

Nonuniformly Dehaze Network for Visible Remote Sensing Images

Zhaojie Chen¹, Qi Li¹, Huajun Feng¹, Zhihai Xu¹, and Yueting Chen¹ 
¹College of Optical Science and Engineering, Zhejiang University, China
 {22030045, liqi, fenghj, xuzh, chenyt}@zju.edu.cn

Abstract

Nonuniform haze on remote sensing images degrades image quality and hinders many high-level tasks. In this paper, we propose a Nonuniformly Dehaze Network towards nonuniform haze on visible remote sensing images. To extract robust haze-aware features, we propose Nonuniformly Excite (NE) module. Inspired by the well-known gather-excite attention module, NE module works in a map-excite manner. In the map operation, we utilize a proposed Dual Attention Dehaze block to extract local enhanced features. In the gather operation, we utilize a strided deformable convolution to nonuniformly process features and extract non-local haze-aware features. In the excite operation, we employ a pixel-wise attention between local enhanced features and nonlocal haze-aware features, to gain finer haze-aware features. Moreover, we recursively embed NE modules in a multi-scale framework. It helps not only significantly reduce network's parameters, but also recursively deliver and fuse haze-aware features from higher levels, which makes learning more efficient. Experiments demonstrate that the proposed network performs favorably against the state-of-the-art methods on both synthetic and real-world images.

1. Introduction

Remote sensing (RS) images have been used in extensive fields including military, forestry, and agricultural monitoring [27]. However, the quality of optical RS images is usually contaminated by haze, and it makes various critical tasks more difficult, such as target detection [41], land cover classification [38], and tiny object detection [31]. Haze is caused by clouds, atmospheric moisture, and bad weather [19], which is considered as a condensation of light water vapor, fine dust, or smoke [18]. Under haze conditions, the imaging light is scattered and attenuated, which causes poor visibility, low contrast, and low intensity of images. Dehazing on optical RS images improves the quality of images and accuracy of various applications.

Different from the haze on ground natural images, be-

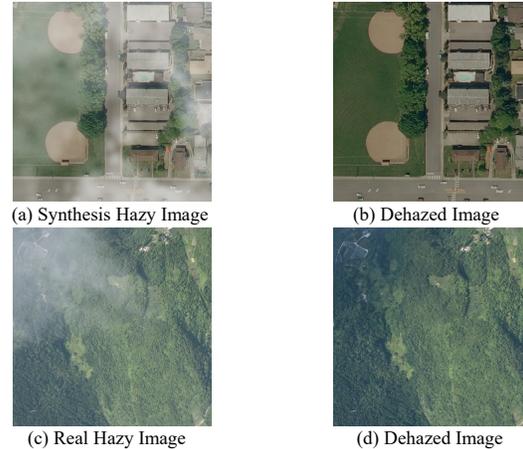


Figure 1. Dehazed results of proposed NDNNet on synthesis hazy image and real-world hazy image.

cause of the vast field of view of RS images and more complicated air conditions, the haze on RS images could be very nonuniform. Nonuniform haze is closer to the real cases than uniform ones [9], and it is different from the uniform ones on the physical aspect.

The traditional physical haze model is defined as

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

where J , t , A denote the clear image, transmission map and global atmospheric light, respectively. For uniform haze, the transmission map is given as

$$t(x) = e^{-\beta d(x)}, \quad (2)$$

where d is the scene depth, and β is the scattering coefficient of the atmosphere.

For remote sensing nonuniform haze, the scene depth can be considered as a constant because of the remote imaging distance. However, the scattering parameter is no longer a constant, and should be only related to the distribution of the haze. That is,

$$t(x) = e^{-\beta(x)d}, \quad (3)$$

which means the transmission map is only related to the distribution of nonuniform haze, and is not related to the information of the scene. Therefore, it means methods should learn less useful information for dehazing from details or edges of scenes, and requires more about the ability of haze awareness, which leads to extra difficulties for dehazing.

Recently, convolution neural networks (CNNs) achieve state-of-the-art (SOTA) performance on ground natural image dehazing [5–7, 24], and the CNN-based method is becoming a promising method for visible RS image haze removal [4, 15, 16, 19, 22, 34]. However, despite the awareness of haze nonuniformity, previous CNN-based works tended to refer to networks from natural image dehazing, and did not propose targeted networks towards nonuniform haze removal. Haze on natural images can be considered uniform, so natural image dehazing methods tend to learn useful information from details or edges of scenes [26, 33], and rarely consider awareness of haze distribution. Therefore, simply referring to networks from natural dehazing methods will lead to low efficiency.

In this work, we propose an end-to-end deep neural network, named Nonuniformly Dehaze network (NDNet), towards RS nonuniform haze removal. To extract robust haze-aware features, we propose Nonuniformly Excite (NE) module. Inspired by the well-known gather-excite attention module [13], NE module works in a map-gather-excite manner. In the map operation, we utilize a proposed Dual Attention Dehaze (DAD) block to extract local enhanced features. Specifically, DAD block is inspired by the physical dehazing model, and utilizes two attention-based branches to fit its two items respectively. In the gather operation, we extract nonlocal haze-aware features with a combination of down-sample and up-sample operations. During the down-sample operation, to encode more nonlocal information and nonuniformly process features, we utilize a strided deformable convolution layer. The deformable convolution layer learns a deformable offset, and thus it is more likely to generate more flexible features and encode nonuniform haze-aware information. Finally, in the excite operation, to nonuniformly excite the extracted features, we utilize a pixel-wise attention between local enhanced features and nonlocal haze-aware features, and output finer haze-aware features. Moreover, we recursively embed NE modules in a multi-scale framework. A lower-level NE module will compute local enhanced features with its own DAD block. However, for the nonlocal haze-aware features, a lower-level NE module will recursively call a higher-level NE module to compute. It helps not only significantly reduce parameters of network, but also fuses lower-level and higher-level haze-aware features, which makes learning more efficient.

The contributions of this work are:

- We propose the Nonuniformly Dehaze network for RS

nonuniform haze removal, which outperforms other SOTA methods on synthetic and real-world images.

- We propose a Nonuniformly Excite (NE) module for nonuniform degradation. In the map operation, we enhance features with Dual Attention Dehaze blocks. In the gather-excite operation, we nonuniformly process features with strided deformable convolution layers and then excite local enhanced features.
- We recursively embed NE modules in a multi-scale framework, which helps not only significantly reduce trainable parameters, but also fuse lower-level and higher-level features.

2. Related work

Early works [10, 17, 36, 39] on single RS image dehazing often use strong priors or assumptions to constrain the optimized result and restore dehazed images. In [17], Jiang *et al.* adopt a proportional strategy in the blue, green, red and near-infrared bands, to infer the haze thickness map. Then they remove haze by subtracting the haze thickness map from each band. In [39], Zhang *et al.* propose the dehazing method based on dark channel prior, and they further propose their proportion-based spatial adaptive strategy and spectral adaptive strategy to diminish the phenomenon of color distortion. In [36], Wen *et al.* utilize a haze optimized transformation (HOT) in bands blue, green and red, to assess haze spatial distribution. Then they apply the HOT image to remove the radiometric effects of haze. In [10], Guo *et al.* propose a novel elliptical boundary-prior to transform the haze thickness in each local patch from the pixels cluster in the spectral space. They estimate the transmission map according to the generated haze thickness map and then remove haze. Single image dehazing is an ill-posed problem, and the priors maybe not be robust in special cases. Sometimes prior-based methods generate halo and gradient reversal artifacts [9].

Recently, the CNN-based method [4, 15, 16, 19, 22] is becoming a promising method for single RS image dehazing. Following the prominent tendency in image reconstruction, most of them are trained in an end-to-end manner without estimating transmission maps and atmospheric lights. In [16], Jiang *et al.* propose a multi-scale residual convolutional neural network (MRCNN). Ke *et al.* [19] develop their network based on the fully convolutional network. Huang *et al.* [15] stack dilated convolutional blocks to gain a bigger receptive field. Li *et al.* [22] stack residual dense blocks (RDB) and proposed their two-stage network to finely dehaze. Chen *et al.* [4] refer to HRNet [35] and further propose their network structure. Nevertheless, these methods tend to refer to network architectures from single natural image dehazing without significant modification and do not propose targeted methods, which are inefficient for remote sensing image dehazing.

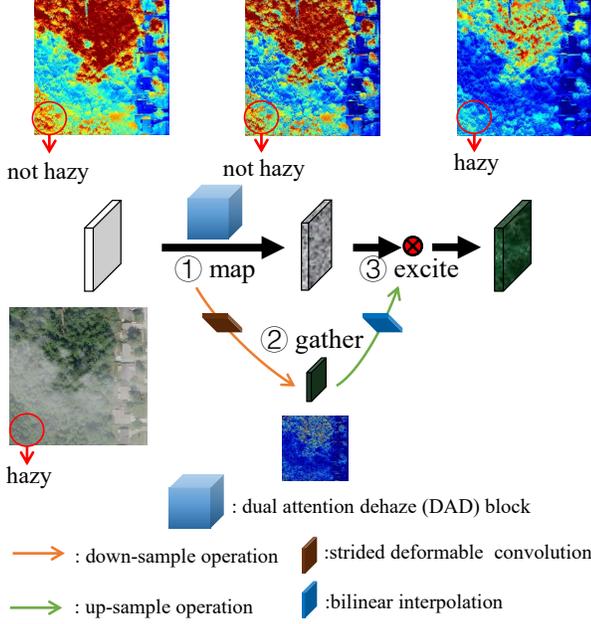


Figure 2. The procedure of the proposed Nonuniformly Excite Module.

3. Proposed Method

In this section, we will describe the proposed network. First, we will introduce the core NE module. Second, we will describe the used recursive structure and further introduce the whole network architecture.

3.1. Nonuniformly Excite Module

Dehazing on RS images is related to the ability of haze awareness. It means the network should encode the global information of haze distribution and thus requires bigger effective receptive fields (ERFs) [28]. Previous work GENet [13] has got a bigger ERF with gather-excite operations. Inspired by the gather-excite attention module, the proposed Nonuniformly Excite (NE) module works in a map-gather-excite manner to extract robust haze-aware features, as shown in Fig.2.

Map Operation. In the map operation, we propose Dual Attention Dehaze (DAD) blocks to extract local enhanced features. We are inspired by the traditional physical dehazing model,

$$J(x) = A + \frac{I(x) - A}{t(x)}. \quad (2)$$

Note that the first item is only related to the channel information, and the second item is related to both channel information and spatial information. Therefore, we utilize a channel attention branch and a pixel-wise attention to fit these two items, respectively. Eventually, we use a concatenation and 1×1 convolution to fuse features.

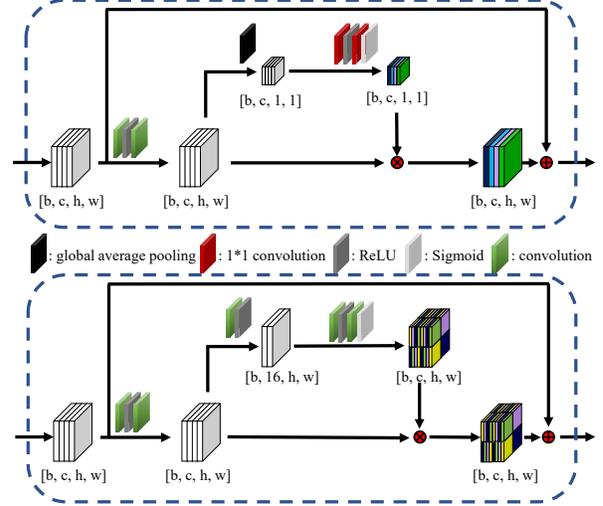


Figure 3. The channel attention branch and the pixel-wise attention branch of Dual Attention Dehaze block.

For the channel attention branch, we employ a typical Residual Channel Attention Block (RCAB) [40], which has a combination of squeeze, excitation, recalibration and residual connection. Following [14], we use the reduction ratio of 16 during excitation operation. For the pixel-wise attention branch, we develop similar processes. To avoid the loss of spatial information during pooling operation, we squeeze features on channel dimension with a convolution. To fuse information from both channel dimension and spatial dimension, we use a common 3×3 convolution instead of a 1×1 convolution. Crossing three 3×3 common convolution layers, it has a theoretic receptive field of 7×7 , as large as that of CBAM [37].

As a feature extraction block, DAD block can effectively process features within its ERF, and thus features extracted by the map operation is local. As shown in Fig.2, the map operation tends to locally enhance original features with finer textures.

Gather-Excite Operation. In the gather operation, we use a combination of down-sample and up-sample operations. During the down-sample operation, to encode more non-local information, we utilize a strided deformable convolution layer. Deformable convolution nonuniformly learns a offset on features, thus generates more flexible features, and can better fit the haze distribution. As shown in Fig.2, the down-sampled features reflected more about the overall haze distribution, and focus less on the local texture of features. Therefore, after the up-sample operation, the gather operation will generate nonlocal haze-aware features.

Thus, in the excite operation, we utilize a pixel-wise attention to nonuniformly excite local features extracted by DAD blocks. Finally, DAD blocks output processed features. As shown in Fig.2, compared with features extracted

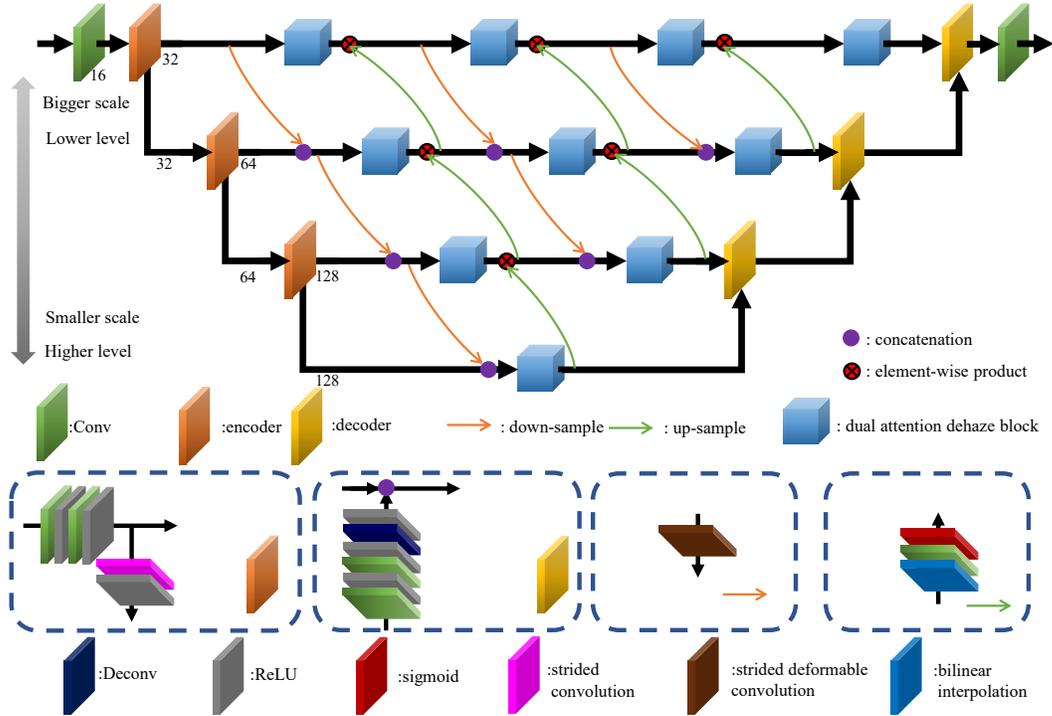


Figure 4. The architecture of the proposed NDNet.

by DAD blocks, the processed features reflect a more precise boundary of haze, and thus can better reflect the haze distribution.

3.2. Network

Recursive Structure. To make NE modules work on a multi-scale framework, a natural idea is to embed NE modules on every level of the framework, as shown in Fig.5(a). To make learning more efficient, we optimize the network design with recursive structure.

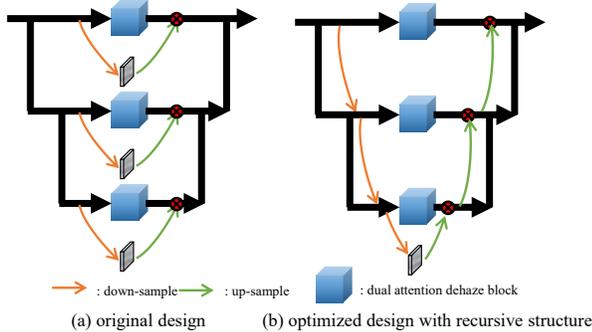


Figure 5. We optimize the original design by recursively embedding NE modules into a multi-scale framework.

As shown in Fig.5(b), a lower-level NE module will generate local enhanced features with its own DAD block.

However, for nonlocal haze-aware features, a lower-level NE module will recursively call a higher-level NE module to compute. It takes several advantages to employ such a recursive structure. First, it significantly reduces trainable parameters of the network. We do not need to compute nonlocal haze-aware features for every level with a sequence of convolution layers, but just directly up-sample higher-level haze-aware features with bilinear interpolation and 1×1 convolution, which has a significantly lower amount of parameters. Second, it helps fuse multi-scale haze-aware features in a bottom-up way. A lower-level NE module will use features generated by higher-level NE modules, and therefore it helps fuse lower-level and higher-level haze-aware features. Moreover, on the right side of DAD blocks, features will be progressively excited in a bottom-up way. Third, it makes learning more efficient. Higher-level haze-aware features have been processed by several NE modules, so they can represent the overall haze distribution coarsely. Therefore, compared with computing nonlocal features from input, directly utilizing higher-level haze-aware features is more efficient.

Whole Network Architecture. The whole network is illustrated in Fig.4. Following the prominent tendency in image reconstruction, we process features in a multi-scale framework. We embed three recursive structures in the framework. To reduce the trainable parameters, we progressively reduce the depth of recursive structures. Features in higher

levels are deeper and more abstract, so we use fewer DAD blocks on higher levels. Previous works [5, 8, 24] usually fuse features in a top-down way on the left of the network, and in a bottom-up way on the right side. However, there are two feature fusion connections staggered inside the network. On the left side of DAD blocks, features are aggregated in a top-down way. On the right side of DAD blocks, features will be progressively excited in a bottom-up way. Moreover, outputs of the last recursive structure will be fed into the next recursive structure, so features will continue to be enhanced and fused. Therefore, crossing 3 recursive structures, features will subsequently evolve.

4. Experimental Results

4.1. Dataset

RSNHaze. Sufficient data is vital for supervised learning. However, the lack of a large-scale dataset for RS dehazing hinders the development of network-based methods. Therefore, we build a new large-scale RS dehazing dataset, named remote sensing nonuniform haze dataset (RSNHaze).

For the clear haze-free images, we utilize data from the publicly available dataset, The Inria Aerial Image Labeling Dataset [29], which offers aerial orthorectified color imagery. It contains 360 aerial images of 36 scenes with an image resolution of 5000×5000 . The abundant haze-free data allow us to build a large-scale dataset. For RS haze, we get data from cloudGAN [11]. CloudGAN uses a pretrained CycleGAN model with data of Sentinel-3 SLSTR Product, and it converts cloud masks into cloud images to generate vivid clouds.

First, we employ 87.5% of the whole data for training and the rest for testing. Second, we split every 5000×5000 haze-free image into 81 patches, and each of them has an image resolution of 512×512 . Third, we up-sample the could images to the resolution of 1536×1536 , and implement data augment of random horizontal flipping, random vertical flipping and random rotation. And we randomly crop a 512×512 area to generate transmission maps. Fourth, we generate corresponding hazy images with haze-free images and transmission maps according to Equation (1). To further increase diversity of image pairs, the atmospheric light we used here is $A \in (0.8, 1.0)$, and the transmission maps will be multiplied by a parameter $para \in (\frac{1}{e}, \frac{1.5}{e})$. Smaller transmission values indicate thicker haze, which increases the difficulty of dehazing. Finally, we get 29160 image pairs. 25920 of them are used for training, and 3240 of them are used for testing.

As shown in Fig.6(a) and (b), synthetic hazy images have a similar pattern to real-world hazy images. Some examples of synthetic image pairs are shown in Fig.6(b) and (c).

NHHaze2 Due to the lack of other available RS nonuni-



(a) real-world hazy image (b) synthetic hazy image (c) ground truth

Figure 6. (a) and (b) demonstrate that synthetic hazy images have a similar pattern to real-world hazy images. Some examples of synthetic image pairs are shown in (b) and (c).

form haze datasets, we evaluate NDNet on a ground natural nonuniformly hazy dataset. Here, we choose NHHaze2 [1] dataset, which contains 25 image pairs. We use images 1-20 for training and 21-25 for testing.

4.2. Implementation Details

The proposed network is trained in an end-to-end manner. We train the network with patches of 128×128 and a batch size of 16. We use Adam an optimizer [20] with default setting of $\beta_1=0.9$ and $\beta_2=0.999$. For RSNHaze, we train the network for 200 epochs. The initial learning rate is 0.0003, and it is reduced by half every 25 epochs. For NHHaze2, we train the network for 10000 epochs. The initial learning rate is 0.0003, and it is reduced by half every 2000 epochs. The loss function in all experiments is a combination of smooth L1 loss SL_1 and VGG-based perceptual loss L_{per} , that is,

$$loss(out, gt) = SL_1(out, gt) + 0.04 \times L_{per}(out, gt). \quad (3)$$

We implement the proposed network based on the PyTorch using a machine with two NVIDIA GTX 3090 GPUs.

4.3. Quantitative Evaluation on Synthetic Dataset

Result on RSNHaze. We evaluate the proposed NDNet on our RSNHaze dataset against SOTA methods of both RS image dehazing methods (SateHaze [15], FCFTNet [22]) and natural image dehazing methods (DCP [12], DehazeNet [2], AODNet [21], GCANet [3], GDNNet [24], PFFNet [30], DuRN [25], MSBDN [6]). Except for DCP, all methods are based on convolution neural networks. Except for SateHaze, all networks are trained with the model code given by authors. SateHaze does not open their code, so we built the network according to the model description from their paper. DehazeNet uses the strategy of first estimating the

Table 1. Quantitative evaluations on the RSNHaze dataset. Red text indicates the best performance.

Methods	natural image dehazing method								RS image dehazing method		
	DCP	DehazeN	AODN	GCANet	GDN	PFFN	DuRN	MSBDN	SateHaze	FCFTN	Ours
PSNR	17.44	22.08	22.76	33.88	33.99	28.10	35.06	35.48	24.87	29.51	36.30
SSIM	0.8217	0.8211	0.9479	0.9921	0.9929	0.9620	0.9921	0.9937	0.956	0.9832	0.9948

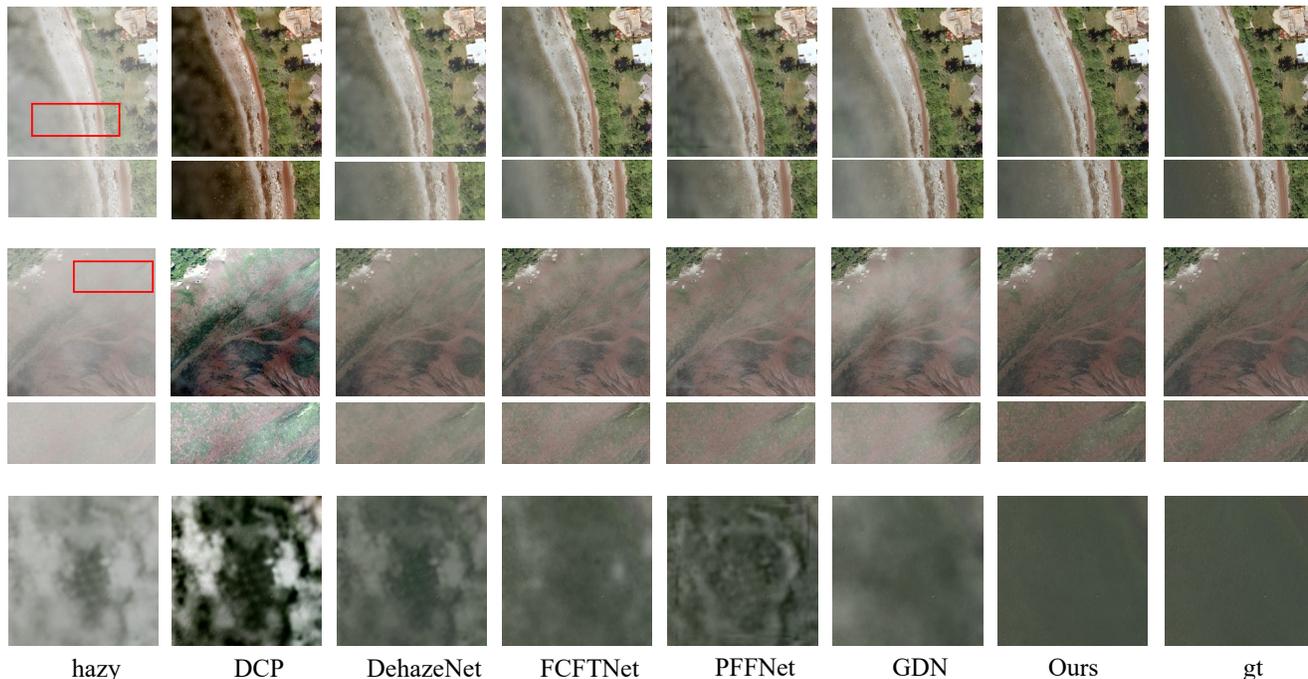


Figure 7. Image dehazing results on RSNHaze dataset.

transmission maps and then recovering the haze-free images. For fair comparisons, we train DehazeNet in an end-to-end way. For further fair comparisons, all networks are retrained and tested on our RSNHaze dataset, and all of them are trained for 200 epochs with the same learning rate scheme. The loss functions of networks are the same. We use the most common metrics, peak signal to ratio (PSNR) and structure similarity (SSIM) for the quantitative assessment of dehazed outputs.

Table 1 shows the experimental results on RSNHaze. As a prior-based method, DCP does not perform well, which achieves the lowest PSNR. DehazeNet and AODNet are early neural networks for haze removal and limited by the depth of networks, they are less effective than other network-based methods. Other SOTA methods deliberately design their network architectures, so they achieve impressive results with high PSNR and SSIM. Our proposed network achieves higher metrics than not only RS image dehazing methods, but also natural image dehazing methods. Especially compared to RS image dehazing methods, our network outperforms other methods by a wide margin.

Fig. 7 shows dehazed outputs of some SOTA methods. Due to the inaccurate estimation of nonuniform haze thickness, DCP leaves a mass of haze and suffers from severe color distortions. Network-based methods generate results with less color distortion. However, they do not get satisfactory details on their results. As shown in the first row, other network-based methods usually remain some haze on their results. However, NDNet remains less haze and generates clearer results. As shown in the second row, other network-based methods may blur textures and overly smooth details. However, NDNet gets more distinct details on its results. The last row shows the strong capacity of NDNet to nonuniformly de haze. Such a hazy input is very difficult to de haze, because the haze is extremely nonuniform. Previous SOTA methods more or less remain haze on their outputs. However, the result of NDNet is clean and haze-free.

Result on NHHaze2. To demonstrate NDNet’s capability of nonuniformly dehazing, we also evaluate it on NHHaze2. We load pretrained parameters on RSNHaze, and then train NDNet on NHHaze2. We perform our network against SOTA methods (AODNet, GCANet, FFANet [32],

Table 2. Quantitate evaluation on real-world dehazed images with non-reference image quality assessment methods. Red text, blue text and green text indicate the best, the second-best and the third-best performance, respectively.

Methods	DCP	DehazeNet	AODNet	FCFTNet	GDN	PFFNet	DuRN	MSBDN	Ours
NIQE↓	12.21	19.24	11.91	11.97	11.75	12.24	12.10	11.80	11.62
BRISQUE↓	54.05	34.94	48.57	47.89	48.47	47.27	54.76	47.89	47.40

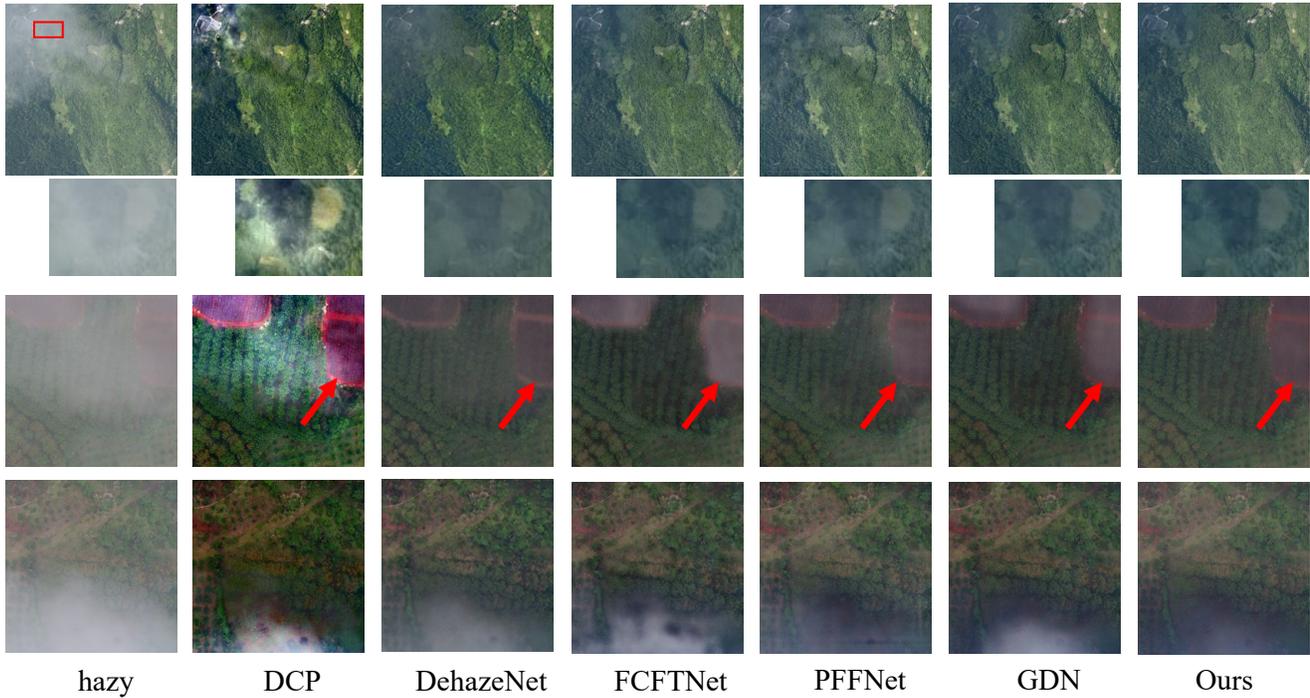


Figure 8. Qualitative comparisons on real RS hazy images.

TDN [23]). The experimental results are listed in Table 3, and it shows that NDNet outperforms previous SOTA by a considerable margin on NHHaze2. Dehazed outputs are shown in Fig. 9, and it shows that NDNet generates visually haze-free images facing nonuniform haze.

Table 3. Quantitative evaluations on the NHHaze2. Red text indicates the best performance.

Methods	AODNet	GCANet	FFANet	TDN	Ours
PSNR	13.30	18.79	20.45	20.23	21.36
SSIM	0.4693	0.7729	0.8043	0.7622	0.8472

4.4. Evaluation on Real-world Images

To further demonstrate the generalization ability of NDNet, we evaluate it on real-world RS hazy images. We select 25 real RS images with nonuniform haze. Fig. 8 shows some dehazed results by different methods. As shown in the first row, DCP, DehazeNet and PFFNet usually leave thin haze on their outputs. FCFTNet, GDN and NDNet generate less



Figure 9. Our dehazed results on images 21-25 of NHHaze2.

haze on their results. However, NDNet usually gets more discriminable edges and enhances higher contrast than other methods. As shown in the second row, all methods perform greatly for dehazing, but they may suffer from color distortions in some areas. NDNet will generate fewer color distortions than other methods. As shown in the last row, when facing extremely nonuniform haze, all the methods do not obtain satisfactory results, but our result leaves less haze.

We further perform a quantitative evaluation on the real-world dehazed images. The indicators include NIQE and BRISQUE, and both of them are non-reference image quality assessment methods. The evaluation is listed in Table 2. Considering the combination of NIQE and BRISQUE, the proposed network generates results with higher quality.

4.5. Ablation Study

We perform ablation experiments to investigate the effectiveness of operations in NE modules, that is the map, gather and excite operation.

Effectiveness of the map operation. We mainly utilize DAD block in the map operation, and therefore we perform two ablation studies for it. The first ablation experiment is that we remove 1 or 2 branches of DAD block. The second ablation experiment is that we replace DAD block with 1 or 2 residual dense blocks (RDBs). The number of convolution layers in RDBs is 4 and the growth rate is the same as the channel number of input features. After modifying, we retrain the network with the same experimental settings as the main experiment. The experimental results are listed in Table 4. If we only use one of the two branches, PSNR will descend by 0.48 or 1.86dB. When removing both two branches of DAD block, PSNR will severely descend by over 3dB, which is more than 1.86+0.48dB and demonstrates the effectiveness of the two branches. The experiment result also shows that the channel attention branch is more important than the pixel-wise attention branch. Because compared with employing no branch, employing channel attention branch achieves an improvement of 2.69dB on PSNR, while employing pixel-wise attention achieves an improvement of 1.31dB. We infer that it is because the squeeze operation in the channel attention branch is more compact, so the information flow to attention maps is more efficient. Moreover, after replacing DAD with RDBs, PSNR descends almost 0.92dB or 0.57dB. We infer that it is because DAD blocks mainly utilize attention mechanism, and attention mechanism is more efficient than stacking convolution layers when facing nonuniform degradation. It further reveals that DAD has a stronger mapping capability than RDB.

Effectiveness of the gather-excite operation. We perform two ablation studies for the used gather-excite operation. In the first experiment, we remove both gather and excite operations, which means there are no connections across levels inside the multi-scale framework. In the second experiment, we change the gather or excite operation. First, we replace the deformable convolution with common convolution. Second, we use the concatenation method, instead of pixel-wise attention in the excite operation. To ensure the channel of features will not change after concatenation, we then use a 1×1 convolution. The experimental results are listed in Table 5. Without gather-excite operations,

Table 4. Results of ablation study for map operation.

channel attention branch	pixel-wise attention branch	PSNR/SSIM
		33.13/0.9942
✓		35.82/0.9945
	✓	34.44/0.9937
✓	✓	36.30/0.9948
DAD	number of RDBs	PSNR/SSIM
	1	35.38/0.9943
	2	35.73/0.9947
✓	0	36.30/0.9948

Table 5. Results of ablation study for gather-excite operation.

Method	Gather-excite	W/o gather-excite	Common convolution	Concatenation
PSNR	36.30	35.18	36.13	35.45
SSIM	0.9948	0.9935	0.9948	0.9943

the multi-scale information will not communicate, which makes learning inefficient, so PSNR decreases over 1dB. With common convolution, the learned features will be not so nonuniform. However, the overall map-gather-excite operations will help subsequently evolve features. So the metrics just decrease slightly. After replacing pixel-wise attention with concatenation, PSNR decreases by over 0.8dB. It shows that pixel-wise attention is more suitable for nonuniform degradation than concatenation.

5. Conclusion

In this paper, we propose an end-to-end network for nonuniform haze removal on visible RS images. For feature extraction, we propose Nonuniformly Excite module, which works in a map-gather-excite operation. In the map operation, we utilize the proposed Dual Attention Dehaze blocks to extract local enhanced features. In the gather-excite operation, we utilize a strided deformable convolution to nonuniformly process features and then excite local enhanced features. For feature fusion, we recursively embed NE modules in a multi-scale framework. It helps not only significantly reduce network’s parameters, but also recursively deliver and fuse haze-aware features from higher levels. Experiments demonstrate that the proposed network performs favorably against SOTA methods on synthetic and real images.

6. Acknowledgement

This work is supported by Civil Aerospace Pre-Research Project (No. D040104) and National Natural Science Foundation of China (No. 61975175).

References

- [1] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluiianu, and Radu Timofte. Ntire 2021 nonhomogeneous dehazing challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 627–646, 2021. 5
- [2] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 5
- [3] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In *2019 IEEE winter conference on applications of computer vision (WACV)*, pages 1375–1383. IEEE, 2019. 5
- [4] Xiang Chen, Yufeng Li, Longgang Dai, and Caihua Kong. Hybrid high-resolution learning for single remote sensing satellite image dehazing. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021. 2
- [5] Qili Deng, Ziling Huang, Chung-Chi Tsai, and Chia-Wen Lin. Hardgan: A haze-aware representation distillation gan for single image dehazing. In *European Conference on Computer Vision*, pages 722–738. Springer, 2020. 2, 5
- [6] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2157–2167, 2020. 5
- [7] Jiangxin Dong and Jinshan Pan. Physics-based feature dehazing networks. In *European Conference on Computer Vision*, pages 188–204. Springer, 2020. 2
- [8] Damien Fourure, Rémi Emonet, Elisa Fromont, Damien Muselet, Alain Tremeau, and Christian Wolf. Residual conv-deconv grid network for semantic segmentation. *arXiv preprint arXiv:1707.07958*, 2017. 5
- [9] Jianhua Guo, Jingyu Yang, Huanjing Yue, Hai Tan, Chunping Hou, and Kun Li. Rsdhazenet: Dehazing network with channel refinement for multispectral remote sensing images. *IEEE Transactions on geoscience and remote sensing*, 59(3):2535–2549, 2020. 1, 2
- [10] Qiang Guo, Hai-Miao Hu, and Bo Li. Haze and thin cloud removal using elliptical boundary prior for remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):9124–9137, 2019. 2
- [11] GANESH. H. Paired training data generator for cloud masking ai. <https://app.gn3.sh/cloudGAN/>. Mar 2, 2023. 5
- [12] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 5
- [13] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Andrea Vedaldi. Gather-excite: Exploiting feature context in convolutional neural networks. *Advances in neural information processing systems*, 31, 2018. 2, 3
- [14] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 3
- [15] Binghui Huang, Li Zhi, Chao Yang, Fuchun Sun, and Yixu Song. Single satellite optical imagery dehazing using sar image prior based on conditional generative adversarial networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1806–1813, 2020. 2, 5
- [16] Hou Jiang and Ning Lu. Multi-scale residual convolutional neural network for haze removal of remote sensing images. *Remote Sensing*, 10(6):945, 2018. 2
- [17] Hou Jiang, Ning Lu, Ling Yao, and Xingxing Zhang. Single image dehazing for visible remote sensing based on tagged haze thickness maps. *Remote Sensing Letters*, 9(7):627–635, 2018. 2
- [18] YJ Kaufman. The atmospheric effect on remote sensing and its correction. *Theory and application of optical remote sensing*, pages 336–428, 1989. 1
- [19] Ling Ke, Puyun Liao, Xiaodong Zhang, Guanzhou Chen, Kun Zhu, Qing Wang, and Xiaoliang Tan. Haze removal from a single remote sensing image based on a fully convolutional neural network. *Journal of Applied Remote Sensing*, 13(3):036505, 2019. 1, 2
- [20] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5
- [21] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pages 4770–4778, 2017. 5
- [22] Yufeng Li and Xiang Chen. A coarse-to-fine two-stage attentive network for haze removal of remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 18(10):1751–1755, 2020. 2, 5
- [23] Jing Liu, Haiyan Wu, Yuan Xie, Yanyun Qu, and Lizhuang Ma. Trident dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 430–431, 2020. 7
- [24] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Grid-dehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7314–7323, 2019. 2, 5
- [25] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7007–7016, 2019. 5
- [26] Yantong Liu, Hui Yin, Jin Wan, Zhihao Liu, and Aixin Chong. Edge aware network for image dehazing. *IEEE Signal Processing Letters*, 29:174–178, 2021. 2
- [27] Jiao Long, Zhenwei Shi, Wei Tang, and Changshui Zhang. Single remote sensing image dehazing. *IEEE Geoscience and Remote Sensing Letters*, 11(1):59–63, 2013. 1
- [28] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. *Advances in neural information processing systems*, 29, 2016. 3
- [29] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In *2017 IEEE International Geoscience and Remote Sensing*

- Symposium (IGARSS)*, pages 3226–3229. IEEE, 2017. 5
- [30] Kangfu Mei, Aiwen Jiang, Juncheng Li, and Mingwen Wang. Progressive feature fusion network for realistic image dehazing. In *Asian conference on computer vision*, pages 203–215. Springer, 2018. 5
- [31] Jiangmiao Pang, Cong Li, Jianping Shi, Zhihai Xu, and Hua-jun Feng. R2-cnn: Fast tiny object detection in large-scale remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):5512–5524, 2019. 1
- [32] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11908–11915, 2020. 6
- [33] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, 2018. 2
- [34] Praveer Singh and Nikos Komodakis. Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1772–1775. IEEE, 2018. 2
- [35] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5693–5703, 2019. 2
- [36] Xingping Wen and Xiaofeng Yang. Haze removal from the visible bands of cbers remote sensing data. In *2009 International Conference on Industrial and Information Systems*, pages 118–121. IEEE, 2009. 2
- [37] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. 3
- [38] Fan Yang, Jianhua Guo, Hai Tan, and Jingxue Wang. Automated extraction of urban water bodies from zy-3 multi-spectral imagery. *Water*, 9(2):144, 2017. 1
- [39] Chi Zhang, Huifang Li, Huanfeng Shen, and Jie Li. A spatial—spectral adaptive haze removal method for remote sensing images. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 6134–6137. IEEE, 2017. 2
- [40] Yulun Zhang, Kungpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 3
- [41] Dan Zhu, Bin Wang, and Liming Zhang. Airport target detection in remote sensing images: A new method based on two-way saliency. *IEEE Geoscience and Remote Sensing Letters*, 12(5):1096–1100, 2015. 1