.0559
FFA-Net[17]	10.43	0.4168	1.7472
DMPHN	**16.94**	**0.6177**	**0.0145**
DMPSN	16.42	0.5991	0.0210

Table 1: Quantitative results on NH-HAZE[3] Validation set.

4.6.2 NTIRE 2020 challenge on NonHomogeneous Image Dehazing:

We participated in NTIRE 2020 challenge on NonHomogeneous Image Dehazing[5]. 27 teams submitted results in test phase, out of which 19 teams don't take help of extra training data like Dense-Haze[2, 6] and OHaze[4, 1]. The test results were evaluated on Fidelity measures as well as Perceptual Measures. Fidelity measures included PSNR and SSIM[21], where LPIPS[24], Perceptual Index(PI)[7] and Mean Opinion Score(MOS) were used as Perceptual metrics. For fair comparison, we note down performances of some submissions that used only NH-HAZE dataset in Table-2. Our DMPHN network produced moderate quality outputs both in Fidelity and Perceptual metrics. Our network is the fastest entry among all the submissions.

4.6.3 Dense Haze Removal:

DMPHN is effective for dense haze removal as well. We trained our network on Dense-HAZE dataset[2]. We train on 50 images for training and use 5 images for test. We compare the performance with AtJ-DH[10], 123-CEDH[9]

and FFA-Net[17]. Quantitative results and GPU runtimes are shown in Table-3. We observe that DMPHN is significantly better than other models both in terms of fidelity measures and runtime. Figure-6 shows qualitative comparison with the said models.

4.7. Conclusion

In this paper, we use a Multi-Patch and a Multi-Scale architecture for Nonhomogeneous haze removal from images. We show that DMPHN is better than DMSHN because DMPHN aggregates local features generated from a finer level to coarser level. Moreover, DMPHN is a fast algorithm and can dehaze images from a video sequence in real-time. We also show that DMPHN performs well for Dense Haze Removal. In future, the effectiveness of DMPHN with more levels can be explored for performance improvement, but the addition of more levels to architecture will subject to sacrifice in runtime.

Team	Fidelity		Perceptual quality			Runtime(s)↓	GPU/CPU
	PSNR↑	SSIM↑	LPIPS↓	PI↓	MOS↓		
method1	21.60	0.67	0.363	3.712	3	0.21	v100
method2	21.91	0.69	0.361	3.700	4	0.22	v100
method3	19.25	0.60	0.426	5.061	12	12.88	v100
method4	18.51	0.68	0.308	2.988	12	13.00	n/a
Ours (DMPHN)	*18.24*	*0.65*	*0.329*	*3.051*	*14*	*0.01*	1080
method5	18.70	0.64	0.328	3.114	14	10.43	1080ti
method6	18.67	0.64	0.303	3.211	16	1.64	TitanXP
method7	17.88	0.57	0.378	2.855	16	0.06	n/a
no processing	11.33	0.42	0.582	2.609	20		

Table 2: NTIRE 2020 Nonhomogeneous challenge[5] Leaderboard. Submissions are sorted in ascending order of MOS.
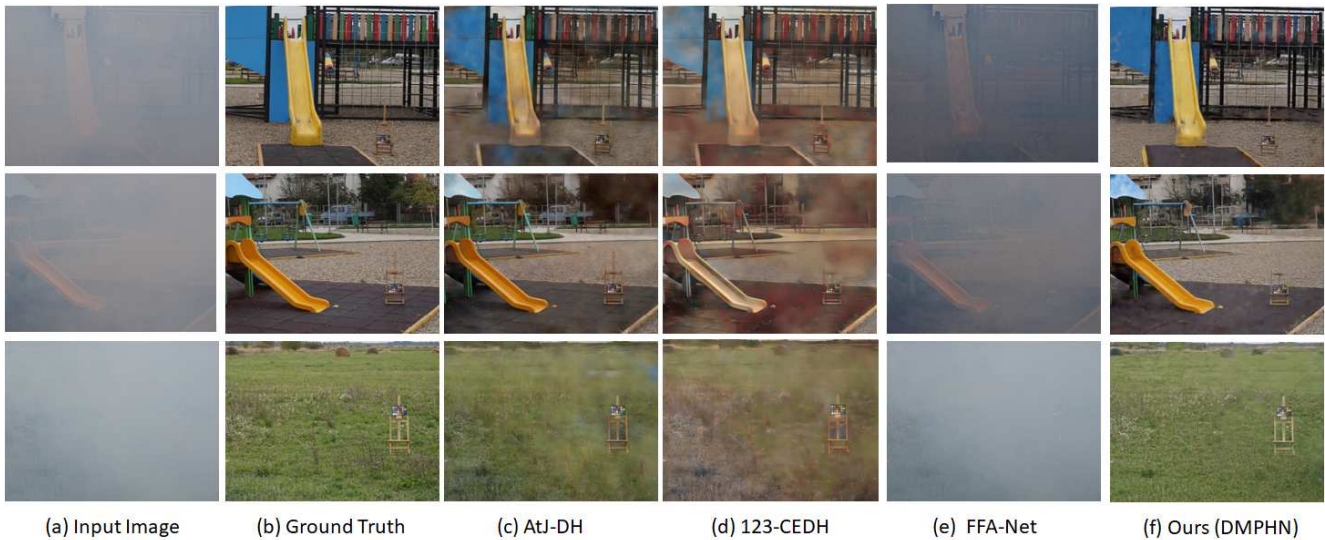

Figure 6: Qualitative results for Dense Haze Removal.

	PSNR	SSIM	Runtime(s)
AtJ-DH[10]	22.54	0.6436	0.0775
123-CEDH[9]	19.63	0.5758	0.0559
FFA-Net[17]	11.93	0.3790	1.7472
Ours(DMPHN)	**23.41**	**0.6669**	**0.0145**

Table 3: Quantitative Comparison on Dense-HAZE[2].

References

[1] C. Ancuti, C.O. Ancuti, R. Timofte, L. Van Gool, and L. Zhang et al. NTIRE 2018 challenge on image dehazing: Methods and results. *IEEE CVPR, NTIRE Workshop*, 2018. 6

[2] 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 *2019 IEEE International Conference on Image Processing (ICIP)*, pages 1014–1018. IEEE, 2019. 6, 7

[3] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. NH-HAZE: An image dehazing benchmark with nonhomogeneous hazy and haze-free images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 4, 6

[4] 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. 6

[5] Codruta O. Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. Ntire 2020 challenge on non-homogeneous dehazing. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 6, 7

[6] C. O. Ancuti, C.Ancuti, R. Timofte, L. Van Gool, and L. Zhang et al. NTIRE 2019 challenge on image dehazing: Methods and results. *IEEE CVPR, NTIRE Workshop*, 2019. 6

[7] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on percep-

tual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018. 6

[8] Zijun Deng, Lei Zhu, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Qing Zhang, Jing Qin, and Pheng-Ann Heng. Deep multi-model fusion for single-image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2453–2462, 2019. 2

[9] Tiantong Guo, Venkateswararao Cherukuri, and Vishal Monga. Dense123'color enhancement dehazing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 5, 6, 7

[10] Tiantong Guo, Xuelu Li, Venkateswararao Cherukuri, and Vishal Monga. Dense scene information estimation network for dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 5, 6, 7

[11] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 4

[12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[13] Yunan Li, Qiguang Miao, Wanli Ouyang, Zhenxin Ma, Huijuan Fang, Chao Dong, and Yining Quan. Lap-net: Level-aware progressive network for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3276–3285, 2019. 2

[14] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Grid-dehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7314–7323, 2019. 2

[15] Yang Liu, Jinshan Pan, Jimmy Ren, and Zhixun Su. Learning deep priors for image dehazing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2492–2500, 2019. 2

[16] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 5

[17] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. *arXiv preprint arXiv:1911.07559*, 2019. 2, 5, 6, 7

[18] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8160–8168, 2019. 2

[19] Prasen Sharma, Priyankar Jain, and Arijit Sur. Scale-aware conditional generative adversarial network for image dehaz-ing. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 2355–2365, 2020. 2

[20] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4

[21] 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. 6

[22] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5978–5986, 2019. 2

[23] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2018. 2

[24] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 6