Gated Context Aggregation Network for Image Dehazing and Deraining

Dongdong Chen[1],   Mingming He[2],   Qingnan Fan[3],   Jing Liao[4]
Liheng Zhang[5],   Dongdong Hou[1],   Lu Yuan[6],   Gang Hua[6]

[1]University of Science and Technology of China,   [2] Hong Kong University of Science and Technology
[3]Shandong University,   [4]City University of Hong Kong
[5]University of Central Florida,   [6]Microsoft Cloud and AI

cd722522@mail.ustc.edu.cn,   hmm.lillian@gmail.com,   fqnchina@gmail.com,
liaojing8871@gmail.com,   lihengzhang1993@knights.ucf.edu,
houdd@mail.ustc.edu.cn,   {luyuan, ganghua}@microsoft.com

Abstract

Image dehazing aims to recover the uncorrupted content from a hazy image. Instead of leveraging traditional low-level or handcrafted image priors as the restoration constraints, e.g., dark channels and increased contrast, we propose an end-to-end gated context aggregation network to directly restore the final haze-free image. In this network, we adopt the latest smoothed dilation technique to help remove the gridding artifacts caused by the widely-used dilated convolution with negligible extra parameters, and leverage a gated sub-network to fuse the features from different levels. Extensive experiments demonstrate that our method can surpass previous state-of-the-art methods by a large margin both quantitatively and qualitatively. In addition, to demonstrate the generality of the proposed method, we further apply it to the image deraining task, which also achieves the state-of-the-art performance.

1. Introduction

Due to the existence of turbid medium (e.g., dusk, smoke, and other particles) in the atmosphere, images taken in such atmospheric phenomena are subject to visible quality degradation, such as contrast and saturation loss. Taking these degraded images as input, many vision-based systems, originally designed with the assumption of clean capture environments, may be easily troubled with drastic performance decrease. Given that, image dehazing has been extensively studied to restore the clean image from the corrupted input, to serve as the preprocessing step of the aforementioned systems.

In this literature, the hazing processing is often represented with the physical corruption model:

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

where $\mathbf{I}(x)$ and $\mathbf{J}(x)$ are the degraded hazy image and the target haze-free scene radiance respectively. $\mathbf{A}$ is the global atmospheric light, and $t(x)$ is the medium transmission map, which is dependent on the unknown depth information. Most previous dehazing methods first estimate the transmission map $t(x)$ or the atmospheric light $\mathbf{A}$, then try to recover the final clean image $\mathbf{J}(x)$. But the first step is a very challenging problem because both the transmission map $t(x)$ and the atmospheric light $\mathbf{A}$ are often unknown in the real scenarios.

To compensate for the lost information during the corruption procedure, many traditional methods [2, 16, 17, 29, 30, 46] leverage some image priors and visual cues to estimate the transmission maps and atmospheric light. For example, [16] maximizes the local contrast of the target image by using the prior that the contrast of degraded images is often drastically decreased. [17] proposes the dark channel prior based on the assumption that image patches of outdoor haze free images often have low-intensity values. [2] relies on the assumption that haze-free image colors are well approximated by a few hundred distinct colors and proposes a non-local prior-based dehazing algorithm. However, these priors do not always hold, so they may not work well in certain real cases.

With the latest advances of deep learning, many CNN-based methods [1, 3, 31, 22, 32, 42] are proposed by leveraging a large scale training datasets. Compared to traditional methods as described above, CNN-based methods attempt to directly regress the intermediate transmission map or the final clean image, and achieve superior performance and robustness. [3] presents an end-to-end network to estimate the intermediate transmission map. [22] reformulates the atmospheric scattering model to predict the final clean image through a light-weight CNN. [32] creates three different derived input images from the

original hazy image and fuses the dehazed results out of these derived inputs. [42] incorporates the physical model in Equation (1) into the network design and uses two sub-networks to regress the transmission map and atmospheric light respectively.

In this paper, we propose a new end-to-end gated context aggregation network (denoted as "**GCANet**") for image dehazing. Since dilated convolution is widely used to aggregate context information for its effectiveness without sacrificing the spatial resolution [41, 25, 36, 15, 9], we also adopt it to help obtain more accurate restoration results by covering more neighbor pixels. However, the original dilated convolution will produce so-called "gridding artifacts" [36, 15], because adjacent units in the output are computed from completely separate sets in the input when the dilation rate is larger than one. Recently, [37] analyzes the dilation convolution in a compositional way and proposes to smooth the dilated convolution, which can greatly reduce this gridding artifacts. Hence, we also incorporate this idea in our context aggregation network. As demonstrated in [42, 27], fusing different levels of features is often beneficial for both low-level and high-level tasks. Inspired by it, we further propose a gated sub-network to determine the importance of different levels and fuse them based on their corresponding importance weights. [32] also uses a gated fusion module in their network, but they directly fuse the dehazing results of different derived input images rather than the intermediate features.

To validate the effectiveness of the proposed **GCANet**, we compare it with previous state-of-the-art methods on the recent dehazing benchmark dataset RESIDE [23]. Experiments demonstrate that our **GCANet** outperforms all the previous methods both qualitatively and quantitatively by a large margin. Furthermore, we conduct comprehensive ablation studies to understand the importance of each component. To show the generality of the proposed **GCANet**, we have also applied it to the image deraining task, which can also obtain superior performance over previous state-of-the-art image deraining methods.

To summarize, our contributions are three-fold as below:

- We propose a new end-to-end gated context aggregation network **GCANet** for image dehazing, in which the smoothed dilated convolution is used to avoid the gridding artifacts and a gated subnetwork is applied to fuse the features of different levels.

- Experiments show that **GCANet** can obtain much better performance than all the previous state-of-the-art image dehazing methods both qualitatively and quantitatively. We also provide comprehensive ablation studies to validate the importance and necessity of each component.

- We further apply our proposed **GCANet** to the image deraining task, which also outperforms previous state-of-the-art image deraining methods and demonstrates its generality.

The remainder of the paper is organized as follows. We will first summarize related work in Section 2, then give our main technical details in Section 3. Finally, we will provide comprehensive experiments results and ablation studies in Section 4 and conclude in Section 6.

2. Related Work

Single image dehazing is the inverse recovery procedure of the physical corruption procedure defined in Equation (1), which is a highly ill-posed problem because of the unknown transmission map and global atmospheric light. In the previous several decades, many different image dehazing methods are proposed to tackle this challenging problem, which can be roughly divided into traditional prior-based methods and modern learning-based methods. The most significant difference between these two types is that the image priors are handcrafted in the former type but are learned automatically in the latter type.

In the traditional prior-based methods, many different image statistics priors are leveraged as extra constraints to compensate for the information loss during the corruption procedure. For example, [11] propose a physically grounded method by estimating the albedo of the scene. [17, 38, 39] discover and improve the effective dark channel prior to calculate the intermediate transmission map more reliably. [34] use Markov Random Field to maximize the local contrast of an image by assuming that the local contrast of a clear image is higher than that of a hazy image. Based on the observation that small image patches typically exhibit a one-dimensional distribution in the RGB color space, [12] recently propose a color-line method for image dehazing and [2] propose a non-local path prior to characterize the clean images. These dedicatedly handcrafted priors , however, hold for some cases, but they are not always robust to handle all the cases.

Recently, learning-based methods are proposed for image dehazing by leveraging the large-scale datasets and the powerful parallelism of GPU. In these type of methods, the image priors are automatically learned from the training dataset by the neural network and saved in the network weights. Their main differences typically lie in the learning targets and the detailed network structures. [3, 31] propose an end-to-end CNN network and multi-scale network respectively to predict the intermediate transmission maps. However, inaccuracies in the estimation of the transmission map always lead to low-quality dehazed results. [22] encode the transmission map and the atmospheric light into one variable, and then use a lightweight network to predict it. [42] design two different sub-networks for the prediction

of the transmission map and the atmospheric light by following the physical model defined in Equation (1). We propose an end-to-end gated context aggregation network for image dehazing but different from these methods, our proposed **GCANet** is designed to directly regress the residue between the hazy image and the target clean image. Moreover, our network structure definitely distinguish from the previous ones, which is quite lightweight but can achieve much better results than all the previous methods.

3. Method

In this section, we will introduce the architecture of the proposed gated context aggregation network **GCANet**. As shown in Figure 1, given a hazy input image, we first encode it into feature maps by the encoder part, then enhance them by aggregating more context information and fusing the features of different levels without downsampling. Specifically, the smoothed dilated convolution and an extra gate sub-network are leveraged. The enhanced feature maps will be finally decoded back to the original image space to get the target haze residue. By adding it onto the input hazy image, we will get the final haze free image.

Smoothed Dilated Convolution Modern image classification networks [21, 33, 18] often integrate multi-scale contextual information via successive pooling and subsampling layers that reduce resolution until a global prediction is obtained. However, for dense prediction tasks like segmentation, the contradiction is the required multi-scale contextual reasoning and the lost spatial resolution information during downsampling. To solve this problem, [41] proposes a new dilated convolutional layer, which supports exponential expansion of the receptive field without loss of resolution or coverage. In the one-dimension case, given a 1-D input f, the output of the regular convolutional layer w with kernel size k is:

$$(f \otimes w)(i) = \sum_{j=1}^{k} f[i+j]w[j] \qquad (2)$$

where one output point cover total k input points, so the receptive field is k. But for the dilated convolution, it can be viewed as "convolution with a dilated filter", which can be represented as:

$$(f \otimes_r w)(i) = \sum_{j=1}^{k} f[i+r*j]w[j] \qquad (3)$$

where r is the dilation rate, and the dilated convolution will degenerate to regular convolution when $r = 1$. To understand the dilated convolution in an intuitive way, we can view it as inserting $r - 1$ zeros between two adjacent

weights of w. In this way, the dilated convolution can increase the original receptive field from k to $r*(k-1)+1$ without reducing the resolution.

Despite of the effectiveness of the dilated convolution, it will produce the so-called gridding artifacts, which is also noticed in previous papers [36, 15]. To understand this issue more clearly, a very recent work [37] analyzes the dilated convolution in a compositional way. The illustration of gridding artifacts is shown in Figure 2, where the case of one dilated convolutional layers with $r = 2, k = 3$ is analyzed. Considering the four neighbor pixels of the next layer, they and their dependent units in the previous layer are marked with four different colors respectively. We can easily find that these four neighor pixels are related to totally different sets of previous units in the previous layer. In other words, there is no dependency among the input units or the output units in the dilated convolution. This is why it will potentially cause the inconsistencies, i.e. gridding artifacts.

To alleviate it, [37] proposes to add interaction among the input units before dilated convolution or output units after dilated convolution by adding an extra convolutional layer of kernel size $(2r - 1)$. In this paper, we choose to add the dependency of input units by default. Need to note that, [37] adopts a separable and shared convolution as the extra convolutional layer rather than the vanilla one. "Separable" means the separable convolution idea from [8], while "shared" means the convolution weights are shared for all the channels. In this way, this special convolutional layer has a constant parameter size $(2r - 1)^2$, which is independent of the feature channel number. Figure 2 is one illustration of smoothed dilated convolution.

Gated Fusion Sub-network As shown in [27, 42], fusing the features from different levels is often beneficial both for low-level and high-level tasks. To implement this idea, [27] uses the feature pyramids to fuse high-level semantic feature maps at all scales, and [42] leverages the densely connected networks. In this paper, we adopt a different way by incorporation of an extra gated fusion sub-network $\mathcal{G}$. Specifically, we first extract the feature maps from different levels F_l, F_m, F_h, and feed them into the gated fusion sub-network. The output of the gated fusion sub-network are three different importance weights $(\mathcal{M}_l, \mathcal{M}_m, \mathcal{M}_h)$, which correspond to each feature level respectively. Finally, these three features maps F_l, F_m, F_h from different levels are linearly combined with the regressed importance weights.

$$\begin{aligned}(\mathcal{M}_l, \mathcal{M}_m, \mathcal{M}_h) &= \mathcal{G}(F_l, F_m, F_h) \\ F_o &= \mathcal{M}_l * F_l + \mathcal{M}_m * F_m + \mathcal{M}_h * F_h\end{aligned} \qquad (4)$$

The combined feature map F_o will be further fed into the decoder to get the target haze residue. In this paper, our gated fusion sub-network consists of only one convolutional
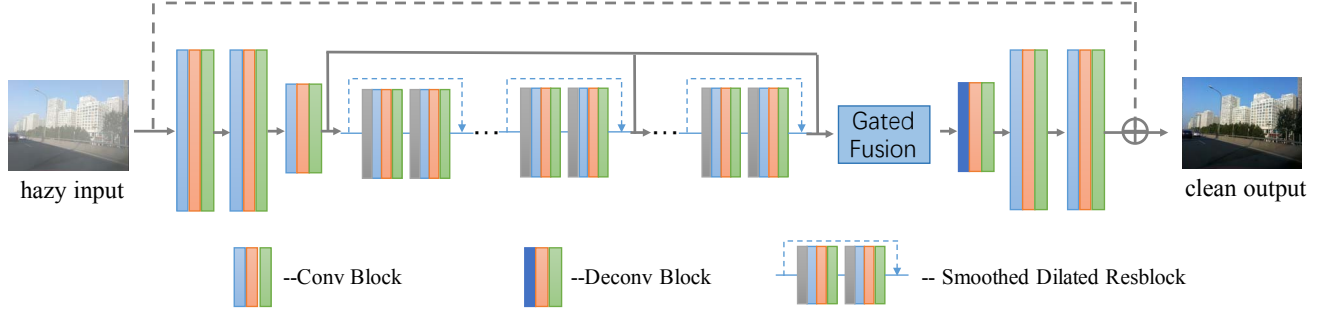
Figure 1. The overall network structure of the proposed **GCANet**, which follows a basic auto-encoder structure. It consists of three convolution blocks as the encoder part, and one deconvolution block and two convolution blocks as the decoder part. Several smoothed dilated resblocks are inserted between them to aggregate context information without gridding artifacts. To fuse the features from different levels, an extra gate fusion sub-network is leveraged. During the runtime, the **GCANet** will predict the residue between the target clean image and the hazy input image in an end-to-end way.
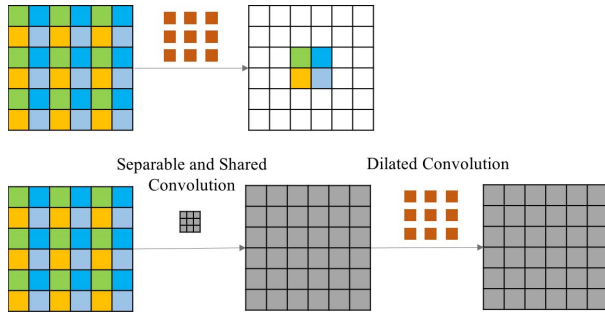

Figure 2. The illustration of gridding artifacts of dilated convolution and the proposed smoothed dilated convolution in [37]: the four different points in next layer i are indicated by different colors, it can be seen that they are related to completely different sets of units of previous layer, which will potential cause the gridding artifacts. By contrast, the smoothed dilated convolution, which adds the dependency among the input units with an extra separable and shared convolutional layer before the dilated convolution.

layer with kernel size 3x3, whose input is the concatenation of F_l, F_m, F_h and output channel number is 3.

Network Structure Following the similar network design principle in [20, 10, 9], our overall network structure are also designed as a simple auto-encoder, where seven residual blocks are inserted between the encoder and decoder to enhance its learning capacity. Specifically, three convolutional layers are first used to encode the input hazy image into the feature maps as the encoder part, where only the last convolutional layer downsamples the feature maps by 1/2 once. Symmetrically, one deconvolutional layer with stride 1/2 is used to upsample the feature map to the original resolution in the decoder part, then the following two convolutional layers convert the feature maps back to the image space to get the final target haze residue. For the intermediate residual blocks, we call them "Smoothed Dilated Resblock", because we

have replaced all the original regular convolutional layers with the aforementioned smoothed dilated convolutional layers. The dilation rates of these seven residual blocks are setted as $(2, 2, 2, 4, 4, 4, 1)$ respectively. To obtain a good tradeoff between the performance and runtime, we set the channel number of all the intermediate convolutional layers as 64. Note that except for the last convolutional layer and every extra separable and shared convolutional layer in the smoothed dilated convolution layer, we put an instance normalization layer [35] and ReLU layer after each convolutional layer. In the experiment part, we will show instance normalization is more suitable than batch normalization for the image dehazing task.

As demonstrated in [10, 9], besides the input image, pre-calculating the edge of the input image and feeding them into the network as the auxiliary information is very helpful to the network learning. Hence, by default, we also adopt this simple idea and concatenate the pre-calculated edge with the input hazy image along the channel dimension as the final inputs of **GCANet**.

Loss Function In previous learning-based image dehazing methods [3, 31, 22, 24, 42, 44], the simple Mean Square Error loss is adopted. Following the same strategy, we also use this simple loss by default. But different from these methods, our learning target is the residue between the haze free image and the input hazy one:

$$
\begin{aligned}
r &= J - I \\
\hat{r} &= GCANet(I) \\
\mathcal{L} &= \|\hat{r} - r\|^2
\end{aligned}
\tag{5}
$$

where r and $\hat{r}$ are the ground truth and predicted haze residue respectively. During runtime, we will add $\hat{r}$ onto the input hazy image to get the final predicted haze free image. Need to emphasize that designing better loss function is not the focus of this paper, but our proposed **GCANet** should

1378

be able to generalize to better designed losses. For example, [24, 42, 44] find the perceptual loss [20] and GAN loss can improve the final dehazing results. However, even only with the above simple loss, our method can still achieve the state-of-the-art performance.

4. Experiments

Implementation Details For experiments, we first validate the effectiveness of the proposed **GCANet** on the image dehazing task, then demonstrate its generality by further applying it to image deraining task. To train these two tasks, we all directly adopt the available benchmark datasets both for training and evaluation. For each task, we compare our method with many previous state-of-the-art methods. Without losing generality, we use almost the same training strategy for these two tasks. By default, the whole network is trained for 100 epochs with the Adam optimizer. The default initial learning rate is set to 0.01 and decayed by 0.1 for every 40 epochs. All the experiments are trained with the default batch size to 12 on 4 GPUs.

Dataset Setup For the image hazing task, we find most previous state-of-the-art methods leverage available depth datasets to synthesize their own hazy datasets based on the physical corruption model in Equation (1), and conduct evaluation only on these specific datasets. Direct comparisons on these datasets are not fair. Recently, [23] proposes a image dehazing benchmark RESIDE, which consists of large-scale training and testing hazy image pairs synthesized from depth and stereo datasets. To compare with state-of-the-art methods, they use many different evaluation metrics and conduct comprehensive comparisons among them. Although their test dataset consists of both indoor and outdoor images, they only report the quantitative results for the indoor parts. Following their strategy, we also compare our method on indoor dataset quantitatively and outdoor dataset qualitatively.

Similar to image hazing, there also exist several different large-scale synthetic datasets for image deraining. Most recently, [43] has developed a new dataset containing raining density labels (e.g. light, medium and heavy) for density-aware image deraining. Although we do not need the rain-density label information in our method, we still adopt this dataset for fair comparison. In this dataset, a total of 12000 training rainy images are synthesized with different orientations and scales with Photoshop.

Quantitative and Qualitative Evaluation for image dehazing In this part, we will compare our method with previous state-of-the-art image dehazing methods both quantitatively and qualitatively.

As shown in Table 1, six different state-of-the-art methods are used for quantitative evaluation: DCP[17], CAP [46], GRM [4], AOD-Net [22], DehazeNet [3], and GFN [32]. Among them, the first three are traditional prior-based methods and the last three are learning-based methods. For convenience, all the results except GFN shown in the Table 1 are directly cited from [23]. For GFN [32], the latest state-of-the-art dehazing method, they have also reported the results on the RESIDE SOTS indoor dataset in their paper. Although various evaluation metrics are proposed in [23], we only adopt PSNR and SSIM, the most widely used metrics in previous methods. It can be seen that our proposed **GCANet** outperforms all previous dehazing methods by a large margin.

We further show the dehazing results of two indoor and three outdoor hazy images in Figure 3 for qualitative comparisons. From these visual results, we can easily observe that DCP [17] and CAP [46] will make the brightness of the dehazed results relatively dark, which is because of their underlying prior assumptions. For AOD-Net [22], we find that it is often unable to entirely remove the haze from the input. Although GFN [32] can achieve quite good dehazing results in some cases, our **GCANet** is the best one which can both preserve the original brightness and remove the haze as much as possible from the input.

Ablation Analysis To understand the importance of each component in our **GCANet**, we have conducted ablation analysis with and without each specific component. Specifically, we focus on three major components: with / without the smoothed dilation, with / without the gated fusion sub-network, and with instance normalization / batch normalization. Correspondingly, four different network configurations are evaluated on the image dehazing task, and we incrementally add one component to each configuration at a time. As shown in Table 3, the final performance keeps raising in these experiments. However, one interesting observation is that it seems the biggest gain comes from instance normalization in place of batch normalization. Therefore, we further add one experiment by using instance normalization only without smoothed dilation and gated fusion network. Unsurprisingly, it can still achieve slightly better results than the first configuration with batch normalization, but the gain is smaller than the aforementioned one. That is to say, by combing all the designed components together, larger gains can be achieved than only applying one or some of them.

To further validate the effectiveness of our smoothed dilated resblock in alleviating the gridding artifacts, we compare it with the previous widely-used exponentially dilated resblock [7, 9, 25], where the dilation rates of adjacent resblocks are increased exponentially (e.g., 2, 4, 8, 16, 32). As shown in the two representative dehazing examples in Figure 5, the gridding artifacts and color shift

	DCP [17]	CAP [46]	GRM [4]	AOD-Net [22]	DehazeNet [3]	GFN [32]	**GCANet**
PSNR	16.62	19.05	18.86	19.06	21.14	22.30	**30.23**
SSIM	0.82	0.84	0.86	0.85	0.86	0.88	**0.98**

Table 1. Quantitative comparisons of image dehazing on the SOTS indoor dataset from RESIDE. Obviously, Our **GCANet** outperforms all the previous state-of-the-art image dehazing methods by a very large margin.
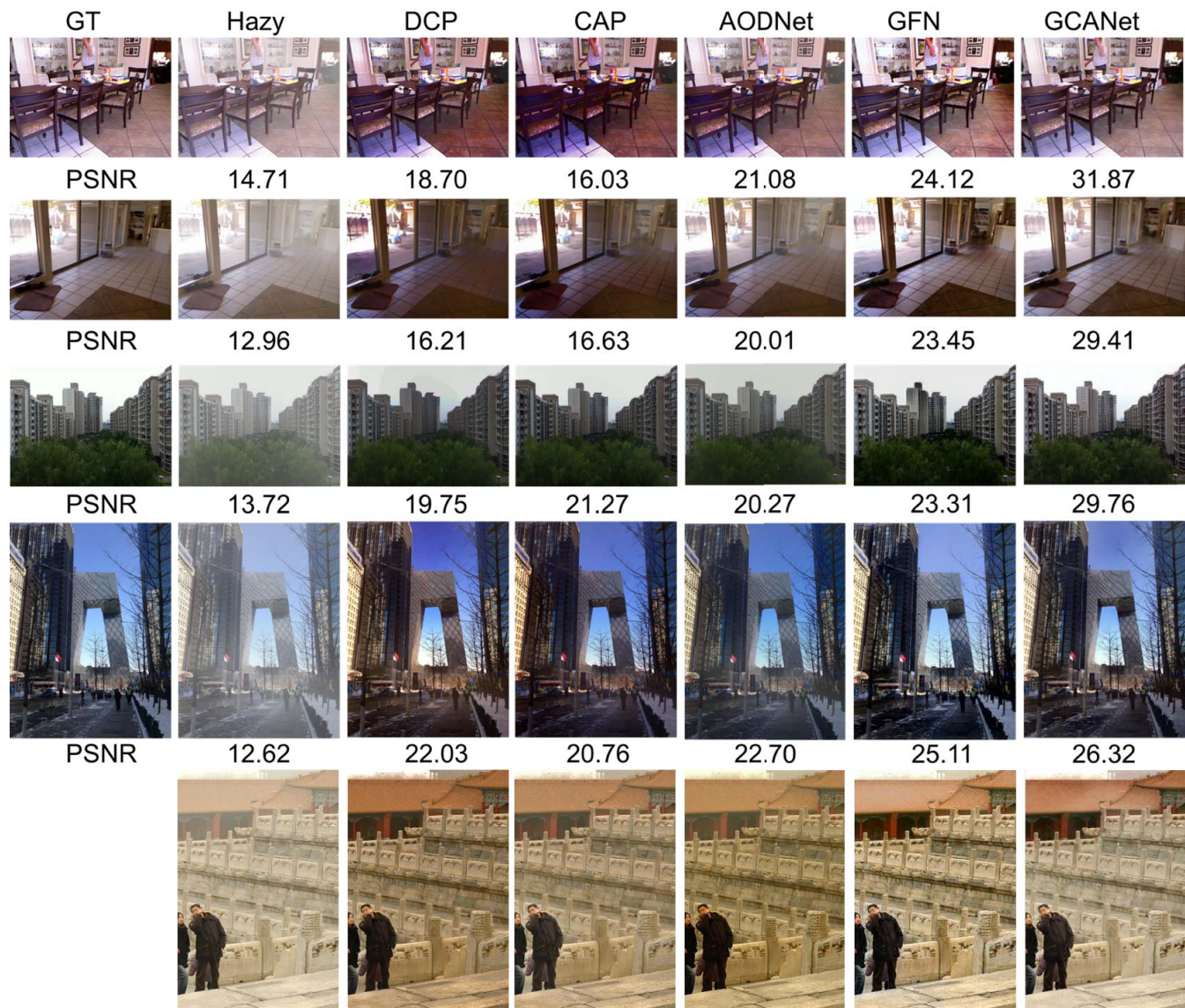

Figure 3. Qualitative comparisons with different dehazing methods for indoor and outdoor hazy images, and the last row is one real hazy example. It can be seen that our *GCANet* is the best one which can remove the underlying haze while maintaining the original brightness.

often happen near the object boundaries and texture regions when the exponentially dilated resblocks are used. By contrast, our smoothed dilated resblocks can address this problem and preserve the original color fidelity.

Generality to Image Deraining Task The task of image deraining is very similar to image dehazing, which aims to remove the rain-streak component from a corrupted image captured in the rainy environment. Though our focus is to design a good network structure for image dehazing, we

1380

DSC[28]	GMM [26]	CNN[13]	JORDER[40]	DDN [14]	JBO [45]	DID-MDN[43]	**GCANet**
21.44	22.75	22.07	24.32	27.33	23.05	27.95	**31.68**

Table 2. Quantitative comparison results (PSNR) of the image deraining task on the DID-MDN test dataset. Although our **GCANet** is mainly designed for image dehazing, it generalizes very well for the image deraining task.

Figure 4. One visual example deraining result for the different state-of-the-art deraining methods. Obviously, previous methods like CNN [13], JORDER [40] tend to under-derain the image, and our **GCANet** can achieve the best deraining results.

smoothed dilation	√	√	√	
gated fusion		√	√	
instance norm			√	√
PSNR	27.57	28.12	28.72	30.23	28.45

Table 3. Detailed ablation analysis for each component with different training configurations, which shows that the combination of all the designed components is the best.
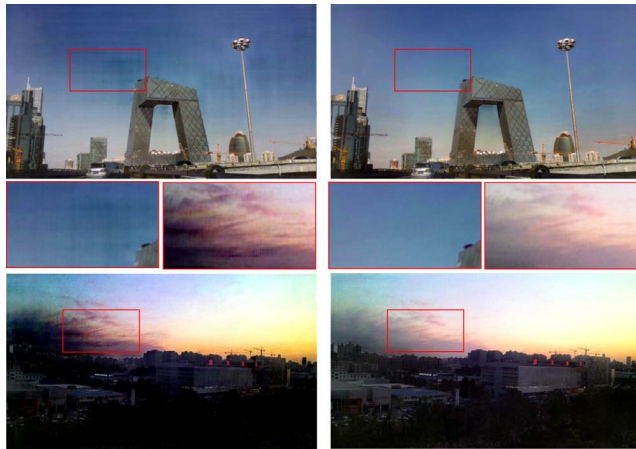

Figure 5. Two dehazing examples to show the superiority of smoothed dilated resblocks (right column) and regular exponentially dilated resblocks (left colum). Obviously, our smoothed dilated resblocks improve the gridding artifacts and produce much better dehazing results.

are also very curious about whether the proposed **GCANet** can be applied to the image deraining task. Specifically, we leverage the training dataset synthesized in [43], and compare our method with seven different image deraining methods: DSC [28], GMM [26], CNN [13], JORDER [40], DDN [14], JBO [45] and DID-MDN [43]. Note that all the results are cited from [43]. Surprisingly, as shown in Table 2, our **GCANet** even outperforms previous best method [43] with more than 3 dB in PSNR.

We also provide one deraining example in Figure 2 for visual comparison. It can be seen that many previous methods like CNN [13, 14] often tend to under-derain the image, and some unexpected patterns may appear in the deraining results of JORDER [40]. To see more details, we crop and zoom-in one local patch from the sky region. It is easy to observe that the deraining result of our **GCANet** is much clearer than other methods.

5. Conclusion

In this paper, we propose an end-to-end gated context aggregation network for image dehazing. To eliminate the gridding artifacts from the dilated convolution, a latest smoothed dilated technique is used. Moreover, a gated sub-network is leveraged to fuse the features of different levels. Despite of the simplicity of the proposed method, it is better than the previous state-of-the-art image dehazing methods by a large margin. We further apply the proposed network to the image deraining task, which can also obtain and state-of-the-art performance. In the future, we will try more facy losses used in [6, 19] and consider to extend to video dehazing like [5].

References

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

[2] D. Berman, S. Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016.

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

[4] C. Chen, M. N. Do, and J. Wang. Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In *ECCV*. Springer, 2016.

[5] D. Chen, J. Liao, L. Yuan, N. Yu, and G. Hua. Coherent online video style transfer. In *Proc. Intl. Conf. Computer Vision (ICCV)*, 2017.

[6] D. Chen, L. Yuan, J. Liao, N. Yu, and G. Hua. Stylebank: An explicit representation for neural image style transfer. In *Proc. CVPR*, volume 1, page 4, 2017.

[7] Q. Chen, J. Xu, and V. Koltun. Fast image processing with fully-convolutional networks. In *IEEE International Conference on Computer Vision*, volume 9, pages 2516–2525, 2017.

[8] F. Chollet. Xception: Deep learning with depthwise separable convolutions. *arXiv preprint*, pages 1610–02357, 2017.

[9] Q. Fan, D. Chen, L. Yuan, G. Hua, N. Yu, and B. Chen. Decouple learning for parameterized image operators. *arXiv preprint arXiv:1807.08186*, 2018.

[10] Q. Fan, J. Yang, G. Hua, B. Chen, and D. P. Wipf. A generic deep architecture for single image reflection removal and image smoothing. In *ICCV*, pages 3258–3267, 2017.

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

[12] R. Fattal. Dehazing using color-lines. *ACM transactions on graphics (TOG)*, 34(1):13, 2014.

[13] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *TIP*, 2017.

[14] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley. Removing rain from single images via a deep detail network. In *CVPR*, 2017.

[15] R. Hamaguchi, A. Fujita, K. Nemoto, T. Imaizumi, and S. Hikosaka. Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1442–1450. IEEE, 2018.

[16] N. Hautière, J.-P. Tarel, and D. Aubert. Towards fog-free in-vehicle vision systems through contrast restoration. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[17] 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.

[18] 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.

[19] M. He, D. Chen, J. Liao, P. V. Sander, and L. Yuan. Deep exemplar-based colorization, 2018.

[20] 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.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[22] 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*, volume 1, page 7, 2017.

[23] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Reside: A benchmark for single image dehazing. *arXiv preprint arXiv:1712.04143*, 2017.

[24] R. Li, J. Pan, Z. Li, and J. Tang. Single image dehazing via conditional generative adversarial network. *methods*, 3:24, 2018.

[25] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. *ECCV*, 2018.

[26] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown. Rain streak removal using layer priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2736–2744, 2016.

[27] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[28] Y. Luo, Y. Xu, and H. Ji. Removing rain from a single image via discriminative sparse coding. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3397–3405, 2015.

[29] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *Proceedings of the IEEE international conference on computer vision*, pages 617–624, 2013.

[30] S.-C. Pei and T.-Y. Lee. Nighttime haze removal using color transfer pre-processing and dark channel prior. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 957–960. IEEE, 2012.

[31] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*, pages 154–169. Springer, 2016.

[32] 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.

[33] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE*

conference on computer vision and pattern recognition, pages 1–9, 2015.

[34] 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.

[35] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Instance normalization: the missing ingredient for fast stylization. corr abs/1607.08022 (2016).

[36] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell. Understanding convolution for semantic segmentation. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1451–1460. IEEE, 2018.

[37] Z. Wang and S. Ji. Smoothed dilated convolutions for improved dense prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2486–2495. ACM, 2018.

[38] B. Xie, F. Guo, and Z. Cai. Improved single image dehazing using dark channel prior and multi-scale retinex. In *Intelligent System Design and Engineering Application (ISDEA), 2010 International Conference on*, volume 1, pages 848–851. IEEE, 2010.

[39] H. Xu, J. Guo, Q. Liu, and L. Ye. Fast image dehazing using improved dark channel prior. In *Information Science and Technology (ICIST), 2012 International Conference on*, pages 663–667. IEEE, 2012.

[40] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan. Deep joint rain detection and removal from a single image. In *CVPR*, 2017.

[41] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[42] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

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

[44] H. Zhang, V. Sindagi, and V. M. Patel. Multi-scale single image dehazing using perceptual pyramid deep network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 902–911, 2018.

[45] L. Zhu, C.-W. Fu, D. Lischinski, and P.-A. Heng. Joint bilayer optimization for single-image rain streak removal. In *ICCV*, 2017.

[46] Q. Zhu, J. Mai, L. Shao, et al. A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Processing*, 24(11):3522–3533, 2015.