Semantic Guidance Learning for High-Resolution Non-homogeneous Dehazing

Hao-Hsiang Yang[2], I-Hsiang Chen[2], Chia-Hsuan Hsieh[4], Hua-En Chang[2],
Yuan-Chun Chiang[2], Yi-Chung Chen[3], Zhi-Kai Huang[2], Wei-Ting Chen[1], Sy-Yen Kuo[2]

[1] Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan
[2] Department of Electrical Engineering, National Taiwan University, Taiwan
[3] Graduate Institute of Communication Engineering, National Taiwan University, Taiwan
[4] ServiceNow, USA

Abstract

High-resolution non-homogeneous dehazing aims to generate a clear image from a 4000 × 6000 image with non-homogeneous haze. To the best of our knowledge, this task is a new challenge that was not addressed in the previous literature. To address this issue, we propose semantic-guided loss functions for high-resolution non-homogeneous dehazing. We find semantic information contains strong texture and color prior. Thus, we proposed to adopt the pre-trained model to generate the semantic mask to guide the neural network during the training phase. On the other hand, to handle the non-homogeneous dehazing process in the high-resolution scenario, we adjust the kernel size of the model to increase the receptive field. Furthermore, to deal with the different image sizes during the training and the testing phase, several post-processing methods are applied to improve the high-resolution non-homogeneous dehazing. Several experiments performed on challenging benchmark show that the proposed model achieves competitive performance in the NTIRE 2023 HR NonHomogeneous Dehazing Challenge.

1. Introduction

Haze or fog is a common phenomenon because of light absorption and scattering in the atmosphere medium in our daily life, and they usually degrade the visibility of images. Furthermore, haze and fog also deteriorate the performance of high-level vision applications like autonomous driving, robot navigation, and object recognition [1]. Therefore, many researchers have endeavored to propose approaches to restoring clean photographs from hazy/foggy ones in both computational photography and vision communities in the past decades. The atmospheric scattering model [2] is used to estimate the clean image from a single hazy input and is expressed as:

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

where $J(x)$ is the captured hazy image, $I(x)$ is the corresponding clear image, $t(x)$ is the medium transmission and A is the global atmospheric light. Single-image dehazing is an ill-posed and challenging problem because multiple mapping solutions are possible from a hazy image to clear images.

Despite its ill-posedness, many efforts on estimating and regularizing the solution space using a variety of statistical and image priors [5–7] are proposed. However, these handcrafted statistical priors are designed based on specific observations, which may not be robust to deal with various situations like the unconstrained environment in the wild. Recently, many researchers proposed neural network solutions [8–10] for single image dehazing because of the superior performance of the learning-based methods and large amounts of data [11, 12]. These methods use convolutional neural networks (CNNs) to extract features and learn the mapping function between hazy and haze-free image pairs and the object loss functions are selected to optimize the network. However, these methods can not handle the high-resolution non-homogeneous dehazing well for some reasons. First, several synthetic datasets are generated based on the assumption that the atmosphere is homogeneous, and many deep learning-based solutions focus on this kind of dataset. On the contrary, non-homogeneous dehazing draws less attention. Second, the image size of the existing dataset is smaller than 4K, while the image size of the high-resolution non-homogeneous dehazing is up to 4000 × 6000. It is impossible to feed the whole image to train the dehazing model. Alternatively, patches are cropped from the training data and fed to the model. In the test phase, we can feed the whole image to the model. However, different image sizes would cause the inconsistency of feature distributions in the training and testing phase [4] and deteriorate the model's performance.

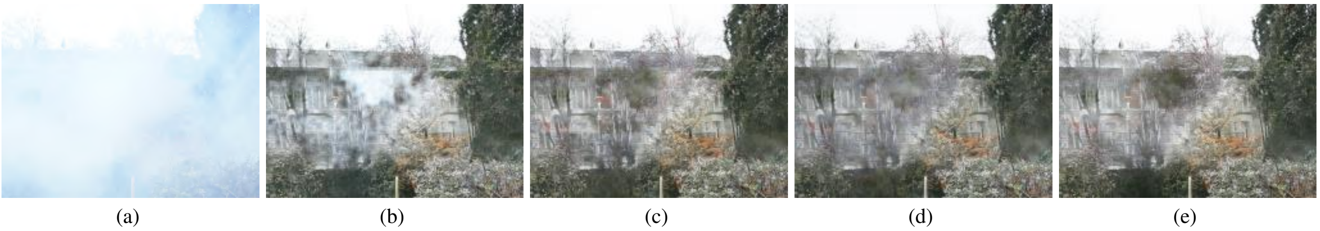

Figure 1. **Visual comparison with different non-homogeneous dehazing methods.** (a): the non-homogeneous haze image. (b): the dehazed result generated by the baseline [3] method. (c): the dehazed result with proposed semantic-guided loss functions and the model. Notably, (b) and (c) are dehazed from inference with overlapping patches and introduce visible boundary artifacts. (d): the dehazed image with the whole input image. (e): the dehazed results with post-processing of TLC [4] and test time augmentation. Please zoom in for a better visual experience.

To address these issues, we propose several strategies for high-resolution non-homogeneous dehazing. We first observe semantic segmentation can provide natural priors and high-level features, which are beneficial to guide the non-homogeneous dehazing network. Though several tasks use loss functions with high-level features [13, 14] to optimize the network, semantic information contains more color, texture, and geometry priors and is more helpful in reconstructing images. Therefore, we use the pre-trained semantic segmentation model [15] to generate semantic maps as extra guidance and utilize them to propose two semantic-based consistent loss functions.

Second, apart from the proposed loss functions, selecting a useful neural network is necessary to handle this task. We review several previous state-of-the-art solutions and select DW-GAN [3] as our backbone. The dehazed results are presented in Figure 1 (b). Compared to Figure 1 (a), DW-GAN can remove the non-homogeneous haze. However, there is residual haze in the middle of the image. The main reason is that when the haze is too dense and the image size is too large, the dehazed network cannot capture large regions with useful information. Thus, we increase the kernel size of the original model to increase receptive fields. The dehazed image by a large model is shown in Figure 1. Third, when inferring Figure 1 (b) and Figure 1 (c), images are cropped to several patches to the model and merged to get the final dehazed results. This inference pipeline may introduce visible boundary artifacts. If the whole image is directly passed to the model, as shown in Figure 1 (d), though the visible boundary artifacts disappear, the contrast and the visual quality are dropped, which is derived from the different feature distributions of different image sizes. Therefore, we also use several post-processing to avoid this issue. As shown in Figure 1 (e), after using post-processing methods, we pass the whole images into the neural network, and the dehazed image contains better color tone and details. To sum up, our main contributions and novelties are as follows:

1. We propose two semantic guidance loss functions for non-homogeneous dehazing. The semantic corresponding loss makes dehazed images contain semantic information like texture or structure. The semantic color tone consistency loss further restricts color tone consistency and smoothness.

2. Combining the proposed loss functions, we modify the backbone [3], and several post-processing methods are implemented to improve the model's performance.

3. We test our proposed method and achieves competitive performance in the NTIRE 2023 HR NonHomogeneous Dehazing Challenge [16].

This paper is organized as follows. In Section 2, related works like single image dehazing and related optimization loss functions are briefly reviewed. Section 3 describes the proposed semantic-guided loss functions, the network architecture, overall optimized loss functions, and post-processing methods in detail. Section 4 introduces our experimental setting and provides several experimental results compared with conventional methods. Finally, the future works and conclusion are presented in Section 5.

2. Related Works

2.1. Single Image Dehazing

As introduced in the previous section, single-image dehazing can be divided from traditional prior-based methods to learning-based methods. Several traditional methods [5–7] estimate transmission maps and the global atmospheric light based on statistical assumption to restore the clear images. Because the prior-based methods are not robust to deal with situations like the non-homogeneous atmosphere, more and more researcher endeavor to purpose deep learning solutions to pursue better performance. On the other hand, due to the prevailing success of deep learning in various tasks and the capability of large datasets, many deep-learning-based [9, 10, 14, 17–19] methods are proposed. For example, FSGDN [10] completely utilizes dual guidance in both the frequency and spatial domains.

In [9], a detail-enhanced attention block is proposed to boost the feature learning for improving the performance of dehazing. For non-homogeneous dehazing, several methods [20] try to estimate density maps of haze and apply them to design the neural network. In [3], DW-GAN contains two branches of neural networks to learn the different features and merge them to deal with non-homogeneous dehazing.

2.2. Loss Functions for Single Image dehazing

Single image dehazing can be seen as a regression task, so absolute difference ($L1$) and mean square error ($L2$) loss functions can be used to optimize the network. Specifically, Chabornnier loss [21] that can avoid unstable value at zero point is proposed. This function makes the overall training phase stable and converges fast. Besides these pixel-wise loss functions, another common visual similarity metric based on image fidelity is given by Structured Similarity (SSIM) [22, 23], whose spirit is the calculation of the covariance within patches. Furthermore, the perceptual loss [13] that measures the feature similarity from the deep neural network is similar to human visual perception. Based on the perceptual loss, the contrastive loss [14] that pushes positives closer to anchors and pushes negatives away from anchors in the feature space has shown impressive performance in image dehazing. Recently, Generative adversarial network (GAN) [3] is proposed to train the dehazing neural network. GANs contain a generator and a discriminator, and the latter can calculate the adversarial loss to optimize the network. But all of them do not consider the benefit of semantic segmentation, which contains more useful information like color and texture of structure for non-homogeneous dehazing. Therefore, in this paper, we leverage the pre-trained semantic segmentation model and extract the semantic feature to optimize the network.

3. Proposed Methods

3.1. Semantic Guided Loss Functions

In this section, we describe the proposed two semantic-based loss functions: *semantic corresponding loss* L_{sem} and *semantic color tone consistency loss* L_{sc}.

Semantic corresponding loss enforces the dehazed images to have an identical semantic representation to that of the ground truth, and it is written as:

$$L_{sem}(I, \hat{I}) = |S(\hat{I}) - S(I)| \tag{2}$$

where $|\cdot|$ is the absolute value. $\hat{I} \in R^{W\times H\times 3}$ and $I \in R^{W\times H\times 3}$ are the dehazed image and the ground truth. S is the semantic segmentation model and $S(.) \in R^{W\times H\times n}$ is the semantic segmentation map with n classes. This loss strengthens the semantic relationship between ground truths and dehazed images. We also plot the semantic segmentation results in Figure 2, and it indicates that the segmentation results are not accurate in dense haze regions. In Figure 2 (a), most regions are predicted as roads in dark magenta, and the predictions from ground truth and the dehazed image are terrains in light green, which demonstrates the semantic consistency loss is helpful to remove haze, especially in dense hazy regions.

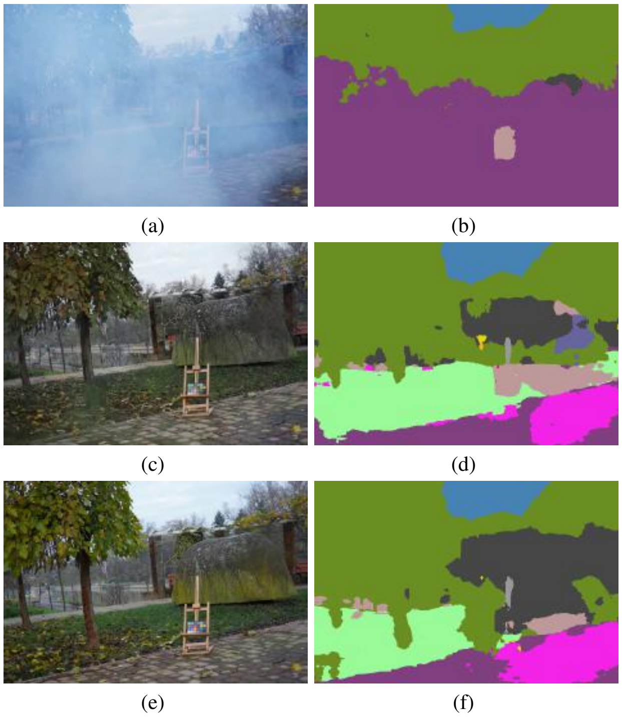

Figure 2. **Semantic segmentation results of different images.** (a): non-homogeneous haze image; (b): semantic segmentation results of (a); (c): dehazed image; (d): semantic segmentation results of (c); (e): ground truth; (f): semantic segmentation results of (e). The visualization of various colors is based on [24].

Semantic color tone consistency loss adjusts the color tones based on separated classes. We notice some pixels in some categories, like the sky, contain similar color tone distribution, which motivates us to consider semantic-based color tone to design the loss function. This loss function is written as:

$$L_{sc}(I, \hat{I}) = \sum_{i=1}^{n} |C_{I,i} - C_{\hat{I},i}| \tag{3}$$

where $C_{I,i}$ means the average color tone of the i^{th} class of I. Specifically, $C_{I,i}$ can be written as:

$$C_{I,i} = \frac{1}{n} \sum_{S(x)\in i} I(x) \tag{4}$$

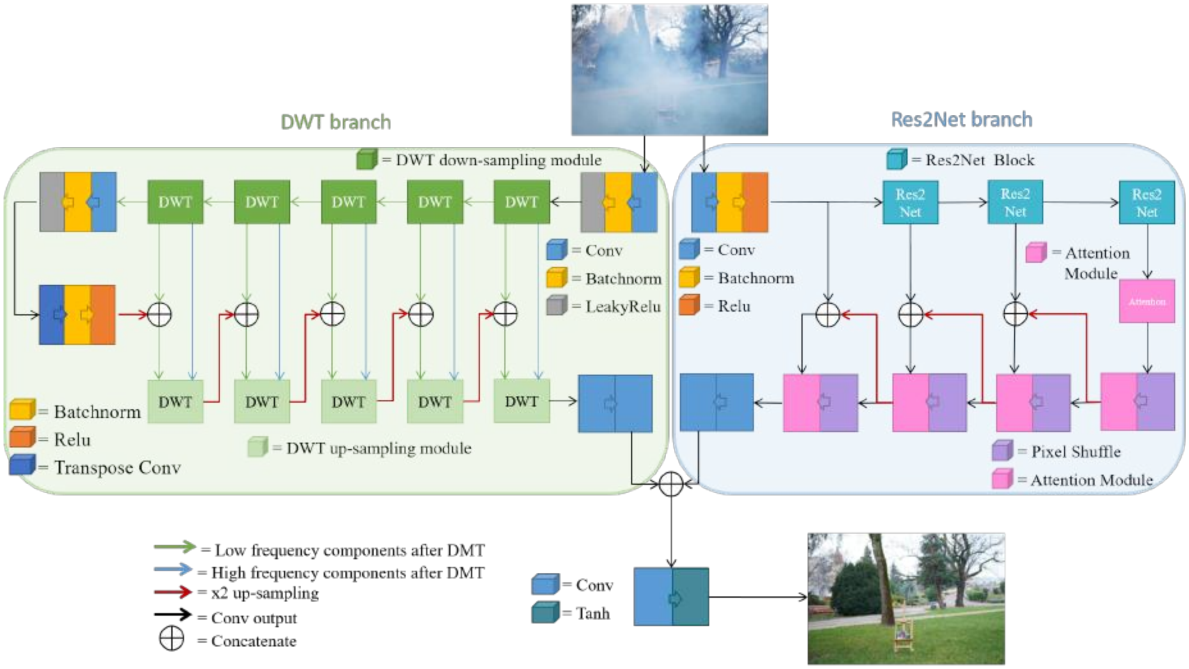

Figure 3. **The proposed network for high-resolution non-homogeneous dehazing.** This model applies the DWT U-net and Res2Net as two branches to extract the different frequency features and high-level features. It is noted all convolutional kernel size is 5 × 5. Two branches are encoder-decoder structures. A 7 × 7 convolution layer is added to fuse feature maps and output the clean images.

3.2. Overall Neural Network

The architecture for non-homogeneous dehazing is presented in Figure 3. This network is based on DW-GAN [3] and contains two branches: Discrete wavelet transform (DWT) branch [25] and Res2Net [26,27] branch. The DWT branch is used to directly extract various frequency features and learn the color tone mapping from hazy to haze-free images. This branch is seen as an encoder-decoder structure based on U-Net [28, 29]. The DWT down-sampling module and DWT up-sampling module are seen as the encoder and decoder modules, respectively. The input feature maps are passed to DWT down-sampling module and decomposed into low-frequency and three high-frequency components by DWT. Low-frequency components are concatenated with convolution output as down-sampling features, and high-frequency components are added to the DWT up-sampling module by skip connection. At the bottom of the U-net, the residual block is used and makes the training process effective, especially in the event of deeper networks. By adding DWT into the neural network, the network replaces the down-sampling and up-sampling with the DWT and inverse discrete wavelet transform (IDWT). Additionally, the network further captures the various frequency features and bi-orthogonal properties of the DWT for signal recovery.

Second, the Res2Net [26] branch also contains the encoder and decoder parts. In the encoder part, the Res2Net50 is applied [26] as the backbone. The Res2Net can represent multi-scale features at a granular level and increases the range of receptive fields for each network layer. After the input is passed through the backbone, multi-scale features are obtained. Note that the Res2Net utilized in this work discards the full connection layer, and the size of final output feature maps from our encoder is $\frac{1}{16}$. We connect the bottom features to the decoder. The decoder consists of stacks of convolution to refine the feature maps. The pixel shuffle [30] is adopted to magnify feature maps. Furthermore, we leverage attention modules to refine intermediate features. Attention modules contain both spatial and channel attention.

Finally, to fuse feature maps from two branches, a simple 7 × 7 convolution layer is added, and the model predicts the final clear images. In NTIRE 2023 HR NonHomogeneous Dehazing Challenge, the image size is up to 4000×6000, and the haze distribution becomes challenging. To address this issue, we increase the kernel size of the original model. We replace 3×3 convolutions with 5×5 convolutions in the whole model to increase the receptive field and make the dehazing network handle the challenging haze distribution.

3.3. Overall Loss Functions

Besides the proposed semantic loss functions L_{sem} and L_{sc}, we also train the network with three extra loss functions. The first function is Charbonnier loss [21], which is considered as the robust L_1 loss function near zero, and the formula of Charbonnier loss can be written as:

$$L_{Cha}(I,\hat{I}) = \frac{1}{T}\sum_{i}^{T}\sqrt{(I_i - \hat{I}_i)^2 + \epsilon^2} \tag{5}$$

where e is seen as a tiny constant (e.g., 10^{-6}) for stable and robust convergence. L_{cha} is used to restore global structure [21] and can handle outliers robustly.

Secondly, we apply the wavelet SSIM loss [23,30]. First, SSIM loss can be expressed as:

$$L_{SSIM}(I,\hat{I}) = -\frac{(2\mu_I\mu_{\hat{I}} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)} \tag{6}$$

where μ and σ mean the covariance, the mean, and the standard deviation of images. In the non-homogeneous image dehazing task, to remove different dense haze from the original image, we extend the SSIM loss function so that our network can restore more detailed parts. The DWT is helpful for both neural network structures and the loss function. We integrate DWT into the SSIM loss. Initially, the DWT decomposes the dehazed image into four different and small sub-band images. The operation can be expressed as

$$\hat{I}^{LL}, \hat{I}^{LH}, \hat{I}^{HL}, \hat{I}^{HH} = \mathrm{DWT}(\hat{I}) \tag{7}$$

where superscripts mean the output from respective filters (e.g., f_{LL}, f_{HL}, f_{LH} and f_{HH}). f_{HL}, f_{LH}, and f_{HH} are high-pass filters for the horizontal edge, the vertical edge, and the corner detection, respectively. f_{LL} is seen as the down-sampling operation. Moreover, the DWT can keep decomposing the $\hat{I}^{LL}$ generating images with different scales and frequency information. This step is written as:

$$\hat{I}_{i+1}^{LL}, \hat{I}_{i+1}^{LH}, \hat{I}_{i+1}^{HL}, \hat{I}_{i+1}^{HH} = \mathrm{DWT}(\hat{I}_i^{LL}) \tag{8}$$

where the subscript i means the output from the i^{th} DWT iteration, and $\hat{I}_0^{LL}$ is the original predicted dehazed image. The SSIM loss terms described above are calculated from the original image pair and various sub-band image pairs. The fusion of the SSIM loss and the DWT is integrated as follows:

$$\begin{aligned} &L_{W-SSIM}(I,\hat{I}) = \sum_{0}^{i}\gamma_i L_{\mathrm{SSIM}}(I_i^w, \hat{I}_i^w), \\ &w \in \{LL, HL, LH, HH\} \end{aligned} \tag{9}$$

where γ_i is based on [23] to control the importance of different patches. Compared to the original L_{SSIM}, L_{W-SSIM} can reconstruct local textures and details better.

The third loss is the perceptual loss [13]. Unlike the aforementioned two loss functions, the perceptual loss leverages multi-scale features based on a pre-trained deep neural network (e.g., VGG19 [31]) to measure the visual feature difference between the ground truth and the estimated image. Formally, in this task, the VGG19 pre-trained on ImageNet is utilized as the loss function network. The perceptual loss is defined as

$$L_{Per}(I,\hat{I}) = |(VGG(I) - VGG(\hat{I})| \tag{10}$$

where $|\cdot|$ is the absolute value. The overall loss function :

$$L_{Total} = \lambda_1 L_{cha} + \lambda_2 L_{W-SSIM} + \lambda_3 L_{Per} + \lambda_4 L_{sem} + \lambda_5 L_{sc} \tag{11}$$

where λ_1, λ_2, λ_3, λ_4, and λ_5 are tuning coefficients and used to adjust the relative weights of the overall loss function.

3.4. Post-Processing for Inference

Because of different image sizes during the training and testing status, several operations like global average and instance normalization would generate different results and impact the final quality of dehazing images as shown in 1 (d). To address these issues, we use Test-time Local Converter (TLC) [4] to adjust the model after the training phase. Specifically, our model contains the global average in the channel attention layer. TLC converts the region of feature aggregations from global to local, which aligns the global distribution with the local one.

Besides TLC bridges the gap of information aggregation between training and testing, another test time augmentation (TTA) method is applied to improve the performance of our model. Specifically, the input hazy images are rotated and flipped vertically and horizontally and passed through the neural network. We flip and rotate back dehazed images and average them as the final prediction. It is noted that there are 8 different orientation images used for TTA. In Section 4, several ablation studies demonstrate the effectiveness of these post-processing methods.

4. Experiments

4.1. Dataset

The NTIRE 2023 HR NonHomogeneous Dehazing Challenge dataset comprises 50 images from different scenes. 40 pairs of clear and non-homogeneous dehazing images are used for training, and five non-homogeneous dehazing images are used for validation and testing, respectively. The resolution of all images is 4000 × 6000. To avoid overfitting the model, we split five pairs of images from the training data and evaluated them to select the best model. Additionally, we also use extra datasets from the previous NTIRE dehazing Challenge: O-Haze [11], DENSE-HAZE [12], NH-HAZE [32] and NH-HAZE2 [33]. There are 45 outdoor images and the corresponding images affected by haze in O-Haze. DENSE-HAZE contains

33 dense hazy and ground-truth images. It is noted both datasets are not for non-homogeneous dehazing, but we also use them to make the model robust against challenging non-homogeneous hazy scenarios. Both NH-Haze and NH-Haze2 consist of all 80 non-homogeneous hazy images and their corresponding ground truth images of the same scene.

4.2. Experimental Setting

During the training phase, the image size is randomly cropped as 512 × 512, and we use the random rotation and the random flip vertically and horizontally as the data augmentation. The AdamW optimizer [34] is utilized with batch size 10 to train the network. We train the network for 1000 epochs with the momentum $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rate is initialed as 10^{-4} and divided by ten after 300, 600, and 900 epochs. We use (11) to optimize the network, and the λ_1, λ_2, λ_3, λ_4, and λ_5 in (11) are set as 1, 1.1, 0.1, 0.1 and 0.1, respectively. We use the pre-trained DeepLab v3 [15] semantic segmentation model trained on the Cityscape dataset [24] to extract the semantic masks. The Cityscape dataset contains 19 categories, and we just select the fence, sky, terrain, vegetation, sidewalk, and road in (2) and (4). In the final 100 epochs, we remove semantic-based loss functions. We perform our experiments on a single Nvidia V100 graphic card and the PyTorch platform. We spend about 40 hours training the model. In the testing phase, we can feed the whole image to the model, and our model takes 6.51 seconds to infer a single image.

4.3. Ablation Experiments

To find the best effectiveness of the proposed solution, we perform some ablation experiments in this section. The peak signal-to-noise ratio (PSNR) and the SSIM are applied as objective metrics for quantitative evaluation.

The ablation experiments contain five experimental settings. First, we use the dehazing neural network of DW-GAN [3] as the baseline. Second, we change the kernel size in DW-GAN from 3 × 3 to 5 × 5. Note that only Charbonnier loss [21], wavelet SSIM loss [23], and perceptual loss [13] functions are used for both experiments. Third, we use the proposed loss functions L_{sem} L_{sc} to train the large model. Forth, we use TLSC [4] as post-processing to make inferences. Last, besides TLSC, TTA is also applied.

The results tested on the validation images are reported in Table 1. The PSNR and SSIM scores of setting 2 can be improved compared with setting 1. It can demonstrate that using large convolutional kernels can increase the receptive field to capture features of non-homogenous haze and improve dehazing performance. Furthermore, compared with setting 2, the performance of setting 3 is improved effectively. It indicates that the L_{sem} and L_{sc} can be more beneficial to reconstruct clear images. Additionally, compared with setting 3, performances of setting 4 and setting 5 are also further improved. It proves that post-processing is essential for performance even though the network is unchanged. To sum up, compared with baseline methods, in setting 3, the proposed semantic-based loss functions new model makes the PSNR score and the SSIM improve by 1.70 and 0.017, respectively. On the other hand, the PSNR and the SSIM are increased by 0.38 and 0.008 between setting 3 and setting 5, which demonstrates the effeteness of our proposed strategies.

Table 1. **The ablation experiment of applying different models, loss functions, and post-processing.**

Index	Method	Metrics	
		PSNR	SSIM
(1)	Baseline [3]	19.3441	0.6846
(2)	(1) + Large kernels	20.5395	0.6964
(3)	(2) + L_{sem}, L_{sc}	21.0412	0.7011
(4)	(3) + TLSC [4]	21.3831	0.7049
(5)	(4) + TTA	21.4235	0.7088

Table 2. **Non-homogeneous dehazing results by state-of-the-art methods.**

Method	**Metrics**	
	PSNR	SSIM
DW-GAN [3]	21.3212	0.7349
FSDGN [10]	21.3826	0.7355
DEA-Net [9]	19.6420	0.7295
Ours	22.5186	0.7477

4.4. Comparison with State-of-the-art Methods

We compare our solution with three state-of-the-art single-image dehazing methods, including DW-GAN [3] DEA-Net [9], and FSDGN [10] as described in Section 2. We use the same training set to train these methods and report PSNR and SSIM on our split validation data. As shown in Table 2, our solution outperforms other methods by a large margin. Our method achieves the best performance on both PSNR and SSIM, which surpasses the second place 1.13 dB and 0.012 in SSIM. It is noted we apply the different datasets to calculate the PSNR and the SSIM, so the numerical results of Table 1 and Table 2 are not aligned.

Some dehazed images are plotted in Figure 4. Compared with other state-of-the-art methods, the proposed method has the best performance in terms of non-homogenous haze removal and artifact/distortion suppression. It is noted that the performance of FSDGN is slightly better than that of DW-GAN, but the dehazed results from FSDGN contain some artifacts and residual haze. Therefore, we select DW-GAN [3] as our baseline.

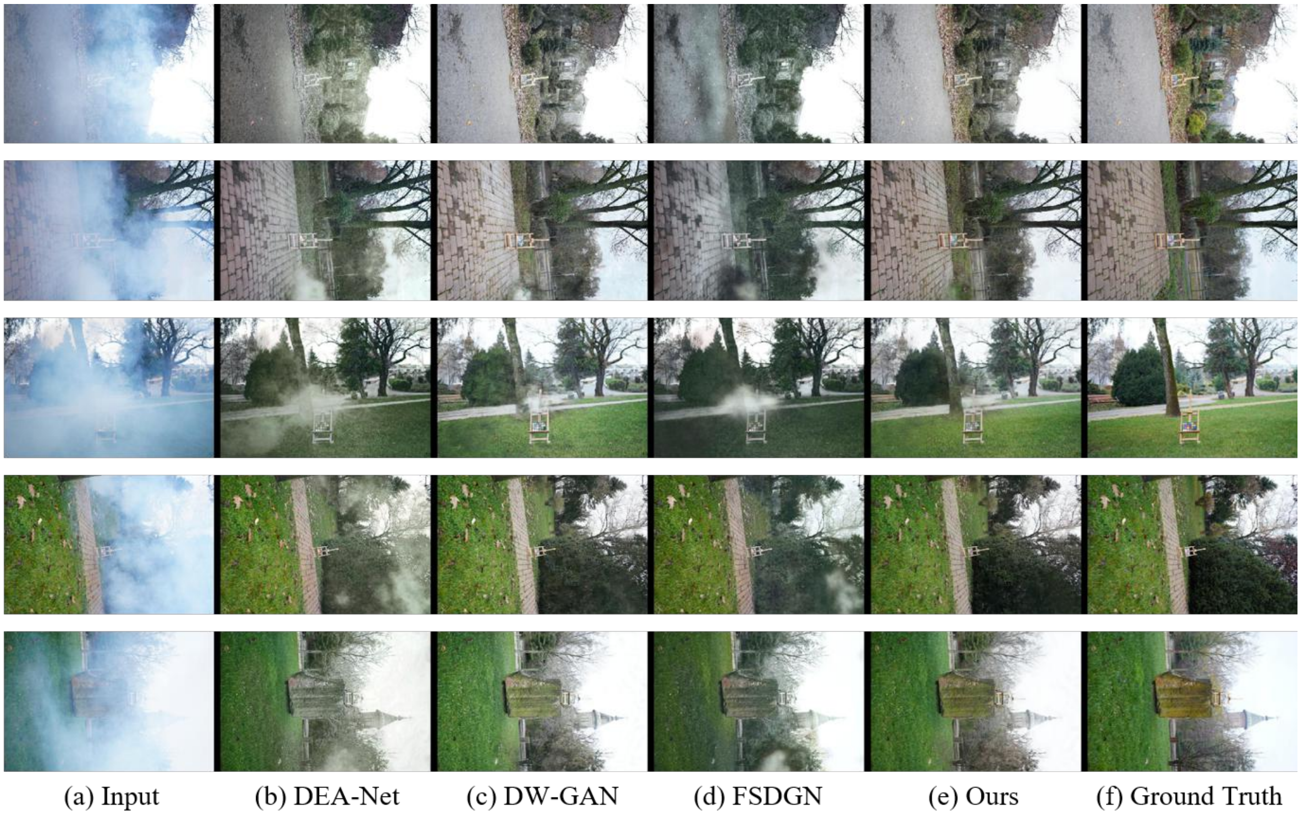

Figure 4. **Visual comparison for the high resolution non-homogeneous dehazing results recovered by our solution and other state-of-the-art solutions.** For convenience, all images are depicted in landscape modes.

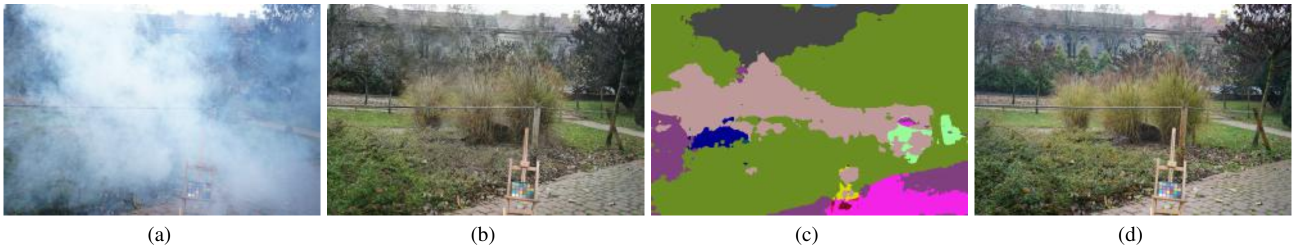

Figure 5. **The failure segmentation results and the corresponding dehazed results.** (a): the non-homogeneous haze image; (b): the dehazed results from our solution; (c): the inaccurate semantic segmentation result. The skin pixels and gray pixels mean the building and fence. (d): ground truth. Compared with the ground truth, there is a color tone shift and residual haze on the building of (b).

4.5. Results of Challenge

We list the results of the proposed solution compared with other competing entries in HR Non-Homogeneous Dehazing of NTIRE 2023 workshop [16] in Table 3. Besides PSNR, SSIM, and LPIPS, the Mean Opinion Score (MOS), as a result of an user study set by the challenge organizers is adopted to evaluate the performance of all submissions. As shown in Table 3, our results obtained a competitive performance in terms of SSIM, PSNR, LPIPS, and MOS.

4.6. Limitations and Discussion

The proposed method contains semantic guidance loss functions and a large kernel neural network, making it easy to learn the mapping functions of non-homogeneous dehazing. Although our method achieves competitive performance in this competition, there are some limitations. First, the Cityscape dataset and dataset from NTIRE 2023 HR NonHomogeneous Dehazing Challenge are collected in different environments, which may cause the domain gap and the inaccurate semantic segmentation prediction as shown

Table 3. The average SSIM, PSN, LPIPS, MOS of top 10 methods over NTIRE 2023 HR NonHomogeneous Dehazing Challenge dataset validation and testing dataset.

User name	PSNR	SSIM	LPIPS	MOS
zhouh115	22.87	0.71	0.346	8.07
lillian	22.96	0.71	0.345	7.85
ShawnDong98	22.18	0.7	0.401	7.125
Yinwei_Wu	21.97	0.68	0.38	7.9
Anas	22.27	0.7	0.439	7.4
YuanGao	21.75	0.7	0.404	6.95
lightdehaze	22.01	0.70	0.384	5.35
xsourse	22.09	0.65	0.556	7.65
CongXiaofeng	21.86	0.67	0.492	6.9
HaoqiangYang	22.11	0.71	0.442	7.1

in Figure 5(c). Furthermore, L_{sc} can not handle certain categories like buildings because buildings would contain various color pixels. These limitations cause color tone shift and residual haze on the building of dehazed images as shown in Figure 5(b). Last but not least, as shown in Table 3, in terms of perceptual properties, the proposed method seems to suffer. It demonstrates that when designing the loss function, it is necessary to consider the impact of perceptual loss and the balance of overall loss functions.

5. Conclusion

In this paper, we propose a novel solution for HR nonhomogeneous dehazing. Our solution mainly contains three parts. First, we propose two semantic-based loss functions: semantic corresponding loss and semantic color tone consistency loss to optimize the dehazing network. Furthermore, we increase the receptive fields of the model to handle complicated high-resolution hazy scenarios. Third, several post-processing methods are applied to further improve the performance of the model. In the NTIRE 2023 HR Non-Homogeneous Dehazing Challenge, our solution achieves competitive performance. In future works, we would collect compact datasets for a better semantic segmentation model. We could integrate the semantic guided loss function to contrastive learning [14]. Furthermore, features from language knowledge or text-image embedding seen as the high-level guidance [35] can be used to further develop the dehazing model. Besides dehazing, the proposed semantic-based loss functions can be leveraged for other image enhancement tasks [36, 37].

6. Acknowledgement

We thank to National Center for High-performance Computing (NCHC) for providing computational and storage resources. This work was supported by the National Science and Technology Council, Taiwan, under Grant MOST 108-2221-E-002-072-MY3, Grant MOST 108-2638-E-002-002-MY2, and Grant NSTC 111-2221-E-002-136-MY3.

References

[1] W.-T. Chen, I.-H. Chen, C.-Y. Yeh, H.-H. Yang, H.-E. Chang, J.-J. Ding, and S.-Y. Kuo, "Rvsl: Robust vehicle similarity learning in real hazy scenes based on semi-supervised learning," in *ECCV*, 2022. 1

[2] E. J. McCartney, "Optics of the atmosphere: scattering by molecules and particles," *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976. 1

[3] M. Fu, H. Liu, Y. Yu, J. Chen, and K. Wang, "Dw-gan: A discrete wavelet transform gan for nonhomogeneous dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2021, pp. 203–212. 2, 3, 4, 6

[4] X. Chu, L. Chen, C. Chen, and X. Lu, "Improving image restoration by revisiting global information aggregation," in *ECCV*, 2022. 1, 2, 5, 6

[5] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1674–1682. 1, 2

[6] C. O. Ancuti and C. Ancuti, "Single image dehazing by multi-scale fusion," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3271–3282, 2013. 1, 2

[7] X. Liu, H. Zhang, Y.-m. Cheung, X. You, and Y. Y. Tang, "Efficient single image dehazing and denoising: An efficient multi-scale correlated wavelet approach," *Computer Vision and Image Understanding*, vol. 162, pp. 23–33, 2017. 1, 2

[8] Y. Song, Y. Zhou, H. Qian, and X. Du, "Rethinking performance gains in image dehazing networks," *arXiv preprint arXiv:2209.11448*, 2022. 1

[9] Z. Chen, Z. He, and Z.-M. Lu, "Dea-net: Single image dehazing based on detail-enhanced convolution and content-guided attention," *arXiv preprint arXiv:2301.04805*, 2023. 1, 2, 3, 6

[10] H. Yu, N. Zheng, M. Zhou, J. Huang, Z. Xiao, and F. Zhao, "Frequency and spatial dual guidance for image dehazing," in *ECCV*, 2022. 1, 2, 6

[11] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 754–762. 1, 5

[12] 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 *2019 IEEE international conference on image processing (ICIP)*. IEEE, 2019, pp. 1014–1018. 1, 5

[13] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711. 2, 3, 5, 6

[14] H. Wu, Y. Qu, S. Lin, J. Zhou, R. Qiao, Z. Zhang, Y. Xie, and L. Ma, "Contrastive learning for compact single image dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 2, 3, 8

[15] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018. 2, 6

[16] C. O. Ancuti, C. Ancuti, F.-A. Vasluianu, and R. Timofte, "Ntire 2023 challenge on nonhomogeneous dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2, 7

[17] H. Wu, J. Liu, Y. Xie, Y. Qu, and L. Ma, "Knowledge transfer dehazing network for nonhomogeneous dehazing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020. 2

[18] W.-T. Chen, J.-J. Ding, and S.-Y. Kuo, "PMS-net: Robust haze removal based on patch map for single images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 2

[19] W.-T. Chen, Z.-K. Huang, C.-C. Tsai, H.-H. Yang, J.-J. Ding, and S.-Y. Kuo, "Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 653–17 662. 2

[20] Y. Jin, W. Yan, W. Yang, and R. T. Tan, "Structure representation network and uncertainty feedback learning for dense non-uniform fog removal," in *Computer Vision–ACCV 2022: 16th Asian Conference on Computer Vision*, 2023. 3

[21] J. T. Barron, "A general and adaptive robust loss function," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 3, 4, 5, 6

[22] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, 2016. 3

[23] H.-H. Yang, C.-H. H. Yang, and Y.-C. J. Tsai, "Y-net: Multiscale feature aggregation network with wavelet structure similarity loss function for single image dehazing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020. 3, 5, 6

[24] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213–3223. 3, 6

[25] H.-H. Yang, C.-H. H. Yang, and Y.-C. F. Wang, "Wavelet channel attention module with a fusion network for single image deraining," in *IEEE International Conference on Image Processing (ICIP)*, 2020. 4

[26] S. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. H. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE transactions on pattern analysis and machine intelligence*, 2019. 4

[27] H.-H. Yang, W.-T. Chen, and S.-Y. Kuo, "S3net: A single stream structure for depth guided image relighting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 276–283. 4

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015. 4

[29] H.-H. Yang and Y. Fu, "Wavelet U-net and the chromatic adaptation transform for single image dehazing," in *IEEE International Conference on Image Processing (ICIP)*, 2019. 4

[30] 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 subpixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. 4, 5

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. 5

[32] C. O. Ancuti, C. Ancuti, and R. Timofte, "Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 444–445. 5

[33] C. O. Ancuti, C. Ancuti, F.-A. Vasluianu, and R. Timofte, "Ntire 2021 nonhomogeneous dehazing challenge report," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 627–646. 5

[34] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017. 6

[35] G. Kwon and J. C. Ye, "Clipstyler: Image style transfer with a single text condition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18 062–18 071. 8

[36] H.-H. Yang, K.-C. Huang, and W.-T. Chen, "Laffnet: A lightweight adaptive feature fusion network for underwater image enhancement," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021. 8

[37] H.-E. Chang, C.-H. Hsieh, H.-H. Yang, I.-H. Chen, Y.-C. Chen, Y.-C. Chiang, W.-T. Huang, Zhi-Kai Chen, and S.-Y. Kuo, "TSRFormer: Transformer based two-stage refinement for single image shadow removal," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 8