MULTIPLE PATCH-AWARE NETWORK FOR FASTER REAL-WORLD IMAGE DEHAZING

Kun Yang, Juan Zhang, Xiaoqi Lang*
Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China

ABSTRACT

This paper proposes a Multiple Patch-aware Dehazing Network (MPADN), to remove haze in real-world images fast and efficiently. Firstly, we design a Multiple Patch-aware Module (MPAM), which utilizes joint decisions from multiple patch awareness to gain more stable local features. Compared with the multi-patch hierarchical strategy, MPAM extremely compresses the size of the model. Besides, we propose a novel data enhancement method called Concentration Sampling Enhancement (CSE), which generates new training samples by haze concentration sampling based on hazy images and clear images. This method makes use of the precious limited real-world paired data to generate more potentially usable data. It is easy to extend CSE to more models or tasks where the ground truth and the input information are similar. Experiments show that MPADN outperforms state-of-art fast dehazing methods, and it only takes 0.0065s to process a 1600x1200 image.

Index Terms— image dehazing, multiple patch-aware, fast, data enhancement

1. INTRODUCTION

Image dehazing is a critical visual task to recover clear images from hazy ones. Because many advanced computer vision tasks[1-3] are trained on clear image data and will essentially degenerate with distinct hazy noise, image dehazing is a feasible solution to deal with this problem.

Dehazing methods are mainly based on prior statistical algorithms and learning-based algorithms. Many of these algorithms are based on atmospheric scattering models[4]:

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

where I is a hazy image, J is a haze-free image, t is the medium transmission map, A represents the global atmospheric light, and x represents the spatial position of the pixel. Eq. (1) indicates that image dehazing is a severely ill-posed problem, which contributes to recover J when t and A are both unknown.

Algorithms based on the statistical prior[5, 6] have to face the problem that there are instances that deviate from the primary data distribution. For example, [5] has no effect on high brightness places, such as the sky. Moreover, due to multiple approximate estimations, the obtained transmission map is not accurate, so the quality of the restored haze-free image is unstable.

*Corresponding author: zhang-j@foxmail.com. This research is supported by Natural Science Foundation of China under Grant 61801288. Any findings and recommendations or conclusions in this paper belong to our own and are not necessarily reflect the views of our sponsors.

The learning-based algorithms directly learn the mapping relationship from the image to the medium transmission map or the haze-free image. Thus, learning-based algorithms avoid the limitations of traditional algorithms based on the statistical prior and may even not rely on atmospheric scattering models. [7, 8] learn the mapping relationship between the hazy image and the medium transmission map. Recent works[9-14] often directly learn the end-to-end transformation from hazy images to haze-free images. [9] reformulates atmospheric scattering model and reconstructs haze-free images via an end-to-end CNN. [10] uses a generative adversarial network(GAN) for dehazing, which does not require paired training data. [11] develops smooth dilation technique and a gated context aggregation to integrate cascading different levels of features. [12] introduces a multi-scale model based on the attention mechanism, which effectively alleviates the bottleneck of the multi-scale structure. [13] employs three sub-nets for reconstruction, high-frequency details, and the haze density map. [14] maintains the input resolution and applies the cascade fusion based on the attention mechanism.

Although dehazing algorithms based on deep learning have achieved high-grade results, most recently proposed model structures are pretty complex. Too many parameters usually lead to inefficiency and sluggish speed. It is hard to interface with many advanced computer vision tasks that emphasize real time, such as semantic segmentation, object detection and tracking. In addition, the training data that drives the dehazing network is synthesized with homogenous haze in most cases, so the dehazing effect is often unsatisfactory on a real non-homogenous hazy image. Our contributions are summarized below:

- In this paper, we propose a Multiple Patch-aware Dehazing Network (MPADN) for fast and effective real-world image dehazing.
- This paper designs a Multiple Patch-aware Module (MPAM), which utilizes the patch-aware decision-making based on various patches to gain smooth local features and fuses them by embedded channel attention.
- This paper proposes a novel data enhancement method called Concentration Sampling Enhancement (CSE) that generates new training samples by haze concentration sampling. Experiments show that it has broad applied value for real-world image dehazing.

2. PROPOSED METHOD

2.1. Multiple Patch-aware Dehazing Network (MPADN)

Multi-patch Hierarchical Strategy The multi-patch hierarchical network structure[15, 16] assists the generation of features in a larger grid with the local features learned in the local restricted grid, and it realizes image restoration by bottom-up iterative fusion. Based on the multi-patch hierarchical strategy, [16] achieves the fastest speed in real-world nonhomogeneous images dehazing[17] and performs competitively in image recovery quality.

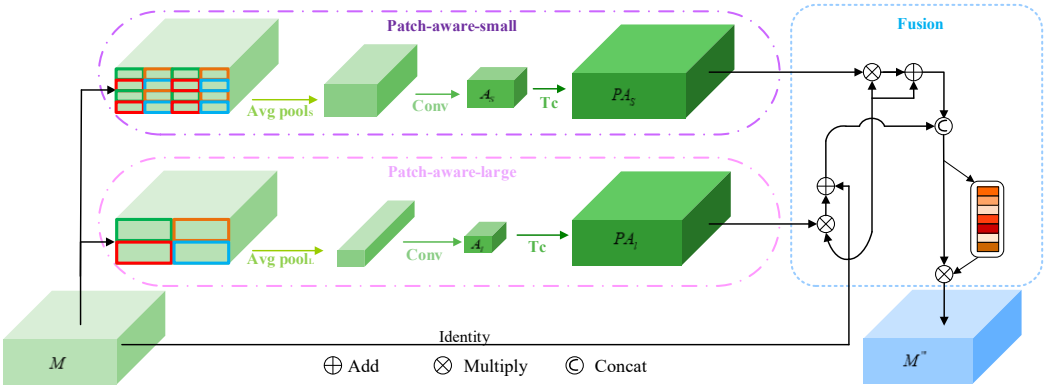

Fig. 1. The structure diagram of Multiple Patch-Aware Module (MPAM).

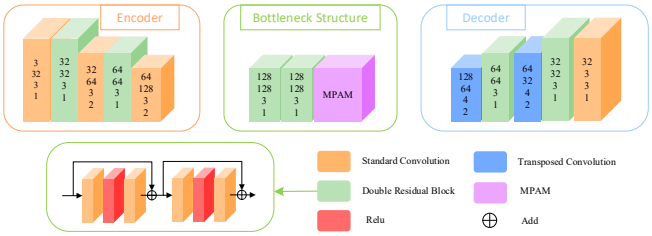

Fig. 2. The architecture of Multiple Patch-aware Dehazing Network (MPADN). The parameters on the encoder and the decoder are the number of input channels, the number of output channels, the size of the convolution kernel and the step size.

In order to limit the receptive field in the patch block, the multi-patch hierarchical strategy cuts the input image based on different grids. In this way, the same input information needs to be repeatedly encoded with the multiple hierarchical structures for fusion. In addition, this forced division in the spatial domain is artificially added, which results the grid effect in restored images. The expansion of this hierarchical structure effectively enhances the restoration quality of the image, but the time and space resources required for the hierarchy expansion increase exponentially.

MPAM In order to solve above problems, we propose the MPAM that directly acts on the encoded feature map. As shown in Fig. 1, the original feature M is spatially divided into patches of different sizes for patch-aware decision-making at different levels. As shown in the Eq. (2):

$$\begin{cases} A_i = Conv(avgpool_i(M)) \\ PA_i = \sigma(Tc(Re(A_i))) \end{cases}, \tag{2}$$

where A_i represents compressed feature weights, $Conv$ represents 1×1 convolution, $avgpool_i$ represents average pooling of i level patch size, M represents original features, PA_i represents patch-aware weights, Tc represents transposed convolution, σ and Re represent Sigmoid and ReLU function, respectively. First of all, we divide the patch-aware area into Patch-aware-small and Patch-aware-large by average pooling with different scales, which is a fixed operation with good robustness. The larger pooling patches have the larger receptive field, so it can perceive a larger local area to make decisions. Because the nonhomogenous hazy block is locally affected by the gradual haze concentration rather than the abrupt spatial distance, local features tend to be smooth and feature representations will benefit from suitable extraction mode of local features. We reduce computational overhead for further inference via 1×1 convolution. Then, transposed convolution is applied to restore the feature map to the original scale. Furthermore, M is multiplied by PA_i to extract the residual features, which is then added to M to generate the patch-aware features $M_i^{'}$:

$$M_i^{'}=PA_i \otimes M + M, \tag{3}$$

where $\otimes$ represents element-wise multiplication, $M_i^{'}$ represents patch-aware features. We combine multiple patch-aware decision-making by concatenating channel-wise features together, and get M^{*}:

$$M^{*} = \bigcup_{i=1}^{n} M_i^{'} . \tag{4}$$

Based on the patch awareness of different sizes, the decision-making of $M_i^{'}$ is diverse and is suitable for different local haze distribution because it takes into account different spatial scales. We further enhance the compelling features and suppress useless features based on the channel attention mechanism[18], to provide high-quality initial features for the subsequent restoration:

$$M^{*} = FC_2(FC_1(avgpool_g(M^{*}))) \otimes M^{*}, \tag{5}$$

where $avgpool_g$ represents global average pooling, FC_1 and FC_2 represent fully connected layers.

MPADN Based on the MPAM, a lightweight network with single-hierarchy called MPADN is designed. As shown in Fig. 2, at the end of the encoder, we add a single MPAM to realize the gradual fusion of local features.

Unlike the multi-patch hierarchical strategy cutting original images, the local receptive field's division of MPADN is based on encoded features. Accordingly, network complexity is attenuated exponentially without hierarchical expansion. In addition, multiple patch awareness is based on the pooling of different sizes. The multiple patch-aware weights based on encoded features are smooth and consistent without feature boundaries from the mandatory division, which alleviates the grid effect. Hence, MPADN has pretty fast speed and considerable image recovery capability.

2.2. Concentration Sampling Enhancement （CSE）

Acquisition cost of paired data in the real world is quite high, which leads to their scarcity. The generalization ability of learning-based algorithms is usually positively correlated with the number of practical training samples. Therefore, making full use of the real-world paired images boosts network performance effectively.

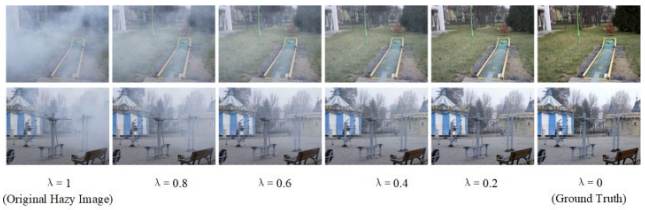

Fig. 3. Images generated by CSE with different λ .

We explore and propose an image generation and enhancement method called CSE, which can effectively utilize the hidden distribution between the hazy domain and the clear domain. More available data can result in a more robust model for data-driven methods, especially when the training data is only a few.

In previous works, only original hazy images are inputted. However, in a robust system, the hazy image should be mapped to the clear image, the clear image should also be mapped to the clear image, so the zone between original hazy images and clear images should also realize the corresponding mapping. CSE based on this hypothesis makes the model more robust and brings more potentially available data to real-world datasets.

CSE explores the hidden domain between the original hazy images and the clear images. The training set S based on CSE can be expressed as:

$$\begin{cases} I(\lambda) = \lambda I + (1-\lambda)J \\ \lambda \sim Sample(\beta, 1), \beta \in [0, 1], \\ S = \sum I(\lambda) \end{cases} \tag{6}$$

where $I(\lambda)$ represents the image generated by CSE, $Sample()$ represents the sampling mode, β is the lower bound of haze concentration whose range is between 0 and 1. λ is the haze concentration (the ratio of the input image in the generated image), which is sampled between β and 1. The sum of the ratio of the original hazy image and the clear image is limited to 1. Accordingly, the generated image is between the original hazy image and the clear image, and will not exceed the pixel representation interval. Since $I(\lambda)$ is obtained by concentration sampling in a continuous distribution, countless images between I and J are generated for training.

Fig. 3 shows some examples generated by CSE. As shown in Fig. 3, generated images based on CSE are as smooth and realistic as the real-word nonhomogeneous hazy images, which are almost consistent with real ones in human eyes.

3. EXPERIMENTS

3.1. Dataset

In order to verify the effectiveness of our algorithm, we conduct extensive ablation and comparison experiments on NH-HAZE[19], which is a real-world nonhomogeneous hazy data set. The public part of this high-resolution dataset includes 55 pairs of images in NH-HAZE20 and 25 pairs of images in NH-HAZE21. In order to conduct a fair generalization test, we use the original validation and test set in NH-HAZE20, and five pairs of images in NH-HAZE21 for testing and evaluation. In addition, we compare various state-of-art methods on Dense-Haze[20]. The validation and test sets are used to evaluate the model's performance, and the rest are used for training.

3.2. Experiment setting

Table 1. Quantitative comparison on NH-HAZE.

Algorithm	PSNR	SSIM	Runtime/s
DCP[5]	11.45	0.5259	>1(CPU)
AOD[9]	16.08	0.6298	**0.0015**
GCA[11]	16.72	0.6587	0.2316
GDN[12]	18.63	0.7260	0.0274
DMPHN[16]	19.23	0.7414	0.0261
ours	**19.86**	**0.7538**	0.0065

Table 2. Quantitative comparison on Dense-Haze.

Algorithm	PSNR	SSIM	Runtime/s
DCP[5]	11.3	0.3812	>1(CPU)
AOD [9]	13.32	0.4559	**0.0015**
GCA[11]	13.31	0.4195	0.2316
GDN[12]	14.45	0.4847	0.0274
DMPHN[16]	14.29	**0.4879**	0.0261
ours	**14.47**	0.4870	0.0065

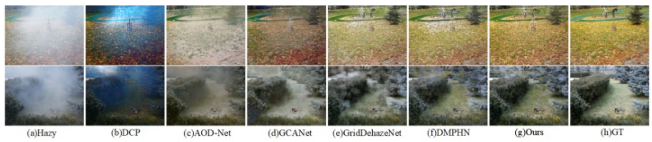

Fig. 4. Qualitative comparison on NH-HAZE.

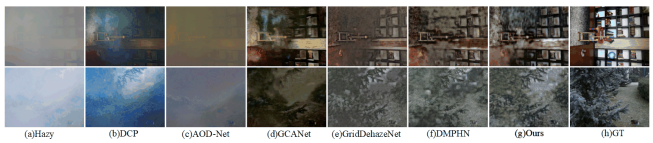

Fig. 5. Qualitative comparison on Dense-Haze.

Training details This paper is based on the PyTorch deep learning framework and the linux16.04 operating system. We train and test our models with one RTX2080. Adam optimizer is applied. The learning rate gradually decays, which is expressed as:

$$Lr = l \times arc\cos((\frac{2epoch - EPOCH}{EPOCH})^3) / \pi, \tag{7}$$

where Lr represents the learning rate, l represents the initial value of the learning rate, $epoch$ represents the current training batch and $EPOCH$ represents the total training batch. We construct a learning rate curve by Eq. (7) that decays rapidly at the beginning and the end, and changes gently in the main part. l is initialized to 0.0001 in our work.

Loss function and evaluation Our loss function includes L_1 loss and L_2 loss for image reconstruction and the inverse of SSIM for structural similarity restoration:

$$L_H = \varpi_1 L_1 + \varpi_2 L_2 + \frac{\varpi_3}{SSIM}, \tag{8}$$

where ϖ_1 、 ϖ_2 and ϖ_3 represent weights of loss function and are set to 0.6、0.4 and 0.01 based on tuning experiments. We use PSNR and SSIM[21] to evaluate the quality of image restoration and use the running time of a single 1600x1200 image inferred on the GPU to evaluate the running speed.

3.3. Results

Table 3 Ablation study results of MPADN.

Algorithm	MPAM			PSNR	SSIM	Rt/s
	S	P_S	P_L			
[16]				19.28	0.7376	0.0261
B				18.46	0.7221	**0.0051**
B+SE	√			19.16	0.7299	0.0056
B+P_S		√		19.32	0.7362	0.0060
B+P_L			√	19.26	0.7353	0.0061
B+P_S+P_L+C		√	√	19.23	0.7403	0.0065
B+P_S+P_L		√	√	19.15	0.7319	0.0065
ours	√	√	√	**19.48**	**0.7459**	0.0065

Note: B, P_S, P_L, C, S, Rt denotes Baseline, Patch-aware-small, Patch-aware-large, SE, Runtime respectively.

Table 4 Ablation study results of CSE with different parameters. Original training strategy without CSE can be seen as a special case of CSE that β is equal to 1.

CSE	PSNR	SSIM
$\beta = 1$	19.48	0.7459
$\beta = 0.9$	19.57	0.7463
$\beta = 0.8$	19.83	0.7497
$\beta = 0.7$	**19.86**	**0.7538**
$\beta = 0.6$	19.65	0.7512
$\beta = 0.5$	19.69	0.7507

Comparison of different methods on NH-HAZE Compared with various methods on NH-HAZE, in the test results, both PSNR and SSIM of ours achieve higher results. The quantitative results are shown in Table 1. DCP[5] is tested on the CPU, and the rest are performed on the GPU. The comparison of the visualization results is shown in Fig. 4. It can be seen that the haze-free image restored by the MPADN is better in terms of local texture and contrast than other methods. Compared with [16], which achieves the fastest dehazing speed in the NH-HAZE competition, ours only needs about a quarter of its inferring time to get better quantitative and qualitative evaluation performance. Specifically, PSNR increases by 0.61db, and SSIM increases by 0.0124. Ours has better performance in fast dehazing methods, whether it is quantitative or qualitative evaluation.

Comparison of different methods on Dense-Haze On Dense-Haze, the effectiveness of ours is further evaluated. Table 2 shows the comparison results of different methods on Dense-Haze. Compared with [16], MPADN gets considerable performance, with higher PSNR, lower SSIM, and four times inferring speed. Fig. 5 shows the visual contrast after dehazing. Compared with other methods, the texture quality and color saturation of ours is better.

3.4. Ablation study

Based on NH-HAZE, we conduct extensive ablation experiments on MPADN and CSE.

Ablation study of MPADN As shown in Table 3, compared with the fast dehazing method based on [16], PSNR of our baseline is 0.82db lower than it and SSIM of our baseline is 0.0155 lower than it. With the MPAM, our final network structure (MPADN) is even slightly better than [16]. Our PSNR increases by 1.5%, SSIM improves by 1.1%, and the time to process images is only 25% of it.

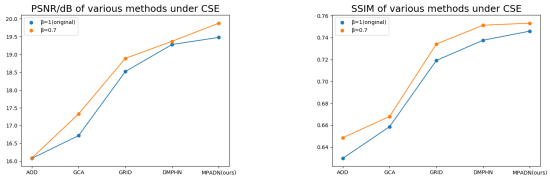

Fig. 6. PSNR and SSIM of various methods with CSE on NH-HAZE.

Our basic network structure is lightweight and single hierarchical, the running time has reached an outstanding level. However, the image restoration effect is still much inferior to [16]. In order to make the encoded features be aware of surrounding haze distribution to make more accurate weight decisions, a single patch-aware-small module is added. Then, the network can achieve a recovery effect equivalent to [16], with the running time increased a little. Similarly, after adding a single patch-aware-large module, the network achieves a similar effect as adding a patch-aware-small module. We further fuse patch-aware features of different scales.

Compared with addition, which limits representation, we concatenate channel-wise patch-aware features of different scales to fuse. Ablation experiments show that this fusion method consumes a little processing time to boost the model's representation.

Ablation study of CSE Random sampling is applied for λ in Eq. (6). We respectively enumerate the performance changes of our model and other methods with or without CSE.

As shown in Table 4, although only a simple implementation of CSE is carried out, our model's performance benefits a lot. As new data is added, the performance of the model first improves, but as the added data moves away from the source domain, the performance degrades. In the optimal situation, when β is equal to 0.7, PSNR increases by 1.95%, SSIM increases by 1.06%. The restored image quality has been dramatically improved without any changes for the network. We further verify the effectiveness of CSE for various models with β equal to 0.7.

The impact of CSE on various models on NH-HAZE Fig. 6 shows that after employing CSE (β=0.7), the performance of various methods has been improved. It is not difficult to see that data is a critical bottleneck that restricts the model's capabilities. CSE directly exploits the potential information between the hazy image and the clear image to enhance the model's performance without additional cost.

4. CONCLUSION

In order to fast and effectively remove the haze in real-world images, we propose the MPADN. Based on the MPAM, MPADN only costs 0.0065s to process a 1600x1200 image with a single RTX2080 GPU, which is relatively fast and achieves a considerable image recovery effect. In addition, we proposed CSE, which utilizes original hazy images and clear images to generate images of various haze concentration to exploit the potential paired data of real-world images. CSE not only significantly improves the performance of MPADN, but it is also effective for other models. However, the impact of different sampling methods on performance remains to be explored. Experiments demonstrates that our method outperforms state-of-art fast dehazing methods.

5. REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans Pattern Anal Mach Intell, pp 1137-1149, 2017.

[2] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp 640-651, 2017.

[3] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp 121-135, 2019.

[4] E. McCartney and F. Hall, "Optics of the Atmosphere: Scattering by Molecules and Particles," New York, John Wiley and Sons, Inc.,1976. 421 p., 1976.

[5] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Miami, pp 1956-1963, 2009.

[6] Q. Zhu, J. Mai, and L. Shao, "A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior," IEEE Transactions on Image Processing, pp 3522-3533, 2015.

[7] 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, pp 5187-5198, 2016.

[8] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single Image Dehazing via Multi-scale Convolutional Neural Networks," European Conference on Computer Vision(ECCV), Springer, Amsterdam, pp 154-169, 2016.

[9] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-One Dehazing Network," 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, pp 4780-4788, 2017.

[10] J. Zhao et al., "DD-CycleGAN: Unpaired image dehazing via Double-Discriminator Cycle-Consistent Generative Adversarial Network," Engineering Applications of Artificial Intelligence, pp 263-271, 2019.

[11] D. Chen et al., "Gated Context Aggregation Network for Image Dehazing and Deraining," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, Waikoloa Village, pp 1375-1383, 2019.

[12] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing," 2019 IEEE International Conference on Computer Vision (ICCV), IEEE, Seoul, pp 7313-7322, 2019.

[13] J. Liu, H. Wu, Y. Xie, Y. Qu, and L. Ma, "Trident Dehazing Network," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Seattle, pp 1732-1741, 2020.

[14] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature Fusion Attention Network for Single Image Dehazing," AAAI Conference on Artificial Intelligence, AAAI, New York, pp 11908-11915, 2020.

[15] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep Stacked Hierarchical Multi-Patch Network for Image Deblurring," 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Long Beach, pp 5971-5979, 2019.

[16] S. D. Das and S. Dutta, "Fast Deep Multi-patch Hierarchical Network for Nonhomogeneous Image Dehazing," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Seattle, pp 1994-2001, 2020.

[17] C. O. Ancuti et al., "NTIRE 2020 Challenge on NonHomogeneous Dehazing," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Seattle, pp 2029-2044, 2020.

[18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," 2018 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), IEEE, Salt Lake City, pp 7132-7141, 2018.

[19] C. O. Ancuti, C. Ancuti, and R. Timofte, "NH-HAZE: An Image Dehazing Benchmark with Non-Homogeneous Hazy and Haze-Free Images," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Seattle, pp 1798-1805, 2020.

[20] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense-Haze: A Benchmark for Image Dehazing with Dense-Haze and Haze-Free Images," 2019 IEEE International Conference on Image Processing (ICIP), IEEE, Taipei, pp 1014-1018, 2019.

[21] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, pp 600-612, 2004.