

Multi-Scale Selective Residual Learning for Non-Homogeneous Dehazing

Eunsung Jo¹ and Jae-Young Sim^{1,2}
Graduate School of Artificial Intelligence¹
Department of Electrical Engineering²

Ulsan National Institute of Science and Technology, Ulsan, Korea

esjo93@unist.ac.kr, jysim@unist.ac.kr

Abstract

As the particles in hazy medium cause the absorption and scattering of light, the images captured under such environment suffer from quality degradation such as low contrast and color distortion. While numerous single image dehazing methods have been proposed to reconstruct clean images from hazy images, non-homogeneous dehazing has been rarely studied. In this paper, we design an end-to-end network to remove non-homogeneous dense haze. We employ the selective residual blocks to adaptively improve the visibility of resulting images, where the input feature and the residual feature are combined with fully trainable weights. Experimental results including the ablation study demonstrate that the proposed method is a promising tool for non-homogeneous dehazing that enhances the contrast of hazy images effectively while restoring colorful appearance faithfully.

1. Introduction

Digital images captured with particles in the air, such as aerosols, smoke, and mist, often suffer from quality degradation due to the scattering and absorption of light as shown in Figure 1. The quality degradation in such hazy images include low contrast, reduced sharpness, and distorted colors. Single image dehazing (SID) methods have been proposed to generate a clean image with improved visibility from a single hazy image. Recently, the performance of SID methods has been improved rapidly through many challenges and competitions [4, 5, 2, 3, 1, 6, 7, 8].

To restore a clean image from a hazy image, SID methods usually employ the image formation model (IFM) given by

$$I(\mathbf{x}) = J(\mathbf{x})e^{-\beta d(\mathbf{x})} + A(1 - e^{-\beta d(\mathbf{x})}), \quad (1)$$

where $I(\mathbf{x})$ and $J(\mathbf{x})$ denote the observed and original intensity at the pixel location \mathbf{x} , respectively. A is the atmospheric light, $d(\mathbf{x})$ is the scene distance at \mathbf{x} , and β is the

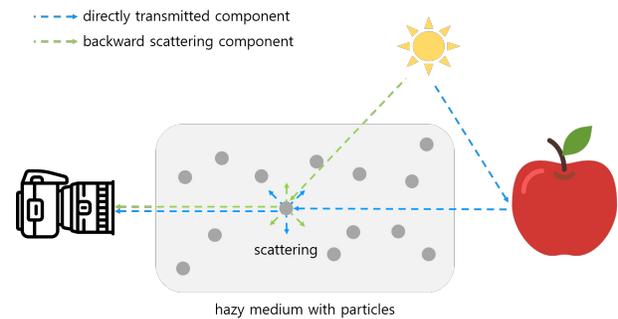


Figure 1: Light scattering in the air with particles.

light attenuation coefficient. The transmission of $e^{-\beta d(\mathbf{x})}$ represents the light conservation rate along the line-of-sight at \mathbf{x} . Most of the conventional SID methods attempt to estimate the unknown parameters of A and $e^{-\beta d(\mathbf{x})}$ to derive J from I . In typical SID methods targeting homogeneous haze, it is usually assumed that A and β are constant in an entire image and the amount of haze depends on the scene distance $d(\mathbf{x})$ only. However, these assumptions do not hold in real situations where a hazy image often exhibits locally different haze densities associated with different values of A and β . More recently, attempt has been made to address this problem via non-homogeneous dehazing challenges [3, 7, 8].

In this paper, we propose an end-to-end multi-scale dehazing network for non-homogeneous dense haze, which employs selective residual blocks where the input feature and the residual feature are combined with different learnable weights. By selecting optimal weighting parameters via network training, we can effectively balance the contributions of the input feature and the residual feature to dehazing. Experimental results show that the proposed network effectively removes dense and non-homogeneous haze from a single image and generates a clear image with improved visibility.

2. Related Work

2.1. Prior Knowledge based Methods

There have been many SID methods using prior knowledges. For example, Tan [25] maximized the contrast of hazy images with a prior that a clean image tend to show high contrast in general. He et al. [11] estimated the unknown parameters in Eq. 1 based on a statistical prior, called dark channel prior (DCP), assuming that at least one of the pixels in a patch has the intensity value of 0 among the three color channels. Zhu et al. [30] estimated the scene depth for dehazing using a color attenuation prior which shows the correlation between the brightness and the saturation. Berman et al. [9] proposed a non-local prior based on the observation that the pixels in a same cluster are distributed along a haze-line. These methods provide good results for homogeneous haze, but have limited performance on *non-homogeneous* haze due to the gap between the image formation model and real situations.

2.2. Deep Learning based Methods

Deep learning based SID methods have been proposed [19, 10, 15, 27, 21, 20, 18, 26]. Cai et al. [10] proposed a dehazing model which estimates the transmission maps of IFM. Zhang et al. [27] used a densely connected pyramid network to estimate the transmission map with adversarial learning. Li et al. [15] reformulated IFM to train deep networks.

On the other hand, end-to-end network architectures and their training models were designed in a data-driven way without prior knowledge [21, 19, 20, 18, 26]. Liu et al. [18] proposed an end-to-end dehazing network composed of three sub-networks. The first sub-network refines the detail of an input hazy image. The second sub-network estimates the hazy density map in an unsupervised way. The last sub-network reconstructs coarse level features of the input image. Wu et al. [26] employed knowledge transfer to construct a model composed of two networks: the teacher network that auto-encodes the clean images and the student network designed to perform dehazing. The feature level knowledge transfer loss also helps the teacher network to guide the student network to learn the information needed to reconstruct clean images. Attention-based models were also proposed to flexibly learn the weights of their models. Ren et al. [21] fused gated features from three enhanced versions of the input, and used adversarial learning to render the dehazed outputs. Liu et al. [19] used channel-wise feature attention to fuse different features together. Qin et al. [20] exploited a feature attention module that fuses the features with pixel attention and channel attention.

3. Proposed Method

We propose a non-homogeneous dehazing network based on the selective residual block. The overall network architecture of the proposed method is shown in Figure 2.

3.1. Network Architecture

Multi-Scale Input and Output In order to effectively exploit the local and global characteristics of hazy images, an input hazy image I_0 is first down-sampled to three smaller images of I_1 , I_2 , and I_3 with $1/2$, $1/4$, and $1/8$ sizes, respectively, which are then fed to the network as shown in Figure 2. The deep intermediate features from the multi-scaled input are fused in the bottleneck. The bottleneck feature is then fed to the decoding blocks with deep intermediate features processed from multi-scaled input. Finally, the model generates multi-scale output images of O_0 , O_1 , O_2 , and O_3 to compute the losses at multi-scale images, respectively, which are then combined together to compute the final loss.

Selective Residual Block Inspired by [17], we devise the selective residual block (SRB) as a core functional module of the proposed network. As illustrated in Figure 2, SRB has two 3×3 convolutional layers where each convolutional layer is followed by the normalization layer. We adopt the concept of the residual block [12] that efficiently extracts the features without causing the gradient vanishing problem. Let the input feature map of SRB be $F(\mathbf{x})$ and the estimated residual feature be $R(\mathbf{x})$. The output of SRB, $O(\mathbf{x})$, is the activation result of the weighted sum of the input feature and the residual feature given by

$$O(\mathbf{x}) = \sigma(\alpha F(\mathbf{x}) + \beta R(\mathbf{x})), \quad (2)$$

where $\sigma(\cdot)$ is the activation function. Note that α and β are trainable parameters in the proposed method. Whereas the conventional residual blocks use scaled values of certain constants as the weights [24, 16, 18], SRB selectively takes the balance between the input and residual features by employing fully trainable weights. The effect of using SRB is illustrated in Figure 3. We see that $F(\mathbf{x})$ yields grid artifacts that can be accumulated by the skip connection and degrades the quality of resulting images. However, by taking the weighted sum of $F(\mathbf{x})$ and $R(\mathbf{x})$ with trained α and β , the activated output feature $O(\mathbf{x})$ shows a smoothed and visually pleasing result.

Encoding Block The encoding block has a series of two SRBs and the 3×3 convolutional layer with stride of 2, which downsamples the feature map into the $\frac{1}{2}$ resolution map, followed by ReLU activation layer.

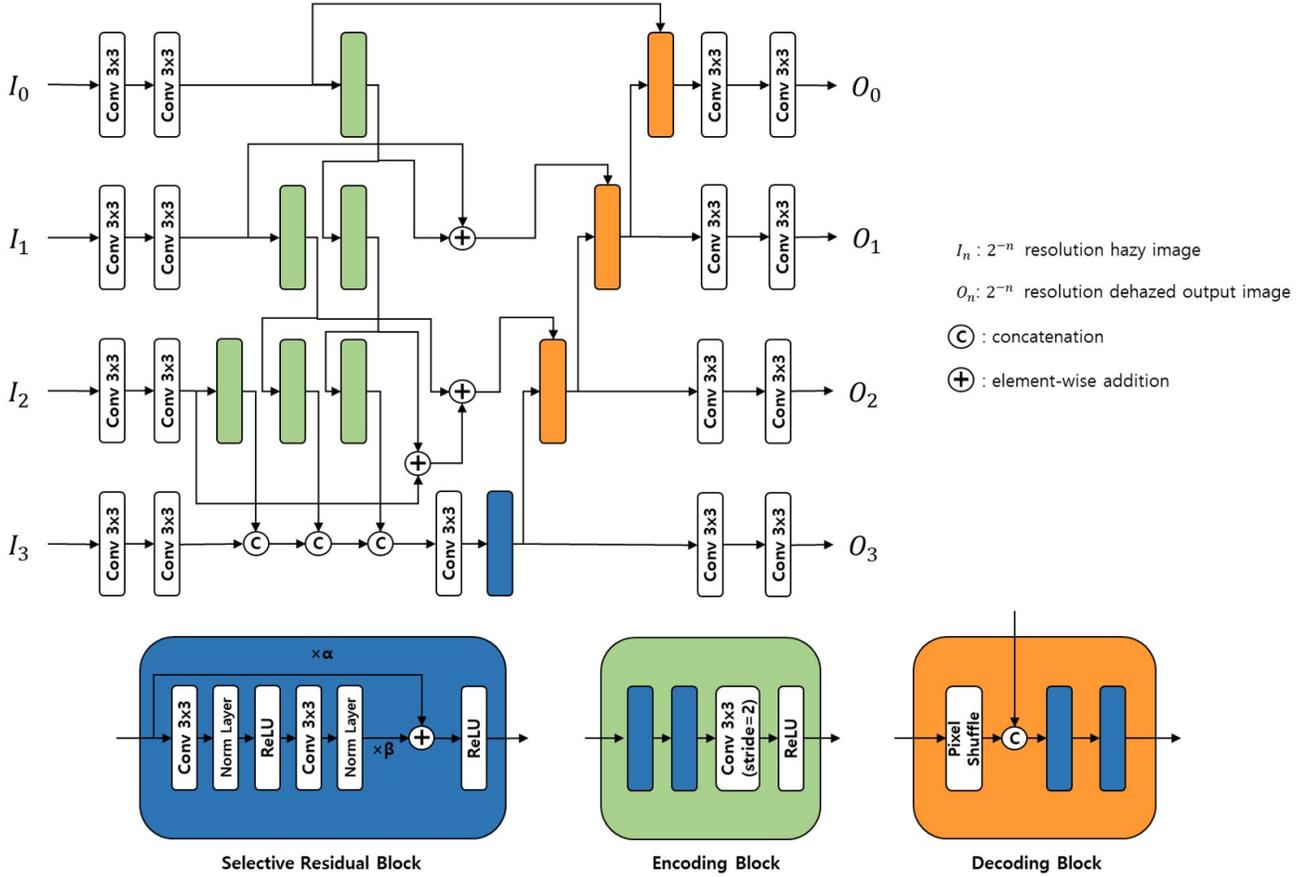


Figure 2: Network architecture of the proposed method.

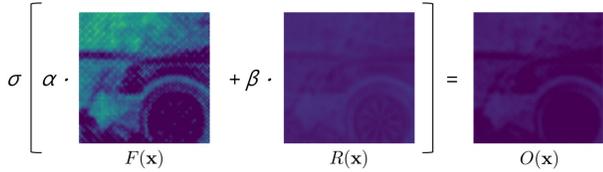


Figure 3: Behavior of selective residual block (SRB). The unwanted artifacts shown in the input feature map $F(\mathbf{x})$ are alleviated in the final output feature map $O(\mathbf{x})$.

Decoding Block The main purpose of the decoding block is up-sampling of the intermediate features. We employ the pixel shuffle layer [22] with the scale factor of 2 to avoid the checkerboard artifacts during up-sampling. The up-sampled feature is concatenated with the feature transferred via the shortcut from the encoding part which is then forwarded to two different SRBs. In addition, SRB takes the skip connection of $F(\mathbf{x})$ selectively to alleviate the uneven artifacts that can be derived from the feature relocation in the pixel shuffle layer.

3.2. Loss Function

We define a total loss function composed of three losses: reconstruction loss, perceptual loss [13], and structural similarity loss. Note that our network returns the multi-scaled outputs, and we compute the loss functions at each image scale, respectively. Finally, the loss functions at multi-scales are combined together with different weights to compute the total loss.

Reconstruction Loss We minimize the reconstruction loss $\mathcal{L}_{\text{recon}}(X, Y)$, that is the L_1 loss or mean absolute error between the reconstructed output image X and the ground truth Y for an input hazy image, given by

$$\mathcal{L}_{\text{recon}}(X, Y) = \frac{1}{3N} \sum_{\mathbf{x}} \sum_c \|X_c(\mathbf{x}) - Y_c(\mathbf{x})\|_1, \quad (3)$$

where N is the number of pixels in an image and $X_c(\mathbf{x})$ and $Y_c(\mathbf{x})$ denote the intensity values of X and Y at the pixel \mathbf{x} in c channel.



Figure 4: Example images in the datasets used in training. The images in the first and third rows show hazy images, and the second and fourth rows show the corresponding ground truth images.

Perceptual Loss The perceptual loss [13] has been widely used to minimize the perceptual differences between two image in style transfer applications. In practice, the intermediate features of the pre-trained VGG network [23] are used to measure the perceptual difference between two images. Let the activated intermediate feature tensor of the pre-trained VGG network in the k -th layer extracted from X be $\phi_k(X)$. The perceptual difference $\mathcal{L}_{\text{perc}}(X, Y)$ between X and Y is defined as

$$\mathcal{L}_{\text{perc}}(X, Y) = \sum_k \frac{1}{C_k H_k W_k} \|\phi_k(X) - \phi_k(Y)\|_1, \quad (4)$$

where C_k is the number of channels and H_k and W_k are the height and width of the feature tensor in the k -th layer, respectively. We set $k \in \{1, 3, 5, 9, 13\}$.

Structural Similarity Loss We also employ the structural similarity loss $\mathcal{L}_s(X, Y)$ between X and Y by computing SSIM[29] as

$$\mathcal{L}_s(X, Y) = 1 - \text{SSIM}(X, Y). \quad (5)$$

Total Loss Function We define the total loss function \mathcal{L}_n at the n -th image scale with $1/2^n$ size by combining the three loss functions together, given by

$$\mathcal{L}_n = \mathcal{L}_{\text{recon}}(O_n, J_n) + \lambda_1 \mathcal{L}_{\text{perc}}(O_n, J_n) + \lambda_2 \mathcal{L}_s(O_n, J_n), \quad (6)$$

where λ_1 and λ_2 are the weighting parameters. We empirically set 0.5 for both λ_1 and λ_2 . The final loss function $\mathcal{L}_{\text{total}}$ is computed as

$$\mathcal{L}_{\text{total}} = \sum_{n=0}^3 \frac{1}{2^n} \mathcal{L}_n \quad (7)$$

by combining the four multi-scale loss functions.

4. Experimental Results

4.1. Datasets

During the 2021 NTIRE Non-Homogeneous dehazing challenge [8], a dataset of 35 different scenes were given. The pairs of a hazy image and its ground truth clear image for 25 scenes were used as training data and only the hazy images for the remaining 10 scenes were used as validation data (5 scenes) and test data (5 scenes). In order to alleviate the effect of overfitting, we internally divided the given training set into a tentative training set (20 pairs) and a tentative validation set (5 pairs), respectively. In addition, we used all the previous NTIRE dehazing challenge datasets of I-HAZE [4], O-HAZE [5], DENSE-HAZE [2], NH-HAZE [3] as training data. Table 1 summarizes the numbers of scenes used in our experiments. Figure 4 also shows some example images of the used datasets.

Datasets	Scenes	Usage	
		Training	Validation
I-HAZE [4]	35	✓	
O-HAZE [5]	45	✓	
DENSE-HAZE [2]	55	✓	
NH-HAZE [3]	55	✓	
NTIRE 2021	20	✓	
NTIRE 2021	5		✓

Table 1: The datasets used in the experiments.

4.1.1 Homogeneous Haze Image Dataset

I-HAZE I-HAZE[4] is an indoor image dataset for single image dehazing. The dataset was introduced in NTIRE 2018 dehazing challenge[1].

O-HAZE O-HAZE[5] is an outdoor image dataset for single image dehazing. O-HAZE was also introduced in NTIRE 2018 image dehazing challenge[1].

DENSE-HAZE DENSE-HAZE[2] was introduced in NTIRE 2019 image dehazing challenge[6]. DENSE-HAZE includes images with thicker haze than that of I-HAZE and O-HAZE.

4.1.2 Non-Homogeneous Haze Image Dataset

NH-HAZE NH-HAZE[3] was introduced in NTIRE 2020 non-homogeneous dehazing challenge[7]. Unlike previously introduced I-HAZE, O-HAZE, and DENSE-HAZE, NH-HAZE shows different haze densities according to local image areas.

NTIRE 2021 Non-homogeneous dehazing challenge held in NTIRE 2021 [8] released an image dataset for single image dehazing. NTIRE 2021 exhibits similar appearance of haze to that of NH-HAZE[3], but includes a smaller number of scenes than that of NH-HAZE[3].

4.2. Training Details

We trained the model on a single Titan RTX. In the training, we used Adam optimizer [14] with $\beta_1 = 0.5$, $\beta_2 = 0.999$, and weight decay (L_2 penalty) coefficient $\lambda = 1 \times 10^{-4}$. We initialized the learning rate to 5×10^{-4} , and the learning rate decays in ratio of 0.5 in every 120th epoch. The batch size is 4. In each experiment, we trained the network for 3000 epochs, and picked the model with the best PSNR score among all the training epochs. The hyper-parameters we used in the training phase are specified in Table 2.

Hyper-parameter	Value
Optimizer	Adam [14]
Initial learning rate	5×10^{-4}
Batch size	4
Weight decay(λ)	1×10^{-4}
Learning rate scheduling	Step
Learning rate decaying step size	120
Training epochs	3000

Table 2: Hyper-parameter setting in training phase.

Loss weights		Block	Metrics		
λ_1	λ_2	Type	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
0	0	RB	19.5602	0.7764	0.3088
0	0	SRB	19.1577	0.7663	0.3187
0	0.1	RB	20.0502	0.8017	0.2773
0	0.1	SRB	19.7538	0.7965	0.2855
0	0.5	RB	20.1578	0.8171	0.2568
0	0.5	SRB	18.5607	0.7779	0.3793
0.1	0	RB	20.3884	0.8097	0.2365
0.1	0	SRB	20.6929	0.8122	0.2266
0.1	0.1	RB	20.3609	0.8172	0.2296
0.1	0.1	SRB	20.3363	0.8183	0.2273
0.1	0.5	RB	20.8112	0.8269	0.2253
0.1	0.5	SRB	20.3635	0.8188	0.2348
0.5	0	RB	20.5445	0.8253	0.2028
0.5	0	SRB	20.5840	0.8257	0.2149
0.5	0.1	RB	20.7056	0.8254	0.2030
0.5	0.1	SRB	21.0842	0.8295	0.2043
0.5	0.5	RB	20.6742	0.8367	0.2084
0.5	0.5	SRB	21.1680	0.8364	0.1964

Table 3: Ablation study on the quantitative performance. In each experiment, the score of the proposed method is calculated from the model that achieves the best PSNR score among the 3000 training epochs on the tentative validation set (#21~25) of 2021 NTIRE non-homogeneous dehazing challenge dataset. The scores written in red and blue denote the best and the second-best scores, respectively.

We also augmented the training data with random cropping to the size of 512×512 , random scaling with the ratios between 0.5 and 1, random identity mapping ($p = 0.1$), and random horizontal/vertical flipping with the probability of 0.5. We did not use any pre-trained model in this work.

4.3. Ablation Study

We evaluated the performance of the proposed method with different hyper-parameter settings on our tentative validation set (#21 ~ 25). The metrics used in the ablation study are PSNR, SSIM [29], and LPIPS [28]. As shown in Table 3, the experiment with $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$ with

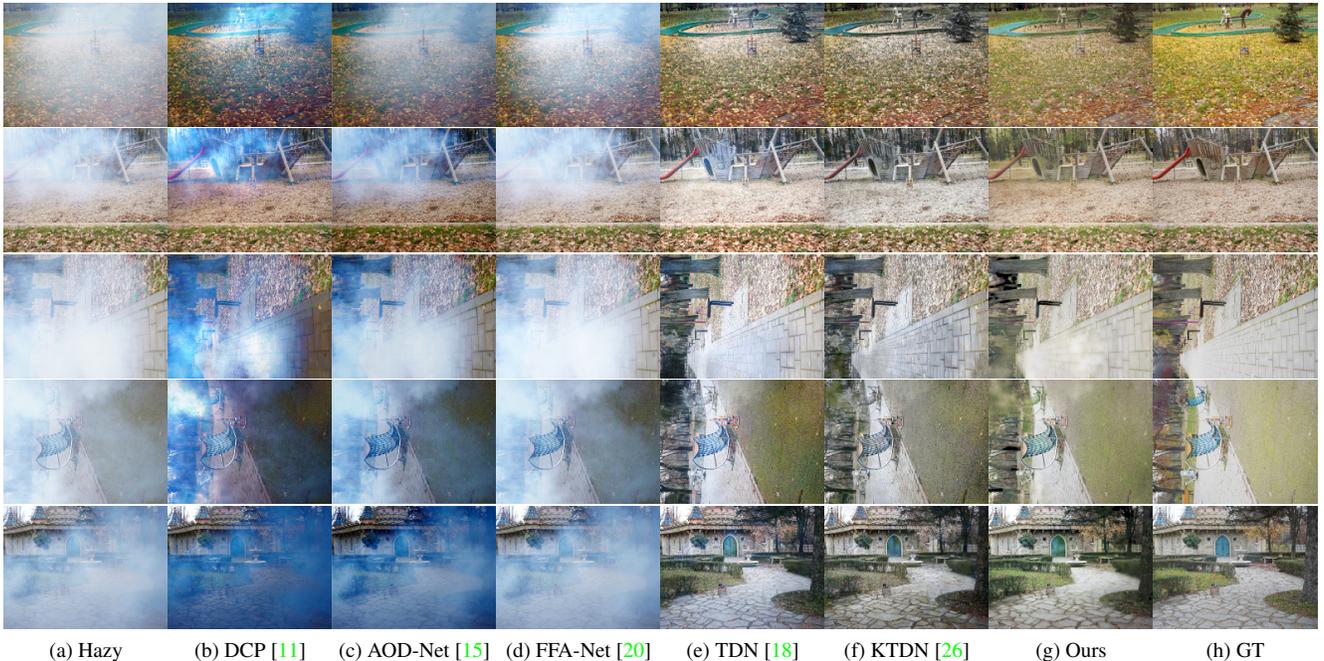


Figure 5: Qualitative comparison of dehazing results on the tentative validation set (#21~25) of 2021 NTIRE non-homogeneous dehazing challenge [8] dataset.

Methods	Metrics		
	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
DCP [11]	11.8310	0.6462	0.4483
AOD-Net [15]	12.3350	0.6312	0.4648
FFA-Net [20]	11.3724	0.6222	0.4851
TDN [18]	18.1288	0.8188	0.2756
KTDN [26]	18.8138	0.7912	0.3029
Ours	21.1680	0.8364	0.1964

Table 4: Quantitative comparison of dehazing results on the tentative validation set (#21~25) of 2021 NTIRE non-homogeneous dehazing challenge dataset. The scores in red and blue denote the best and the second-best scores, respectively.

SRB showed the best quantitative performance in terms of the fidelity related score of PSNR and the perceptual score of LPIPS.

4.4. Comparison with Existing Methods

Tentative Validation Set Figure 5 and Table 4 compare the qualitative and quantitative results obtained by using the state-of-the-art methods and the proposed method, respectively, on a tentative validation set (#21~25). Table 4 shows that [11, 15, 20] yield poor qualitative performances

since they assume the haze is homogeneous or use the images with homogeneous haze in the training phase. On the other hand, we see that the learning-based methods trained on *non-homogeneous* dataset achieve better scores in both of the fidelity-related scores and the perceptual score. The proposed method shows the best quantitative performance in terms of all the three metrics as shown in Table 4.

Figure 5 also visually compares the resulting images generated by using the proposed method and the existing methods. Previously mentioned homogeneous dehazing methods do not remove the haze completely, and often yield bluish appearance. Though TDN [18] and KTDN [26] remove the *non-homogenous* haze remarkably, their results tend to lose colorfulness or show different color tones from that of the ground truth images. In contrary, the proposed method reconstructs qualitatively good images.

Validation and Test Sets During the NTIRE 2021 image dehazing competition [8], there were no ground truth images for the hazy images in the validation set (#26 ~ 30) and the test set (#31 ~ 35). For these datasets, we conduct qualitative comparison only between the existing SID methods [11, 15, 20, 18, 26] and the proposed method. Figures 6 and 7 represent the qualitative results on the validation and test sets, respectively. The homogeneous dehazing method [11] and the learning-based method trained on the

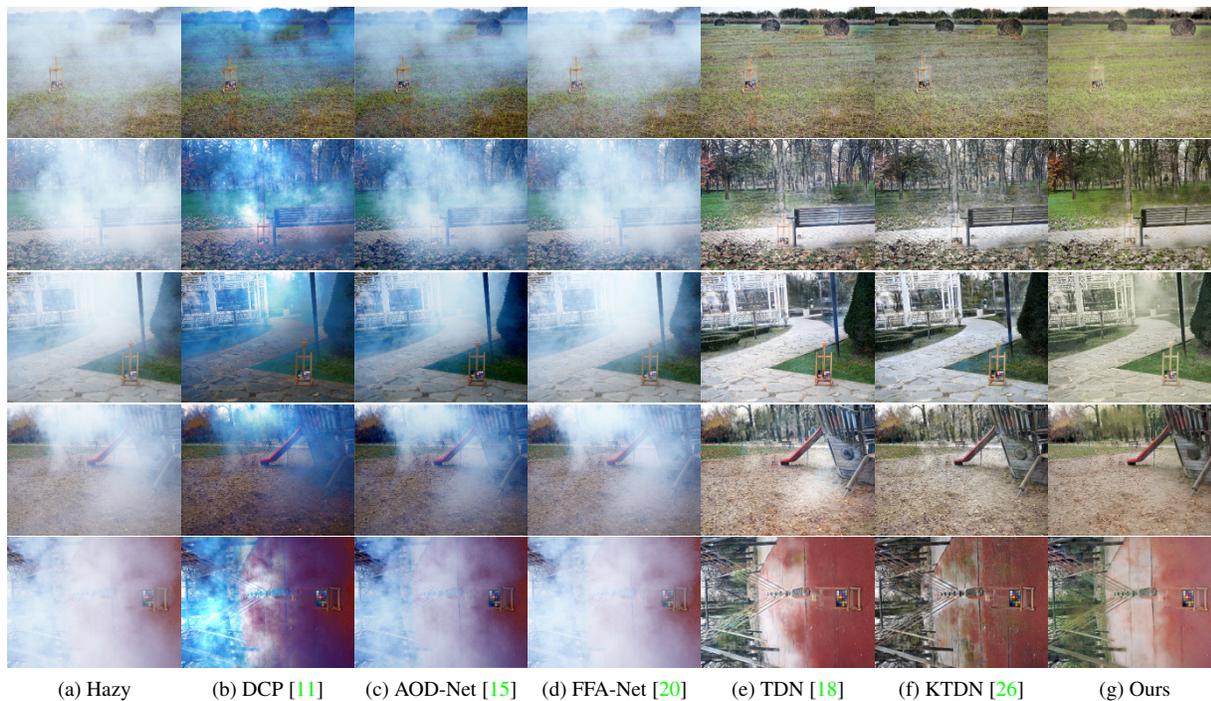


Figure 6: Qualitative comparison of dehazing results on the validation set (#26~30) of 2021 NTIRE non-homogeneous dehazing challenge dataset.

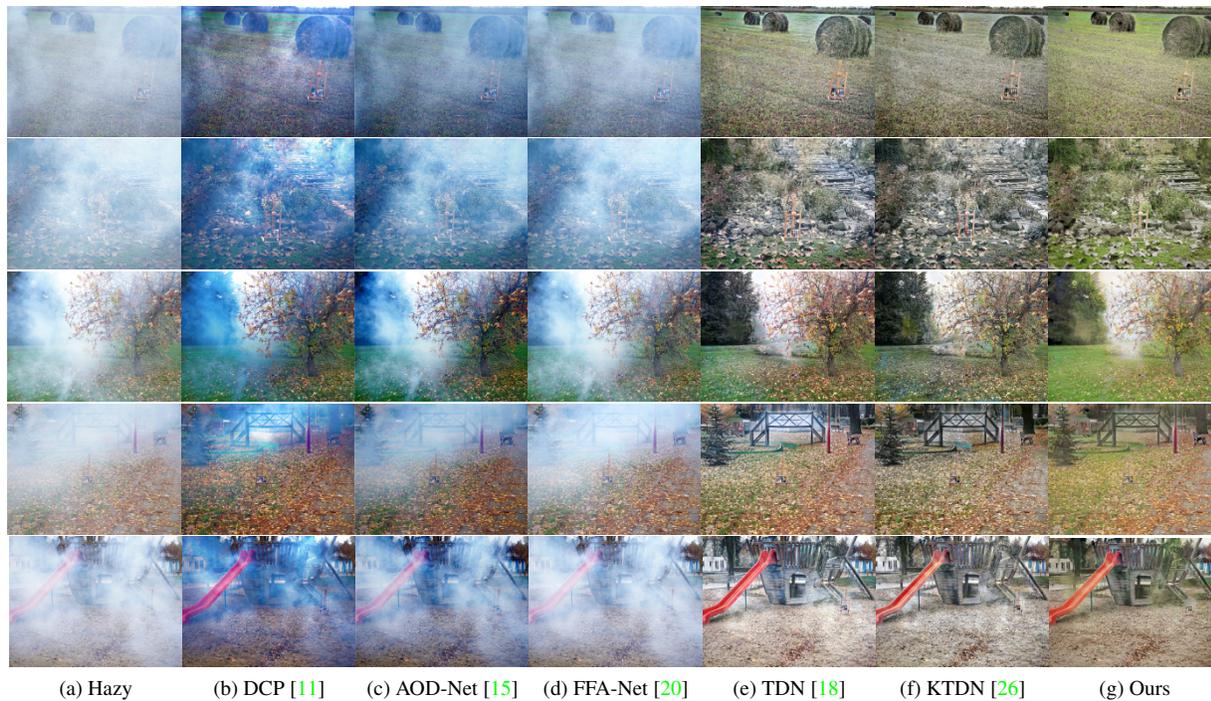


Figure 7: Qualitative comparison of dehazing results on the test set (#31~35) of 2021 NTIRE non-homogeneous dehazing challenge dataset.

synthetic homogeneous dehazing dataset [15, 20] fail to remove locally dense haze completely. The models [18, 26] trained on the *non-homogeneous* dataset [3] remove most of the haze successfully, however their results lose the colorfulness of images. Note that the proposed method not only effectively removes the *non-homogeneous* haze, but also reconstructs vivid colors.

5. Conclusion

In this paper, we proposed a multi-scale end-to-end dehazing network based on selective residual blocks. Whereas the conventional residual modules employ constrained weights, the proposed method trains the weighting parameters for the input and residual features adaptively. Without any physics-related or statistical prior knowledge, the proposed method shows good quantitative performances as well as provides visually pleasing images with vivid colors. We also demonstrated that the selective residual blocks achieve higher performance compared with the original residual blocks.

Acknowledgments

This work was supported in part by the Samsung Research Funding and Incubation Center of Samsung Electronics under Project SRFC-IT1802-08, and in part by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No.20200013360011001, Artificial Intelligence graduate school support(UNIST)).

References

- [1] C. Ancuti, C. O. Ancuti, R. Timofte, L. Van Gool, L. Zhang, et al. Ntire 2018 challenge on image dehazing: Methods and results. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 1004–1014, 2018. 1, 5
- [2] C. O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *IEEE Int. Conf. Image Process.*, pages 1014–1018, 2019. 1, 4, 5
- [3] C. O. Ancuti, C. Ancuti, and R. Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 1798–1805, 2020. 1, 4, 5, 8
- [4] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer. I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In *arXiv:1804.05091v1*, 2018. 1, 4, 5
- [5] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer. O-haze: A dehazing benchmark with real hazy and haze-free outdoor images. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 867–8678, 2018. 1, 4, 5
- [6] C. O. Ancuti, C. Ancuti, R. Timofte, L. Van Gool, L. Zhang, et al. Ntire 2019 image dehazing challenge report. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 2241–2253, 2019. 1, 5
- [7] C. O. Ancuti, C. Ancuti, F. Vasluianu, R. Timofte, et al. Ntire 2020 challenge on nonhomogeneous dehazing. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 2029–2044, 2020. 1, 5
- [8] C. O. Ancuti, C. Ancuti, F. Vasluianu, R. Timofte, et al. NTIRE 2021 nonhomogeneous dehazing challenge report. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, 2021. 1, 4, 5, 6
- [9] D. Berman, T. Treibitz, and S. Avidan. Single image dehazing using haze-lines. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(3):720–734, 2020. 2
- [10] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Trans. Image Process.*, 25(11):5187–5198, 2016. 2
- [11] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(12):2341–2353, 2011. 2, 6, 7
- [12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 770–778, 2016. 2
- [13] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Eur. Conf. Comput. Vis.*, 2016. 3, 4
- [14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent.*, 2015. 5
- [15] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *Int. Conf. Comput. Vis.*, pages 4780–4788, 2017. 2, 6, 7, 8
- [16] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 1132–1140, 2017. 2
- [17] F. Liu, M. Gao, Y. Liu, and K. Lei. Self-adaptive scaling for learnable residual structure. In *CoNLL*, pages 862–870, 2019. 2
- [18] J. Liu, H. Wu, Y. Xie, Y. Qu, and L. Ma. Trident dehazing network. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1732–1741, 2020. 2, 6, 7, 8
- [19] X. Liu, Y. Ma, Z. Shi, and J. Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Int. Conf. Comput. Vis.*, pages 7313–7322, 2019. 2
- [20] X. Qin, Z. Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI*, 2020. 2, 6, 7, 8
- [21] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M. Yang. Gated fusion network for single image dehazing. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3253–3261, 2018. 2
- [22] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1874–1883, 2016. 3

- [23] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *Int. Conf. Learn. Represent.*, 2015. 4
- [24] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, page 4278–4284, 2017. 2
- [25] R. T. Tan. Visibility in bad weather from a single image. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1–8, 2008. 2
- [26] H. Wu, J. Liu, Y. Xie, Y. Qu, and L. Ma. Knowledge transfer dehazing network for nonhomogeneous dehazing. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 1975–1983, 2020. 2, 6, 7, 8
- [27] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3194–3203, 2018. 2
- [28] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 586–595, 2018. 5
- [29] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. 4, 5
- [30] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(11):3522–3533, 2015. 2