Ultra-High-Definition Image Dehazing via Multi-Guided Bilateral Learning

Zhuoran Zheng[1,2], Wenqi Ren[2*], Xiaochun Cao[2], Xiaobin Hu[3], Tao Wang[4], Fenglong Song[4], and Xiuyi Jia[1*]
[1]CSE, Nanjing University of Science and Technology [2]SKLOIS, IIE, CAS
[3]Informatic, Technische Universität München [4]Huawei Noah's Ark Lab

Abstract

Convolutional neural networks (CNNs) have achieved significant success in the single image dehazing task. Unfortunately, most existing deep dehazing models have high computational complexity, which hinders their application to high-resolution images, especially for UHD (ultra-high-definition) or 4K resolution images. To address the problem, we propose a novel network capable of real-time dehazing of 4K images on a single GPU, which consists of three deep CNNs. The first CNN extracts haze-relevant features at a reduced resolution of the hazy input and then fits locally-affine models in the bilateral space. Another CNN is used to learn multiple full-resolution guidance maps corresponding to the learned bilateral model. As a result, the feature maps with high-frequency can be reconstructed by multi-guided bilateral upsampling. Finally, the third CNN fuses the high-quality feature maps into a dehazed image. In addition, we create a large-scale 4K image dehazing dataset to support the training and testing of compared models. Experimental results demonstrate that the proposed algorithm performs favorably against the state-of-the-art dehazing approaches on various benchmarks.

1. Introduction

Images captured in outdoor usually suffer from noticeable degradation of contrast and visibility since the light scattering and absorption, especially in foggy and hazy weather conditions. Even on a clear day, the impurities of the aerosol attenuate the reflected lights of distant objects reaching the camera lens [27, 30].

Single image dehazing aims to estimate the sharp image given a hazy input, which is a highly ill-posed problem. Conventional approaches are physically inspired and apply various sharp image priors [3, 18, 31, 44] to regularize the solution space. Most of these methods involve heuristic parameter-tuning and expensive computation. Further, these hand-crafted image priors struggle while applying on different types of real-world images, where the haze is far more complicated than modeled [1].

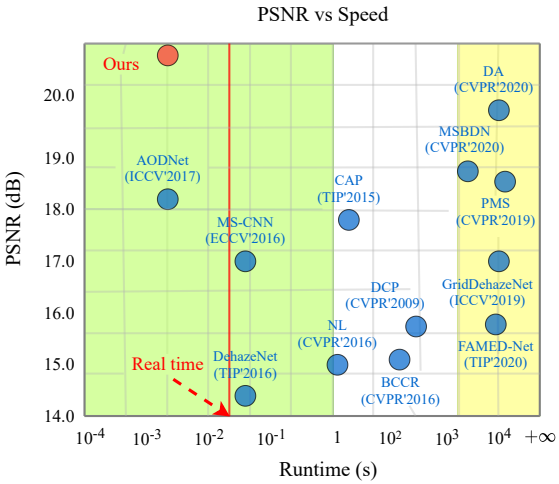

Figure 1. Speed and accuracy trade-off of the state-of-the-art dehazing methods and the proposed algorithm on the 4K dehazing dataset. The red dash line denotes the real-time approach for 4K resolution images at 30 fps. The green region indicates the inference process at the millisecond (ms) level. The yellow region represents the methods that cannot deal with 4K images directly and needs to use the DDU strategy. For example, the maximum resolution can be handled by GridDehazeNet [32] and MSBDN [16] is around 1024×1024, while PMS [14] can only run on images around 640×480. The recent method of DA [40] can dehaze 2K images, i.e., 2560×1440. The proposed method generates dehazed images efficiently and accurately at 4K resolution.

Recently, CNNs have been applied in numerous computer vision problems, including low-level image reconstruction tasks [22, 42, 48, 46, 49], and showed promising results. Therefore, learning-based methods have also been proposed for dehazing. Early algorithms [9, 26, 38] substitute a few modules or steps (e.g., estimating transmission maps or atmospheric lights) in traditional frameworks with learned parameters to make use of external data. More recent researches employ trainable end-to-end networks for dehazing images [15, 20, 24, 28, 37]. Although these approaches achieve state-of-the-art results on single image dehazing task, their models usually have a large size of parameters and are computationally expensive.

*Corresponding authors.

Existing CNN-based methods have two major limitations. First, since learned weights of existing deep models are fixed, these networks need to remove different fog and haze concentrations with the same weights. Therefore, recovering detailed colors and edges from hazy images is still a non-trivial problem. Second, existing models usually implement a highly non-linear mapping from hazy inputs to dehazed results by learning a large number of filters, which inevitably increases the computational consumption. For example, the recent dehazing methods of GridDehazeNet [32], PMS [14], MSBDN [16], and Domain Adaption (DA) [40] cannot directly dehaze 2K or 4K resolution images on an NVIDIA Titan RTX GPU shader with 24G RAM. Although the early light-weight deep models [9, 26, 38] can run on 4K images at the millisecond level, the PSNR results are much lower than the recent work of MSBDN [16], PMS [14], and DA [40] as shown in Figure 1.

Reaching a trade-off between the accuracy and efficiency of a network is a non-trivial task. To achieve this, we propose an efficient and interpretable filtering module in bilateral space [11]. Since haze effect is inherently contrast (or edges) and color degradations [2, 19, 39], we design edge-aware affine modules for different color channels, which adjust the bilateral grid to be applied at different regions and channels. Specially, our algorithm extracts high-resolution and low-resolution features for guidance map learning and affine model fitting, respectively. Then, we employ the multi-guided bilateral interpolating into this array of affine models as a function of position and intensity of each color channel that can be applied to restore high-quality feature maps. Finally, we fuse the multiple high-resolution features into the dehazed result. Our analysis shows that the benefits of the proposed multi-guided bilateral learning for the dehazing task are two-fold: i) It efficiently enables spatially varying restoration since changes in the bilateral grid occur according to the edges and colors in the local region; ii) The proposed model takes less than 8 ms to process a 4K image on a single Titan RTX GPU, which is highly efficient for deployment in practical applications.

The contributions of this paper are as summarized as:

- We propose a multi-guide bilateral learning framework for 4K resolution image dehazing, which can process 4K (3840×2160) resolution images at 125 fps.
- We deploy edge-aware modules that adjust the bilateral affine grid to be applied on multi-guided matrices from different color channels for restoring high-quality features. The method is able to generate more edges and high-frequency details from the hazy input.
- We establish a large and high-quality 4K image dataset for image dehazing. Experimental results on synthetic and real-world images demonstrate the proposed algorithm performs favorably against the state-of-the-art methods on arbitrary spatial input sizes.

2. Related work

Single Image Dehazing. A milestone in single image haze removal was made by Tan [43] and Fattal [17] that can dehaze a single image without multiple inputs or additional information. Since then, numerous methods propose various constrains on the transmission map, atmospheric light, and image to remove haze from a single image by optimizing complex objective functions.

In [23], dark channel prior (DCP) is proposed to estimate the depth and latent image jointly. An image guided depth-edge-aware smoothing algorithm [10] is introduced to refine the transmission map estimated by DCP based on local priors. Zhang et al. [50] propose a maximum reflectance prior for night image dehazing. This method assumes that the local maximum intensities of each color channel of the nighttime haze image are mainly contributed by the ambient illumination. Recently, Berman et al. [7] observe that colors of a sharp image can be well approximated by a few hundred distinct colors, and then propose a dehazing algorithm based on this prior. Although these methods are successful in varying degrees, their performance is limited by the accuracy of the assumptions used in the target scenes.

With the advances in deep learning techniques and the availability of large datasets, recent years have witnessed the increasing popularity of data-driven methods for image dehazing [29, 33, 45, 47, 51]. These methods focus on the deep convolution model to extract haze-relevant features rather than on manual extraction. For instance, DehazeNet proposed in [9] uses a three-layer CNN to estimate the transmission map from a hazy image. Another line of research [25, 26, 36, 39] represents a departure from the conventional atmospheric model-based strategy. Specifically, the estimation of the transmission map and the atmospheric light is explicitly discarded. These methods use end-to-end networks to estimate dehazed images and also reach excellent results. Recently, a multi-scale boost dehazing network (MSDBN) [16] is proposed to directly recover sharp images by incorporating the strengthen-operate-subtract boosting strategy in the decoder of a U-Net. However, we note that most of the CNN-based dehazing methods use a large number of training parameters and result in an expensive runtime (see Figure 1).

Bilateral filtering. Numerous schemes have been proposed for image processing acceleration. The bilateral filter/grid, in particular, has attracted long-term attention in its acceleration [6, 12, 13], which is an edge-aware manipulation of images in the bilateral space [5]. Several methods [6, 41, 52] used a bilateral filter to accelerate the tasks of colorization, depth super-resolution, and semantic segmentation. Surprisingly, the only attempt to learn the bilateral filter for image enhancement we found is [21] that casts the enhancement problem in the bilateral space by the local affine color transformation. However, there are two draw-

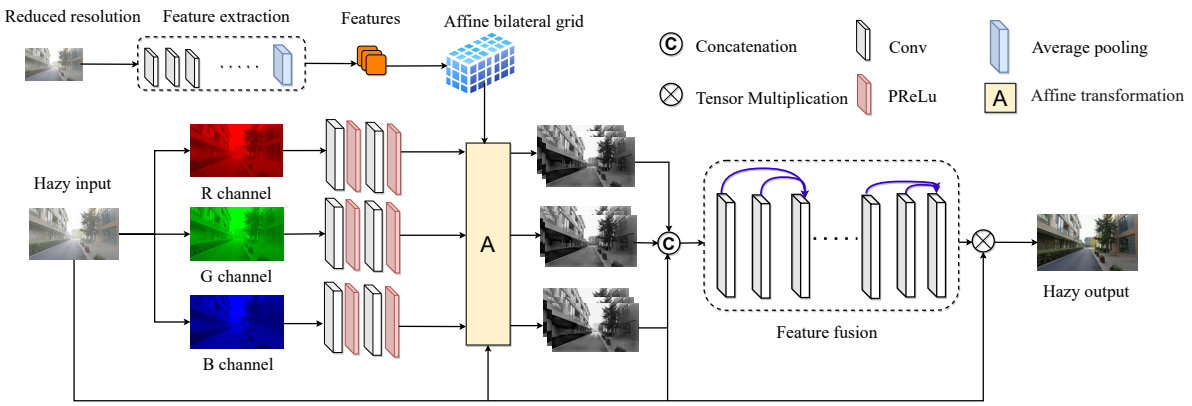

Figure 2. The architecture and configuration of the proposed single image dehazing method, which consists of three parts. The first step starts with a low-resolution coefficient prediction stream that uses a feature extraction block and a downsampling layer to fit an affine bilateral grid. Then we use three guidance matrices and the affine bilateral grid to recover different high-quality features from interpolation. Finally, this feature fusion block combines three high-quality features with the full-resolution input as four heterogenous data to yield the dehazed result. Our proposed algorithm supports dehazing of 4K images at 125 fps on a single Titan RTX GPU shader.

backs in this locally-affine model. First, the extraction of the high-resolution guidance map collapses it directly into a 2D map, and a large amount of color information is lost in the input. Second, [21] uses a linear method of slicing the higher-order features in the application part of the affine model, which has a considerable consumption cost. In contrast, our proposed method makes use of the color information and speed up the slicing operation in [21].

3. Proposed Method

Given a 4K resolution hazy input, our network first reconstructs bilateral coefficients using a feature extraction block on a reduced resolution of the input. Capitalizing on the regressed affine bilateral grid, we can generate high-quality feature maps under the guidance of full-resolution features. Moreover, to provide richer color and edge information, we take all the RGB channels into consideration such that the dehazing network can better recover details. Figure 2 illustrates the architecture of the proposed 4K resolution image dehazing network, which consists of two branches: a low-resolution feature extraction and affine bilateral grid learning (upper stream) and a full-resolution dehazing branch (bottom stream).

3.1. Bilateral Grid Learning

Since the degradation of contrast and visibility of hazy inputs, two pixels across a week edge are close in the spatial dimension. However, these two pixels are distant from the perspective of bilateral filter since their values differ widely in the range dimension. Therefore, we consider predicting an affine bilateral model to restore sharp image structures and edges by fitting a 3D array of affine functions (i.e., fitting different affine functions for different intensity range [11]) on the bilateral space.

First, we reduce the 4K hazy input to a fixed resolution of 910×512 and extract low-level features by two convolutional blocks followed by an average pooling layer as shown in Figure 2. This yields a $32 \times 56 \times 3$ array of the feature map F that contains rich patterns of the input. Then we learn an affine bilateral grid, where each coordinate is augmented with a third dimension (i.e., a function of the pixel intensity of the feature map F), to the scene structures. Under this interpretation, F can be viewed as a $7 \times 8 \times 8$ bilateral grid, where each grid cell contains 12 numbers, one for each coefficient of a 3×4 affine transformation matrix. Specifically, we treat the low-resolution features F as a multi-channel bilateral grid B whose third dimension has been unrolled:

$$B_d[x', y', c'] \leftrightarrow F[x, y, c] \tag{1}$$

where $[x'$=7, y'=8, c'=8$]$ denote the coordinates of the bilateral grid cell, and each cell has d=12 number. In addition, $[x$=32, y=56, c=3$]$ correspond to the height, width, and channels of the feature map F, respectively. This operation is more expressive than conventional bilateral grid splatting which discretizes the degraded input into several intensity bins then box filters the result [12].

3.2. High-Quality Features Reconstruction

Capitalized on the predicted bilateral grid coefficient B by the low-resolution branch, we now need to transfer this information back to the full-resolution space of the original input to produce high-quality dehazed features. To this end, we introduce a bilateral grid slicing operation, which takes guidance maps and the learned bilateral grid to perform a data-dependent lookup in the grid B.

As shown in Figure 3, we use trilinear interpolation to insert grid data into the guidance tensor and employ an additional convolution to obtain high-quality features. First,

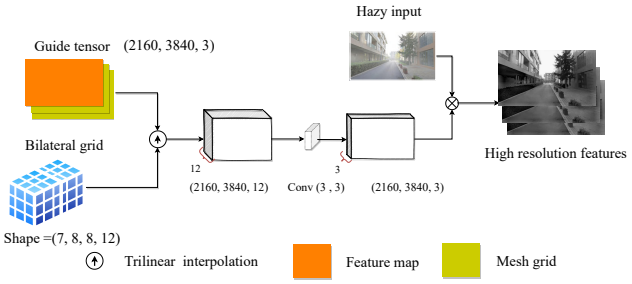

Figure 3. The pipeline of high-quality dehazed features reconstruction.

two maps with coordinate guidance are constructed, and the dimensions are the same as the guidance matrix to stack them into a tensor. Second, we fill the coordinate positions of the pilot tensor with the bilateral grid and use trilinear interpolation to construct a tensor with the same depth (12 channels) as the bilateral grid (we use the PyTorch function of *grid_sample*).

The next step is to compress the tensor by a convolution layer with kernel size of 3×3 and stride of 1 to the depth of 3, which we call the coefficient tensor. Using the convolution layer to compress the high-dimensional tensor is about $2\times$ faster than the linear application operation used in [21] on a GPU shader. Finally, we multiply the coefficient tensor with the full-resolution hazy input to construct new high-quality features.

Multi-guided affine transformation. So far, we have introduced the process of generating and upsampling the bilateral grid of affine coefficients. However, we note that the guidance information obtained by compressing the color space to a single map results in information loss (e.g., color, contrast, and edge). In addition, to reduce the computational cost, we employ a smaller size of the bilateral grid than that proposed in [21]. Therefore, each cell in the bilateral grid would not have enough data to adequately fit an affine transformation for color information.

To address this problem, we propose to process each color channel by different guidance maps as shown in the bottom stream of Figure 2. We use two convolutional layers ($3 \times 3 \times 3$ and $3 \times 3 \times 1$) followed by PReLU to extract the guidance map from each color channel, which is then used to generate high-quality dehazed features for three R, G, and B channels. We demonstrate the effectiveness of multi-guided maps in Figure 4. Take the hazy image in Figure 6(a) as an example. Figure 4(a) shows the edges of the hazy input. With the single-guided map strategy in [21], the recovered edges and structures (Figure 4(b)) are much less than the scheme of using multi-guided maps (Figure 4(a)-(e)). For example, Figure 4(e) reveals more edge and structure information of the house than the one in (b) as shown in the marked green and blue boxes.

To further demonstrate the effectiveness of the proposed multi-guided maps, we randomly select 100 hazy images

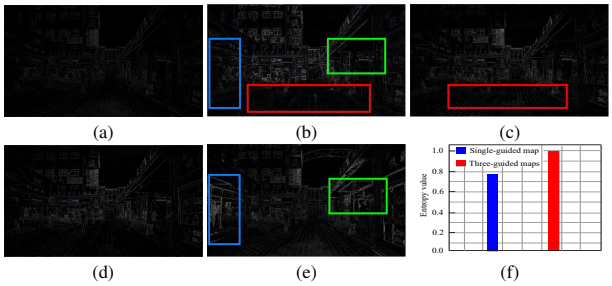

Figure 4. (a) is the edge map of the hazy input in Figure 6(a); (b) is the edge map of the recovered high-quality features produced by single-guided map; (c)-(e) are the edge maps of the high-quality features generated by multi-guided maps from three color channels, respectively. Red, green, and blue boxes indicate that using the multi-guided maps generates richer edge and structure details than the single-guided map. In (f), we compare the entropies of the high-quality features produced by the two methods.

and calculate the entropies[1] of reconstructed features by single- and multi-guided maps, respectively. As shown in Figure 4(f), multiple guidances for the bilateral grid can obtain a larger entropy value, which indicates that more sharp edges and structures are recovered. As a result, our dehazing process follows the learned edges and structural details, thereby regularizing our final predictions towards edge-aware solutions.

3.3. Feature Fusion

To blend effectively the reconstructed high-quality features to preserve the regions with good visibility, we filter their important features by a feature fusion scheme. We concatenate ten multilayer convolution blocks and add skip connections in each block to obtain a coefficient tensor, which acts as a function of the haze concentration. Finally, the predicted coefficient tensor is multiplied by the hazy input to estimate the clean image. We found that using the coefficient tensor can significantly accelerate the convergence of the network and help generate a much clear dehazed image. We also tried to recover the final result by directly using a convolutional layer or linear combination, but we did not obtain improvement.

We optimize the weights and biases of the proposed network by minimizing the L_2 loss on the training set:

$$\mathcal{L} = \frac{1}{D} \sum_{i=1}^{D} \|I_i - J_i\|^2 \qquad (2)$$

where D is number of training images, I is the dehazed result by our network, J is the corresponding ground truth. We also tried the total variation and adversarial losses, but we notice that the L_2 loss is good enough to generate clear and vivid colors in the dehazed results.

[1] A statistical measure [8, 35] for the information content of the image signal, such as structures and details.

(a) GridehazeNet (SDS) (b) GridehazeNet (DDU) (c) PSNR

(d) PMS (SDS) (e) PMS (DDU) (f) SSIM

Figure 5. Qualitative and quantitative comparisons of 4K image processing strategies (SDS and DDU). The SDS scheme cannot consider global information and tends to generate artifacts.

4. Experimental Results

In the section, we evaluate the proposed method by conducting experiments on both synthetic datasets and real-world images. All the results are compared against eleven state-of-the-art dehazing methods: DCP [23], BCCR [34], CAP [53], NLD [7], MSCNN [38], AODNet [26], DehazeNet [9], PMS [14], GridDehazeNet [32], MSBDN [16], and DA [40]. In addition, we conduct ablation studies to demonstrate effectiveness of each module of our network.

4.1. Training data

To train and evaluate the proposed network as well as the comparison methods, we build a new 4K (3840×2160) resolution image dehazing (4KID) dataset, which consists of 10,000 frames hazy/sharp images extracted from 100 video clips at 4K resolution by several different mobile phones. We synthesize hazy images based on the atmospheric scattering model [23]. Following the prototype in [14, 24, 27, 37], we randomly sample atmospheric light conditions $A \in [0.8, 1]$ and scattering coefficient $\beta \in [0.4, 2.0]$ to generate the corresponding hazy images. We further use the translation module in [40] to translate the synthetic hazy images from synthetic domain to real domain, which can improve the generalization of deep models in real cases. Finally, we randomly select 8000 images as the training set. Since the conventional dehazing methods take too long to process a 4K image, we randomly selected 200 images from the remaining images as the test data.

4.2. Implementation Details

The proposed model is implemented in PyTorch and Adam optimizer is used to train the network. We use the resolution 910×512 images with a batch size 6 to train the network. The initial learning rate is set to 0.001. We train the network for 800 epochs in total. At the same time, we include the decay rate in the loss function section.

For DCP [23], we employ different window sizes of 15×15, 45×45, and 65×65. We find that to increase the window size could improve PSNR and brightness of dehazed images, but the time consumption is drastically increased. For instance, the run times are 2.4×10^3 and 5.3×10^3 seconds for the window sizes of 45×45 and 65×65 on a 4K image, respectively. Therefore, we use the window size of 15×15 for DCP [23] model in the paper.

Due to the recent dehazing models of PMS [14], GridDehazeNet [32], MSBDN [16], and DA [40] cannot dehaze 4K images directly, we design two strategies to remove haze from 4K images for these approaches. The first scheme is the downsample-dehazing-upsample (DDU) method which apply dehazing approaches at a low-resolution, then upsample the result. The second one is splitting-dehazing-stitching (SDS), which splits images in patches then stitch the dehazed patch to the full-resolution. For both these two schemes, we resize or split the 4K hazy input to the resolution close to the corresponding maximum size of each dehazing model can handle. We compare these two schemes on our 4KID. As shown in Figure 5, SDS has serious artifacts of color aberration because it does not take into account the global structure of the image. Figure 5 also demonstrates that DDU obtains higher PSNR and SSIM on the test data. Therefore, we adopt the DDU strategy for these four deep models [14, 16, 32, 40]. We fine-tuned deep learning based models of [9, 14, 16, 26, 32, 38, 40] by employing random clipping and resizing. The fine-tuned models can learn both global and local features of the image.

4.3. Evaluation and Results

Quantitative Evaluation. The proposed method is evaluated on two datasets: the 4KID and the O-HAZE dataset [4]. All the deep learning based approaches are fine tuned on our 4KID training data. Sample results for the proposed method and the comparison approaches, on three 4K images and one image from the O-HAZE dataset, are shown in Figures 6 and 7. It can be observed that traditional physics-based methods [7, 23, 34, 53] tend to over-enhance the results, while the recent deep models [16, 26, 32] still remain some haze in the results. However, the dehazed results generated by our algorithm in Figure 6-7(m) are close to the ground truth haze-free images in Figure 6-7(n). The quantitative results on 4KID and O-HAZE datasets reported in Table 1 demonstrate the effectiveness of the proposed method.

Qualitative Evaluation. We then evaluate the proposed algorithm on real-world hazy images. First, we use the real captured 4K resolution hazy images and compare with different state-of-the-art methods. Figure 8 shows the qualitative comparison of results on three challenging real-world images. As shown, DCP and BCCR darken some regions in the dehazed results, CAP and NLD suffer from the color distortions, while the results generated by AOD, DehazeNet, PMS, MSBDN, and DA have some residual haze. In contrast, our algorithm is able to generate realistic colors while better removing haze as shown in the zoomed-in regions of Figure 8(l).

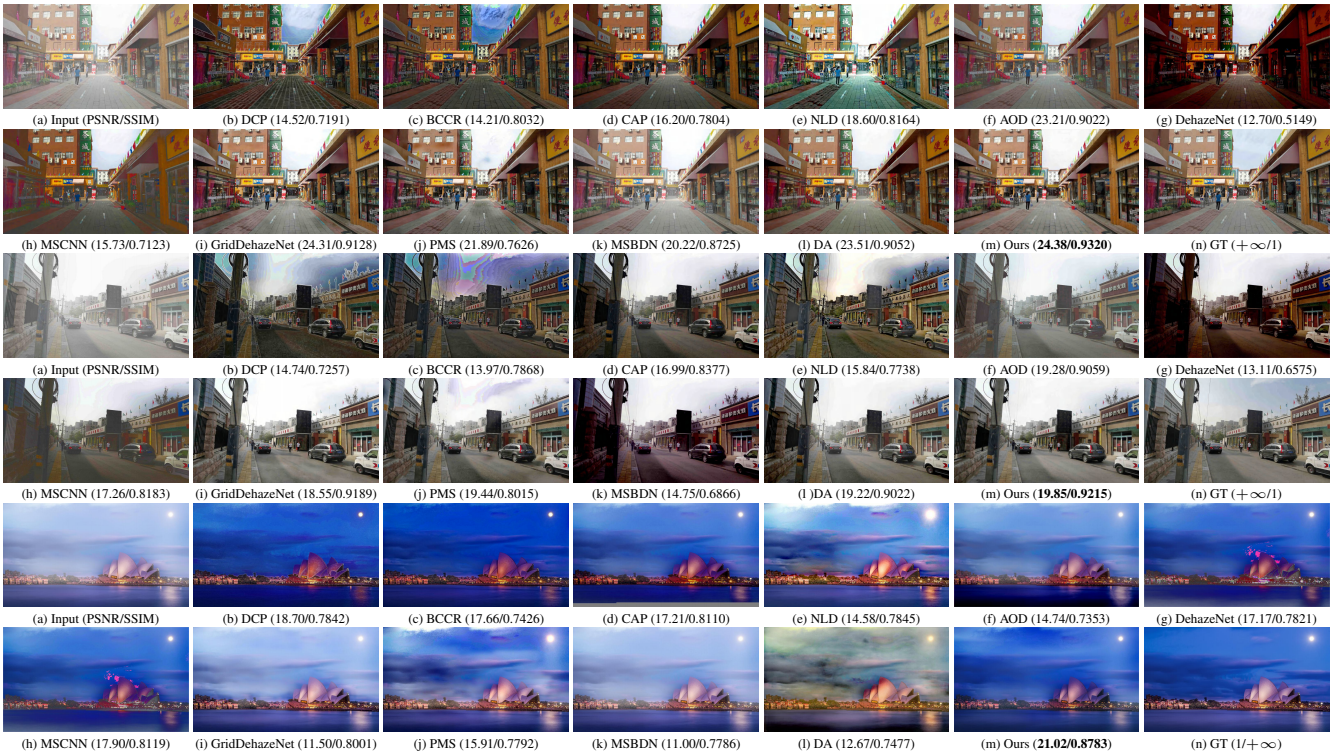

Figure 6. Dehazed results on the 4KID test dataset. Our method obtains better visual quality and recovers more image details compared with other state-of-the-art methods (AODNet [26], MSCNN [38], DehazeNet [9], CAP [53], NLD [7], BCCR [34], DCP [23], MSBDN [16], DA [40], PMS [14] and GridDehazeNet [32]).

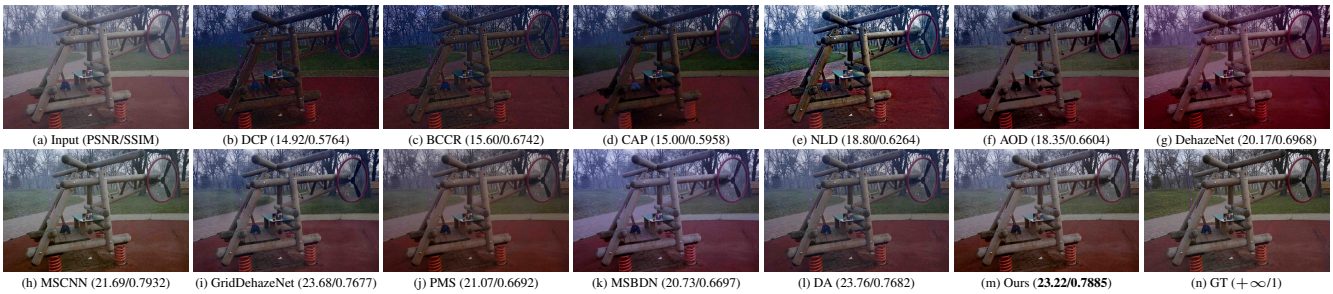

Figure 7. Dehazed results on the O-HAZE dataset [4]. Our method obtains better visual quality and recovers more image details compared with other state-of-the-art methods.

Table 1. Quantitative evaluations on our 4KID (3840×2160) test data and the O-HAZE dataset (This resolution ranges from 1 megapixels (947 ×1286) to 19 megapixels (3612×5456)) [4] in terms of PSNR, SSIM, and run time.

		Input	CAP [53]	NLD [7]	DCP [23]	BCCR [34]	DehazeNet [9]	AOD [26]	MSCNN [38]	PMS [14]	Griddehaze [32]	DA [40]	MSBDN [16]	Ours
4KID	PSNR	11.86	17.67	15.15	15.32	15.01	14.02	17.86	16.41	19.07	16.85	19.26	19.04	**20.56**
	SSIM	0.7421	0.8361	0.7167	0.7511	0.7882	0.7644	0.8519	0.8354	0.7595	0.8752	0.8341	0.7912	**0.8823**
	Time	-	23s	286s	414s	55s	97ms	26ms	39ms	-	-	-	-	**9ms**
O-HAZE	PSNR	13.11	14.55	17.57	16.57	15.01	17.57	15.10	16.63	17.78	16.66	16.86	17.76	**18.43**
	SSIM	0.5643	0.5674	0.7519	0.7350	0.78826	0.7676	0.7448	0.7650	0.7764	0.7436	0.7436	0.7989	**0.8138**
	Time	-	39s	506s	617s	268s	105ms	34ms	47ms	-	-	-	-	**16ms**

In addition to 4K images, we also evaluate our method on several low resolution hazy images downloaded from other public databases. The dehazed results are shown in Figure 9. It can be seen that the outputs from DCP [23] have some halo artifact, while the results by CAP [53], GridDehazeNet [32], and DA [40] contain some color distortions. Contrarily, our method generates better dehazed results by removing the haze and yielding realistic colors effectively.

4.4. Ablation Study

To demonstrate the effectiveness of each module introduced in the proposed network, we perform an ablation study involving the following two experiments:

i) w/o guidance: we remove the low-resolution stream and directly regress the final output without using the bilateral grid and the affine transformation;

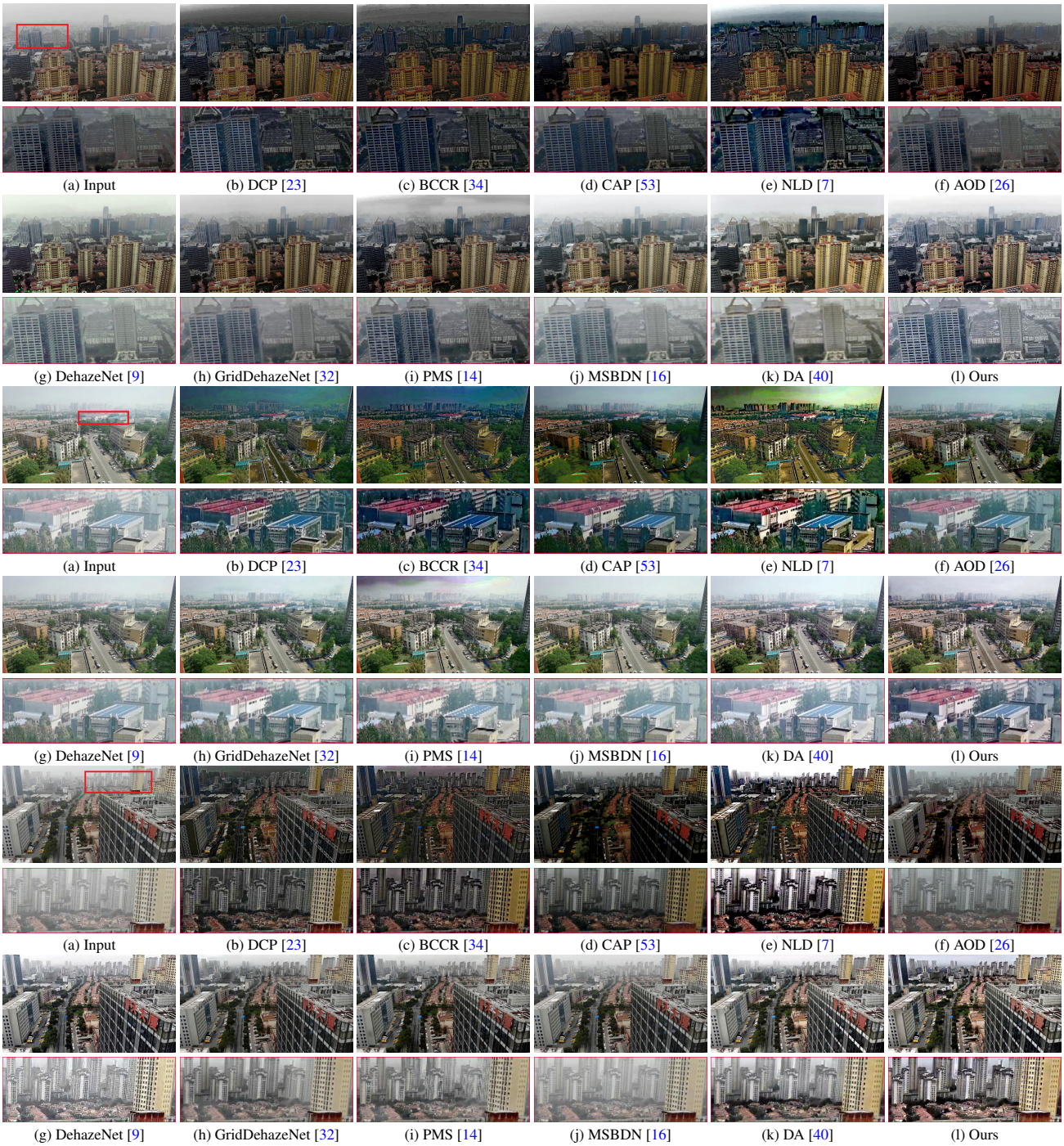

Figure 8. Dehazed results on real captured 4K images. The proposed method generates much clear images.

Table 2. Effectiveness of the multiple guidance scheme.

	w/o guidance	Single guidance	Ours
PSNR	18.29	20.80	**21.52**
SSIM	0.8407	0.8428	**0.8823**

ii) Single guidance: we only use a single guidance matrix to produce a high-resolution feature, where the single guidance matrix is compressed from the original high-resolution image by convolution layers.

Table 2 compares our method against these two baselines on the 4KID dataset. Corresponding visual comparisons are shown in the Figure 10. As observed, without using guidance tends to generate dark result, while the result by using single guidance has some remaining haze. Both the quantitative and qualitative results demonstrate that the proposed multi-guided bilateral grid has a significant improvement than all two baselines.

(a) Input (b) CAP [53] (c) DCP [23] (d) GridDehazeNet [32] (e) DA [40] (f) Our

Figure 9. Dehazed results on low-resolution hazy images. The proposed method recovers high-quality images with clearer details.

(a) Input (PSNR/SSIM) (c) w/o guidance (17.42/0.84) (b) Single guidance (21.02/0.88) (d) Ours (**21.39/0.93**) (e) Ground truth ($+\infty$/1)

Figure 10. Effectiveness of the multiple guidance maps.

4.5. Run Time

We evaluate all the deep models on the same machine with an Intel(R) Xeon(R) CPU and an NVIDIA Titan RTX GPU. The run time is only the processing time of the GPU without considering the I/O operations. The average run times for the 4KID and O-HAZE datasets are shown in Table 1. The conventional methods of [7, 23, 34, 53] solve complex energy functions which inevitably increases the computational cost. The early dehazing approaches of [9, 26, 38] perform faster than conventional methods but still have less efficiency than ours. Since the methods of [14, 16, 32, 40] cannot directly handle 4K images, we do not show their run times in Table 1.

5. Conclusion

In this paper, we propose a ultra-high-definition image dehazing method via multi-guided bilateral learning. The key to our method is using deep CNN to build an affine bilateral grid, which is an effective feature storage containers that maintain detailed edges and textures in the image. At the same time, we establish multiple guidance matrices to assist the affine bilateral grid to restore high-quality features, providing rich color and texture information for image dehazing. Quantitative and qualitative results show that the proposed network performs favorably against the state-of-the-art dehazing methods in terms of accuracy and inference speed (at 125 fps), and can generate visually-pleasing results on real-world 4K hazy images.

Acknowledgment. This work is supported by the National Key Technology R&D Program under Grant 2019QY(Y)0207, National Natural Science Foundation of China (No. 61802403, 61773208, 62025604, U1803264), Beijing Nova Program (No. Z201100006820074), Youth Innovation Promotion Association CAS, the Natural Science Foundation of Jiangsu Province (BK20191287), the Fundamental Research Funds for the Central Universities (30920021131).

References

[1] Cosmin Ancuti, Codruta O Ancuti, and Radu Timofte. Ntire 2018 challenge on image dehazing: Methods and results. In *CVPRW*, 2018. 1

[2] Codruta Orniana Ancuti and Cosmin Ancuti. Single image dehazing by multi-scale fusion. *IEEE TIP*, 22(8):3271–3282, 2013. 2

[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 *ACCV*, 2010. 1

[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 *CVPR*, 2018. 5, 6

[5] Jonathan T Barron, Andrew Adams, YiChang Shih, and Carlos Hernández. Fast bilateral-space stereo for synthetic defocus. In *CVPR*, 2015. 2

[6] Jonathan T. Barron and Ben Poole. The fast bilateral solver. In *ECCV*, 2016. 2

[7] Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *CVPR*, 2016. 2, 5, 6, 7, 8

[8] Christopher M. Bishop and Nasser M. Nasrabadi. *Pattern Recognition and Machine Learning*. *J. Electronic Imaging*, 16(4):049901, 2007. 4

[9] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE TIP*, 25(11):5187–5198, 2016. 1, 2, 5, 6, 7, 8

[10] Chen Chen, Minh N Do, and Jue Wang. Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In *ECCV*, 2016. 2

[11] Jiawen Chen, Andrew Adams, Neal Wadhwa, and Samuel W. Hasinoff. Bilateral guided upsampling. *ACM TOG*, 35(6):203:1–203:8, 2016. 2, 3

[12] Jiawen Chen, Sylvain Paris, and Frédo Durand. Real-time edge-aware image processing with the bilateral grid. *ACM TOG*, 26(3):103–es, 2007. 2, 3

[13] Qifeng Chen, Jia Xu, and Vladlen Koltun. Fast image processing with fully-convolutional networks. In *ICCV*, 2017. 2

[14] Wei-Ting Chen, Jian-Jiun Ding, and Sy-Yen Kuo. Pms-net: Robust haze removal based on patch map for single images. In *CVPR*, 2019. 1, 2, 5, 6, 7, 8

[15] 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 *ICCV*, 2019. 1

[16] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *CVPR*, 2020. 1, 2, 5, 6, 7, 8

[17] Raanan Fattal. Single image dehazing. *ACM TOG*, 27(3):1–9, 2008. 2

[18] Raanan Fattal. Dehazing using color-lines. *ACM TOG*, 34(1):1–14, 2014. 1

[19] Adrian Galdran, Aitor Alvarez-Gila, Alessandro Bria, Javier Vazquez-Corral, and Marcelo Bertalmío. On the duality between retinex and image dehazing. In *CVPR*, 2018. 2

[20] Yossi Gandelsman, Assaf Shocher, and Michal Irani. "double-dip": Unsupervised image decomposition via coupled deep-image-priors. In *CVPR*, 2019. 1

[21] Michaël Gharbi, Jiawen Chen, Jonathan T. Barron, Samuel W. Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM TOG*, 36(4):118:1–118:12, 2017. 2, 3, 4

[22] Yuanbiao Gou, Boyun Li, Zitao Liu, Songfan Yang, and Xi Peng. Clearer: Multi-scale neural architecture search for image restoration. In *NeurIPS*, 2020. 1

[23] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *CVPR*, 2009. 2, 5, 6, 7, 8

[24] Ming Hong, Yuan Xie, Cuihua Li, and Yanyun Qu. Distilling image dehazing with heterogeneous task imitation. In *CVPR*, 2020. 1, 5

[25] Aupendu Kar, Sobhan Kanti Dhara, Debashis Sen, and Prabir Kumar Biswas. Transmission map and atmospheric light guided iterative updater network for single image dehazing. In *CVPR*, 2020. 2

[26] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *ICCV*, 2017. 1, 2, 5, 6, 7, 8

[27] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE TIP*, 28(1):492–505, 2018. 1, 5

[28] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In *CVPR*, 2018. 1

[29] 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 *ICCV*, 2019. 2

[30] Yu Li, Shaodi You, Michael S Brown, and Robby T Tan. Haze visibility enhancement: A survey and quantitative benchmarking. *CVIU*, 165:1–16, 2017. 1

[31] Zhuwen Li, Ping Tan, Robby T Tan, Danping Zou, Steven Zhiying Zhou, and Loong-Fah Cheong. Simultaneous video defogging and stereo reconstruction. In *CVPR*, 2015. 1

[32] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *ICCV*, 2019. 1, 2, 5, 6, 7, 8

[33] Yang Liu, Jinshan Pan, Jimmy Ren, and Zhixun Su. Learning deep priors for image dehazing. In *ICCV*, 2019. 2

[34] Gaofeng Meng, Ying Wang, Jiangyong Duan, Shiming Xiang, and Chunhong Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *ICCV*, 2013. 5, 6, 7, 8

[35] Yi Niu, Xiaolin Wu, and Guangming Shi. Image enhancement by entropy maximization and quantization resolution upconversion. *IEEE TIP*, 25(10):4815–4828, 2016. 4

[36] Yanwei Pang, Jing Nie, Jin Xie, Jungong Han, and Xuelong Li. Bidnet: Binocular image dehazing without explicit disparity estimation. In *CVPR*, 2020. 2

[37] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *CVPR*, 2019. 1, 5

[38] Wenqi Ren, Si Liu, Hua Zhang, Jin-shan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *ECCV*, 2016. 1, 2, 5, 6, 8

[39] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *CVPR*, 2018. 2

[40] Yuanjie Shao, Lerenhan Li, Wenqi Ren, Changxin Gao, and Nong Sang. Domain adaptation for image dehazing. In *CVPR*, 2020. 1, 2, 5, 6, 7, 8

[41] Hang Su, Varun Jampani, Deqing Sun, Orazio Gallo, Erik Learned-Miller, and Jan Kautz. Pixel-adaptive convolutional neural networks. In *CVPR*, 2019. 2

[42] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In *CVPR*, 2020. 1

[43] Robby T Tan. Visibility in bad weather from a single image. In *CVPR*, 2008. 2

[44] Ketan Tang, Jianchao Yang, and Jue Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *CVPR*, 2014. 1

[45] Dong Yang and Jian Sun. Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In *ECCV*, 2018. 2

[46] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE TPAMI*, abs/2009.14468, 2020. 1

[47] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *CVPR*, 2018. 2

[48] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*, 2018. 1

[49] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE TCSVT*, 30(11):3943–3956, 2019. 1

[50] Jing Zhang, Yang Cao, Shuai Fang, Yu Kang, and Chang Wen Chen. Fast haze removal for nighttime image using maximum reflectance prior. In *CVPR*, 2017. 2

[51] Jing Zhang and Dacheng Tao. Famed-net: A fast and accurate multi-scale end-to-end dehazing network. *IEEE TIP*, 29:72–84, 2020. 2

[52] Yinda Zhang and Thomas Funkhouser. Deep depth completion of a single rgb-d image. In *CVPR*, 2018. 2

[53] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE TIP*, 24(11):3522–3533, 2015. 5, 6, 7, 8