

Efficient Re-Parameterization Residual Attention Network for Nonhomogeneous Image Dehazing

Erkang Chen ^{1,2,†} , Tian Ye ^{1,†}, Jingxia Jiang ^{1,†}, Lihan Tong ¹ and Qiubo Ye ^{1,2,*} 

¹ School of Ocean Information Engineering, Jimei University, Xiamen 361021, China; ekchen@jmu.edu.cn (E.C.); 201921114031@jmu.edu.cn (T.Y.); 202021114006@jmu.edu.cn (J.J.); 202121301035@jmu.edu.cn (L.T.)

² Fujian Provincial Key Laboratory of Oceanic Information Perception and Intelligent Processing, Xiamen 361021, China

* Correspondence: qbye@jmu.edu.cn

† These authors contributed equally to this work.

Abstract: Real-world nonhomogeneous haze brings challenges to image restoration. More efforts are needed to remove dense haze and thin haze simultaneously and efficiently. However, most existing dehazing methods do not pay attention to the complex distributions of haze and usually suffer from a low runtime speed. To tackle such problems, we present an efficient re-parameterization residual attention network (RRA-Net), whose design has three key aspects. Firstly, we propose a training-time multi-branch residual attention block (MRAB), where multi-scale convolutions in different branches cope with the nonuniformity of haze and are converted into a single-path convolution during inference. It also features local residual learning with improved spatial attention and channel attention, allowing dense and thin haze to be attended to differently. Secondly, our lightweight network structure cascades six MRABs followed by a long skip connection with attention and a fusion tail. Overall, our RRA-Net only has about 0.3M parameters. Thirdly, two new loss functions, namely the Laplace pyramid loss and the color attenuation loss, help train the network to recover details and colors. The experimental results show that the proposed RRA-Net performs favorably against state-of-the-art dehazing methods on real-world image datasets, including both nonhomogeneous haze and dense homogeneous haze. A runtime comparison under the same hardware setup also demonstrates the superior efficiency of the proposed network.

Keywords: image dehazing; nonhomogeneous haze; re-parameterization network; residual attention block



Citation: Chen, E.; Ye, T.; Jiang, J.; Tong, L.; Ye, Q. Efficient Re-Parameterization Residual Attention Network for Nonhomogeneous Image Dehazing. *Appl. Sci.* **2023**, *13*, 3739. <https://doi.org/10.3390/app13063739>

Academic Editors: Zheng Wang, Xian Zhong, Liang Liao and Kui Jiang

Received: 1 February 2023

Revised: 26 February 2023

Accepted: 8 March 2023

Published: 15 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Clear visibility is crucial to the success of outdoor computer vision tasks [1]. However, the quality of outdoor images often deteriorates due to haze, i.e., the existence of smoke, dust, fumes, mist, and other floating particles in the atmosphere. Such haze reduces the performance of a subsequent visual analysis. Moreover, haze distribution often has a non-homogeneous character in many real scenes, posing additional challenges [2]. Single-image dehazing as a fundamental low-level vision task aims to recover the latent haze-free image and has attracted increasing attention in the computer vision community over the past few decades [3–8].

In the early dehazing methods [9–11], image priors are usually used to estimate important parameters of the imaging model for hazy scenes. Then, image dehazing is fulfilled by solving the inverse problem of such an imaging model. However, these image priors are not always valid in complex, real hazy scenes. Additionally, the nonhomogeneous distribution of haze is certainly one factor that adds to the difficulties.

Recently, deep-learning-based methods [12–20] have directly learned the latent clean image from a single hazy image in an end-to-end manner and have shown promising performance in haze removal. However, the current deep-learning-based dehazing models

often have a large number of parameters (e.g., the numbers of parameters of the recent two-branch neural network (TBNN) [19] and the Dehazer [20] network are around 50.40M and 132.45M, respectively) or suffer from a low runtime speed. In addition, to overcome the complex distributions of non-homogeneous images, a lightweight network to deal with nonhomogeneous haze effectively and efficiently is of great value. To this end, we design an efficient re-parameterization residual attention network featuring a training-time multi-branch residual attention block, an end-to-end lightweight network structure and two new loss functions to achieve a good balance between the dehazing performance and model complexity. We conduct comparative experiments on both non-homogeneous and homogeneous haze scenes to demonstrate the superiority of the proposed RRA-Net. An ablation study is also conducted to validate the main modules of the proposed network.

Our main contributions are summarized as follows:

(1) A training-time multi-branch residual attention block is designed for coping with the non-uniformity of haze.

(2) An end-to-end lightweight network structure (with only about 0.3M parameters) is proposed, which cascades six MRABs followed by a long skip connection with attention and a fusion tail.

(3) Two novel loss functions, namely the Laplace pyramid loss and the color attenuation loss, are employed to train our RRA-Net.

2. Related Works

There are mainly two types of methods for single-image dehazing: prior-based methods and deep-learning-based methods.

2.1. Prior-Based Methods

This type of dehazing method utilizes prior-information based on observations about characteristics of haze-free images or haze-degraded images in order to solve the inverse problem of the atmospheric scattering model. He et al. [9] proposed the dark channel prior (DCP) for single-image haze removal. The DCP is based on the observation that the pixels of a haze-free image tend to have very low intensity in at least one color (red, green, or blue) channel. Tan [10] developed a cost function in the framework of Markov random fields, which was based on two basic observations: haze-free images have more contrast, and the airlight tends to be smooth. Zhu et al. [11] presented the color attenuation prior which states that the relationship among the scene depth, the brightness, and the saturation is linear. They estimated the depth map of the scene using this prior and a supervised linear model. Berman et al. [21] observed that the pixels of a haze-free image in a given cluster are often non-local, which was dubbed as the non-local prior. Bui et al. [22] constructed color ellipsoids that were statistically fitted to haze pixel clusters in the RGB space and then calculated the transmission values through color ellipsoid geometry. Yuan et al. [23] proposed a confidence prior to accurately estimate scene transmissions for image dehazing. However, these hand-crafted priors do not always hold true for diverse real-world scenes with different haze statistics. For example, the DCP is unreliable for sky regions or white objects.

2.2. Deep-Learning-Based Methods

With the availability of large scale datasets and the development of deep neural networks, data-driven deep-learning-based methods have achieved promising results in single-image dehazing.

At first, deep learning was used to estimate variables of the atmospheric scattering model in Equation (1) or certain derived variables [24–27]. For example, Cai et al. [24] proposed DehazeNet, which takes in a hazy image and outputs its transmission map. Apart from the transmission map, the global atmospheric light also needs to be estimated separately in order to recover a haze-free image via the atmospheric scattering model. Li et al. [25] reformulated the atmospheric scattering model and designed the all-in-one

dehazing network (AOD-Net) to estimate a K variable that effectively combines the global atmospheric light and transmission map into one.

Later on, end-to-end convolutional neural networks (CNNs) were devised to directly estimate the haze-free image from a haze-degraded image [12–19]. Liu et al. [12] proposed GridDehazeNet, which implements a novel attention-based multi-scale estimation on a grid network for single-image dehazing. Qin et al. [13] presented the feature fusion attention network (FFA-Net) in which features at different levels are fused by an attention-based feature fusion structure combining channel attention with a pixel attention mechanism. While FFA-Net considers different channel-wise features and an uneven haze distribution, it carries out convolutional operations at the resolution of the original image, resulting in a large amount of calculation. Dong et al. [14] designed the multi-scale boosted dehazing network (MSBDN) with dense feature fusion based on the U-Net architecture, which incorporates the strengthen–operate–subtract boosting strategy in the decoder of the model. Das et al. [15] proposed a fast deep multi-patch hierarchical network (DMPHN) to restore non-homogeneous hazed images by aggregating features from multiple image patches from different spatial sections of the hazed image. Yu et al. [19] introduced a two-branch neural network to separately deal with non-uniformly distributed haze and the limited data challenge and fuse features from these two branches. Bu et al. [16] designed a generative adversarial network with residual guided filters that effectively obtains the contour information of a hazy image. Apart from designing more powerful CNNs, advanced learning techniques are also exploited for single-image dehazing. For example, contrastive learning is used to exploit both the information of hazy images and clear images as negative and positive samples [18]. While achieving promising dehazing results, these networks often have a large number of parameters or perform poorly on nonhomogeneous haze situations in the real world. Unlike the representative end-to-end CNN-based dehazing techniques, the proposed method takes into consideration both a lightweight network design and the ability to cope with the nonuniformity of haze. Specifically, we use multi-scale convolutions and an improved attention mechanism to remove dense haze and thin haze simultaneously.

Recently, visual transformers have also been utilized for image dehazing due to their capabilities for global modeling [20,28–31]. To take the advantage of both visual transformers and CNNs, Xu et al. [28] proposed a transformer–convolution fusion dehazing network. Guo et al. [20] brought a haze density-related prior into the transformer via a novel transmission-aware 3D position-embedding module and modulated the CNN features via learning modulation matrices conditioned on transformer features, instead of simple feature addition or concatenation. However, this transformer relies on large-scale training data for optimal performance. Gao et al. [29] proposed a transformer-based channel attention module combined with a spatial attention module to enhance a CNN-based backbone network. Despite the impressive performance improvements brought by combing transformers and CNNs, these hybrid networks usually have a much higher complexity with more parameters and a slower runtime speed, hindering their applications in real-time scenarios, especially when dehazing high-definition images is necessary.

3. Materials and Methods

Despite the development of single-image dehazing techniques, effectively and efficiently recovering the haze-free images from under nonhomogeneous haze is still a challenge from a technological perspective. Thus, we propose the following methodology.

3.1. Case Study and Definitions

Nonhomogeneous haze is common in real-world hazy weather conditions. One example of hazy images taken in such scenes is shown in Figure 1a, where the haze density is nonuniform across the image. It can be seen that the image suffers from low contrast and detail loss. Additionally, the degree of image degradation is closely related to the haze density.

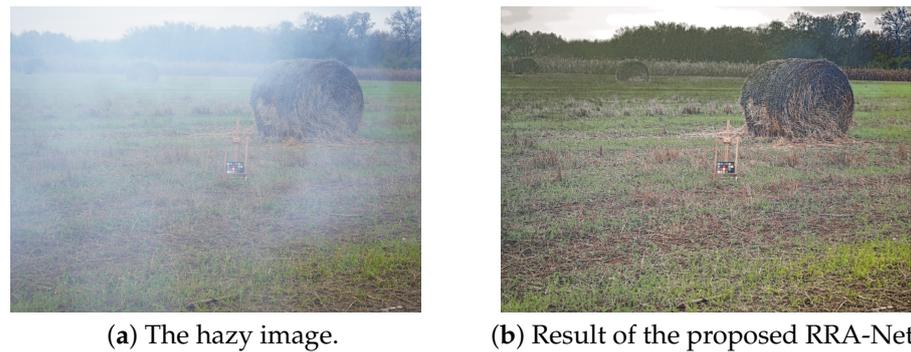


Figure 1. Visual result of the proposed method on NH-Haze (2021) dataset. Our method is able to produce haze-free images with high perceptual quality.

It is well-established that such image degradation caused by haze can be formulated by the atmospheric scattering model [9]:

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

where $I(x)$ is the observed hazy image, $J(x)$ is the latent haze-free image, $t(x)$ is the transmission map, and A is the global atmospheric light. In previous studies, the transmission map t is sometimes assumed to be attenuated exponentially with the scene depth d , i.e., $t(x) = e^{-\beta d(x)}$, where β is the scattering coefficient of the atmosphere. However, this relation is invalid under nonhomogeneous hazy scenes since the assumed transmission map is only distance-dependent, and hence inconsistent, with the randomness of nonhomogeneous haze.

Given an observed hazy $I(x)$, the task of single-image dehazing aims to recover the latent haze-free $J(x)$. To this end, we take the deep learning approach and propose an end-to-end RRA-Net model to tackle the challenge of nonhomogeneous degradations. In the meantime, high dehazing efficiency is taken into consideration to make sure of its practicability in real-time applications.

3.2. Proposed Model

Figure 2 shows the architecture of our RRA-Net, which has a lightweight, shallow structure. Specifically, we design a novel training-time MRAB as the basic block. Our RRA-Net cascades 6 MRABs, followed by a long skip connection with attention and a fusion tail. In this subsection, the MRAB is introduced first, followed by the details of the RRA-Net's lightweight structure and the loss function used during training.

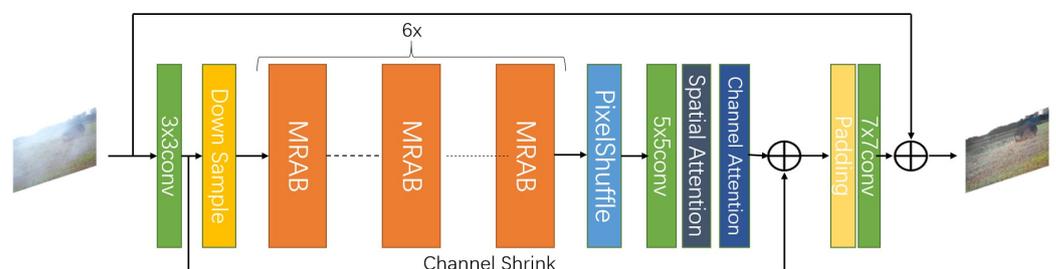


Figure 2. The efficient RRA-Net architecture. It cascades 6 MRABs, followed by a long skip connection with attention and a fusion tail.

3.2.1. Training-Time Multi-Branch Residual Attention Block

Inspired by RepVGG [32], we design a training-time multi-branch structure with an improved batch normalization (BN) strategy. It consists of one 3×3 convolution, one 1×1 convolution, and one identity connection, as shown in Figure 3. This parallel multi-branch structure allows the convolutions of different receptive field sizes to extract

features at different scales and to cope with the nonuniformity of haze. During inference, the multi-branch convolutions and the following BN are converted into a single-path 3×3 convolution through the technique of re-parameterization, largely reducing the inference time.

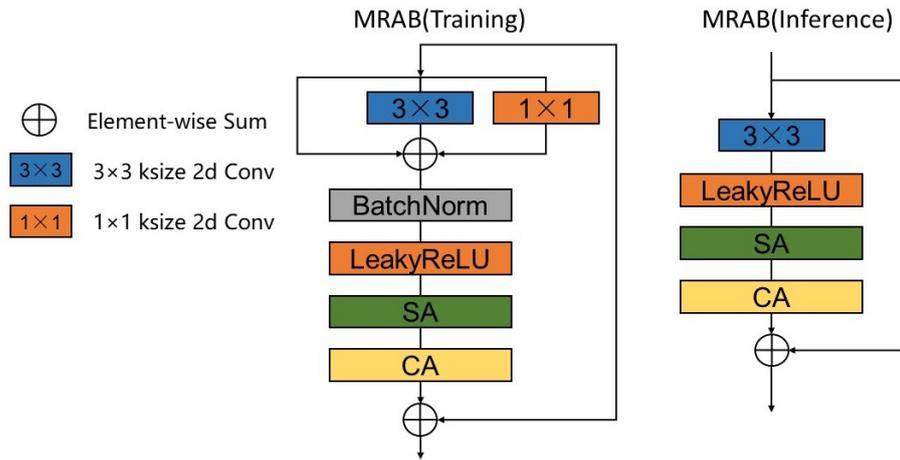


Figure 3. Structures of MRAB during training (left) and during inference (right).

Furthermore, we do not use BN in each branch, as was used in RepVGG, but rather, we use a single BN layer to stabilize the training after summing up the outputs of 3 convolutions. Since BN actually destroys the internal features of the sample, we believe that if individual BN layers are present in different branches, the connection between the features in different branches will be degraded more seriously, incurring a performance loss.

The MRAB also utilizes local residual learning with attention. The local residual connection allows features from haze-free areas to pass directly without processing so that the multi-branch convolutions can deal with the hazy areas in the image. Additionally, our attention module, consisting of an improved spatial attention (SA) layer and a channel attention (CA) layer, allows dense and thin hazy areas in the image be attended to differently.

In Section 3.3, detailed ablation studies are conducted to analyze the MRAB and verify the effectiveness of its structure and the BN strategy. Below, we provide the details of this basic block.

Inference-Time Re-Parameterization of MRAB

During inference, the multi-branch convolutions and the following BN are converted into a single-path 3×3 convolution through the technique of re-parameterization. Let X denote the input of the MRAB, and Y denote the output of BN. The process of re-parameterization can be described as:

$$Y = BN(W^3 * X + W^1 * X + X), \tag{2}$$

where W^3 denotes the kernel of the 3×3 convolution branch, and W^1 denotes the kernel of the 1×1 convolution branch. Since the identity branch can be expressed as a particular 1×1 convolution, and the 1×1 convolution can be considered a 3×3 convolution with zero padding, the re-parameterization during inference is:

$$\begin{aligned} Y &= BN(W^3 * X + W^1 * X + X) \\ &= BN(W^3 * X + W_1^3 * X + W_{identity}^3 * X) \\ &= BN((W^3 + W_1^3 + W_{identity}^3) * X), \end{aligned} \tag{3}$$

where W_1^3 and $W_{identity}^3$ denote the equivalent kernels of the 1×1 convolution branch and the identity branch after re-parameterization.

Furthermore, BN in the above equation can also be integrated into convolution during inference. Let $\{W, b\}$ represent the 3×3 convolution kernel and bias obtained by the three-way branch fusion in Equation (3), and $\mu, \sigma, \gamma, \beta$, respectively, represent the accumulated mean, standard deviation, learned scaling factor, and bias of the BN layer. Then, the convolution kernel and bias obtained after integrating BN are:

$$\begin{aligned} W' &= \frac{\gamma}{\sigma} W, \\ b' &= \frac{\gamma(b - \mu)}{\sigma} + \beta. \end{aligned} \quad (4)$$

Spatial Attention

Since the haze distribution often has a non-homogeneous character in many real scenes, a spatial attention mechanism is advantageous for dealing with such a situation. We design an improved spatial attention module, where 3D MaxPool, 1×1 convolution, ReLU, and 1×1 convolution are applied subsequently, and then a Sigmoid operation is used to output the attention weights:

$$Y_{sa} = \sigma(\text{Conv}(\text{Relu}(\text{Conv}(\text{MaxPool}(X))))), \quad (5)$$

where X is the input feature, σ is the Sigmoid function, and Y_{sa} is the spatial attention weights. The output Y is derived as the element-wise product of the input X and the spatial attention weights Y_{sa} :

$$Y = Y_{sa} \circ X. \quad (6)$$

Channel Attention

Different channels of a feature map often have different degrees of importance; thus, the channel attention mechanism is necessary for low-level vision tasks, such as dehazing. We follow the channel attention design in [13].

3.2.2. Details of RRA-Net's Lightweight Structure

In RRA-Net, as shown in Figure 2, a 3×3 convolution is firstly applied to the hazy image, outputting a feature map of the shape $64 \times H \times W$ where H and W are the height and width of the input image. Then, a 3×3 , stride 2 convolution is used to downsample the feature map, which doubles the receptive fields of the subsequent convolutions. After that, 6 MRABs are applied subsequently to gradually extract features, followed by a pixel shuffle layer, which upscales the feature map back to the original spatial size $H \times W$ and reduces the number of channels to 16. To avoid the loss of shallow features, before the final fusion tail, a long skip connection with attention is used to introduce shallow features. Specifically, the feature map after the pixel shuffle passes through a 5×5 convolution, a spatial attention layer, and a channel attention layer. In the meantime, the shallow features from the first 3×3 convolution are channel shrunk and long skip connected. The final fusion tail consists of a reflection padding, a 7×7 convolution, and a residual connection. The reflection padding operation alleviates the image boundary distortion problem caused by the following convolution of the large kernel size. The 7×7 convolution outputs the difference Y_{res} between the input hazy image I_{hazy} and the recovered clean image $RRA(I_{hazy})$. Thus, the output of the proposed RRA-Net is

$$RRA(I_{hazy}) = Y_{res} + I_{hazy}. \quad (7)$$

Such a multi-stage fusion strategy has the advantages of combining shallow and deep features and avoiding the instability caused by the upsampling process. The struc-

ture of RRA-Net achieves a good balance between the dehazing performance and model complexity. Overall, it only has about 0.3M parameters.

3.2.3. Loss Function

To recover high-frequency details and colors, we design two new loss functions, namely the Laplace pyramid loss and the color attenuation loss. During training, the total loss function is

$$L_{\Theta} = L_1 + \alpha_1 L_{\text{Laplace}} + \alpha_2 L_{\text{CA}}, \quad (8)$$

where Θ denotes the parameters of RRA-Net, L_1 is the L1-norm loss, L_{Laplace} is the Laplace pyramid loss, L_{CA} is the color attenuation loss, and α_1, α_2 are the balancing coefficients. It is worth noting that we do not adopt the commonly used perceptual loss or GAN loss.

Laplace Pyramid Loss

The Laplace pyramid is commonly used to extract high-frequency features from an image. To calculate the Laplace pyramid loss, the recovered clean image $RRA(I_{\text{hazy}})$ and its ground-truth I_{gt} undergo the same process to build two 3-layer Laplace pyramids. Let G_1 denote $RRA(I_{\text{hazy}})$ or I_{gt} . G_1 is downsampled 3 times successively by Gaussian filtering to generate G_k , $k = 2, 3, 4$. The Laplace pyramid is constructed as $LP_k = G_k - u(G_{k+1})$, $k = 1, 2, 3$, where u is the bilinear interpolation upsampling operation. Finally, the proposed Laplace pyramid loss is calculated as

$$L_{\text{Laplace}} = \frac{1}{N} \sum_{k=1}^3 \sum_{i=1}^N \left\| LP_k(RRA(I_{\text{hazy}}^i)) - LP_k(I_{\text{gt}}^i) \right\|_2^2. \quad (9)$$

Color Attenuation Loss

The saturation and brightness of an image are influenced by nonhomogeneous haze. The saturation of the hazy area usually is decreased because the color fades under haze, while the brightness of the hazy area is usually increased. A color attenuation prior was presented in [11] where the haze concentration $P(x)$ at position x can be calculated by the following formula:

$$P(x) = \|S(I(x)) - V(I(x))\|_1, \quad (10)$$

where $S(I(x))$ is the saturation at position x of image I , and $V(I(x))$ is the brightness at position x of image I . Inspired by this prior, we designed a color attenuation loss L_{CA} to regulate the RRA-Net to recover the saturation and brightness. L_{CA} is calculated as:

$$L_{\text{CA}} = \alpha \frac{1}{N} \sum_{i=1}^N \left\| S(RRA(I_{\text{hazy}}^i)) - S(I_{\text{gt}}^i) \right\|_2^2 + \beta \frac{1}{N} \sum_{i=1}^N \left\| V(RRA(I_{\text{hazy}}^i)) - V(I_{\text{gt}}^i) \right\|_2^2, \quad (11)$$

where α and β are the balancing coefficients. Empirically, the values of α and β are set to 1 and 0.5, respectively.

3.3. Experimental Validation

To evaluate the performance of RRA-Net on the single-image dehazing task, we compare it with other state-of-the-art methods both quantitatively and qualitatively. To this end, we construct the training dataset to train the models and the testing dataset for model evaluation. Two evaluation metrics are calculated to compare different models. We also conduct a series of ablation studies to verify the effectiveness of the main components of the proposed RRA-Net.

In this subsection, the datasets and evaluation metrics used in the experiments are introduced first, followed by some implementation details. Finally, the comparative experiments and ablation studies are described.

3.3.1. Datasets

Since RRA-Net is designed to dehaze realistic images, we train and test RRA-Net on real-world hazy image datasets. We take the images from the I-Haze [33] dataset, O-Haze [34] dataset, and NH-Haze (2020) [2] dataset together to form a training set. The I-Haze dataset contains 35 image pairs of hazy and corresponding haze-free (ground-truth) indoor images. The O-Haze dataset consists of 45 pairs of hazy and corresponding haze-free outdoor images. In practice, the hazy images have been captured in presence of real haze, generated by professional haze machines. The NH-Haze (2020) dataset is the first non-homogeneous image-dehazing dataset and contains 55 outdoor image pairs, among which 50 pairs are usually used for training and 5 pairs are for testing. The non-homogeneous haze has been introduced in the scene using a professional haze generator that imitates real conditions of hazy scenes. In our experiments, the constructed training set consists of 130 image pairs in total, including 35 pairs from the I-Haze dataset, 45 pairs from the O-Haze dataset, and 50 pairs from the NH-Haze (2020) dataset.

As for the test dataset, we use the NH-Haze (2021) [35] dataset and the Dense-Haze [36] dataset. The NH-Haze (2021) dataset is a non-homogeneous realistic dataset with pairs of real hazy and corresponding haze-free images. Additionally, the Dense-Haze dataset is characterized by dense and homogeneous hazy scenes.

3.3.2. Evaluation Metrics

To quantitatively evaluate the performance of our RRA-Net, we adopt the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [37] as evaluation metrics.

The PSNR is the ratio between the maximum possible power of a signal and the power of discrepancy that affects the quality of its representation. It is defined via the Mean Squared Error (MSE). Given an $H \times W$ dehazed image $I_{dehazed}$ and its corresponding haze-free image I_{gt} , the MSE is defined as

$$MSE = \frac{1}{HW} \sum_x \| I_{dehazed}(x) - I_{gt}(x) \|_2^2, \quad (12)$$

where x indexes the pixel position of the image. Then, the PSNR (in dB) is defined as

$$PSNR = 10 \cdot \log_{10} \left(\frac{255^2}{MSE} \right). \quad (13)$$

On the other hand, the SSIM is a quality assessment metric based on the degradation of structural information. It extracts three key features from an image, namely the luminance, contrast, and structure. The comparison between $I_{dehazed}$ and I_{gt} is performed using three comparison functions on the basis of these three features. Then, the SSIM is defined as

$$SSIM = [l(I_{dehazed}, I_{gt})]^\alpha \cdot [c(I_{dehazed}, I_{gt})]^\beta \cdot [s(I_{dehazed}, I_{gt})]^\gamma, \quad (14)$$

where $l(I_{dehazed}, I_{gt})$, $c(I_{dehazed}, I_{gt})$, and $s(I_{dehazed}, I_{gt})$ are the luminance, contrast, and structure comparison functions, respectively, as defined in [37], and $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ are the parameters used to adjust the relative importance of the three components. We set $\alpha = \beta = \gamma = 1$ in the experiments since this setting is commonly used.

3.3.3. Implementation Details

During training, since there are only 130 pairs of a hazy image and its ground-truth, the training dataset is augmented with random rotation and horizontal flip. Then, we divide the hazy image and its ground-truth into image patches of the size 128×128 and feed them to RRA-Net. The network is trained for 5×10^5 steps on the training dataset. We use the Adam optimizer whose parameters β_1 and β_2 are set to 0.9 and 0.999, respectively. The initial learning rate is set to 6×10^{-4} , and we adopt the cyclical learning rate strategy [38] to adjust the learning rate from the initial value to 1.2×10^{-3} with a step size of 10. The RRA-Net model is implemented in Pytorch [39] and run on 2 RTX 2080Ti GPUs.

During testing, we pass each whole image from the testing dataset into RRA-Net and other models directly and average the PSNR and SSIM metrics over the testing dataset.

3.3.4. Comparative Experiments and Ablation Studies

Comparison with State-of-the-Art Methods

We compare our RRA-Net with 7 single-image dehazing methods, including the traditional DCP [9] and SOTA deep-learning-based methods. Among the latter type of methods, AOD-Net [25], FFA-Net [13], MSBDN [14], DMPHN [15], and TBNN [19] are based on CNNs, and the Dehamer [20] is based on a transformer. The evaluations are conducted on the NH-Haze (2021) and Dense-Haze datasets.

For each compared model, the PSNR and SSIM metrics are calculated between the dehazed results and the haze-free images of the testing set. A higher PSNR score or a higher SSIM score indicates a better dehazing performance.

To validate the efficiency of the proposed RRA-Net, we also conduct a comparative runtime test. Specifically, images of the size 1600×1200 are passed into each compared model. Additionally, then, the average processing FPS of each model is calculated. A higher FPS score indicates a better dehazing speed and, hence, a higher efficiency.

Ablation Studies

To validate our design choices for RRA-Net, we conduct a series of ablation studies on modules of the MRAB structure, two newly designed loss functions, and the batch normalization strategy. In each ablation study, the PSNR and SSIM metrics are calculated to show how these design choices affect the dehazing performance of RRA-Net.

4. Results

In this section, the results of the comparative experiments are presented and discussed first to evaluate the dehazing performance of the proposed RRA-Net. Then, the results of the ablation studies are given to validate several design choices for the proposed model.

4.1. Results of Comparative Experiments

As shown in Table 1, our RRA-Net outperforms all of the listed SOTA methods on both datasets in terms of the PSNR and SSIM metrics. For the NH-Haze (2021) dataset with nonhomogeneous haze scenes, RRA-Net has a 0.5dB lead in terms of the PSNR over the second-best method. Additionally, note that RRA-Net works well on both nonhomogeneous hazy images and homogeneous dense hazy images. For the Dense-Haze dataset, RRA-Net achieves a PSNR of 15.78 dB, exceeding the second-best method by 0.37 dB.

Table 1. Quantitative comparisons with SOTA methods on the realistic datasets. Best results are underlined. The second-best results are in **Bold**.

Method	NH-Haze (2021)		Dense-Haze		Param	Runtime (fps) @1600 × 1200
	PSNR↑	SSIM↑	PSNR↑	SSIM↑		
DCP [9] (TPAMI'10)	10.57	0.52	10.06	0.3856	-	-
AOD-Net [25] (TIP'17)	14.104	0.552	13.34	0.4244	0.002M	2598.3
FFA-Net [13] (AAAI'20)	19.1	0.748	14.31	0.4797	4.68M	1.28
MSBDN [14] (CVPR'20)	19.31	0.759	15.41	0.4858	31.35M	36.84
DMPHN [15] (CVPRW'20)	18.184	0.745	14.01	0.4436	5.424M	135.46
TBNN [19] (CVPR'21)	15.622 ¹	0.707	14.60	0.4829	50.40M	- ¹
Dehamer [20] (CVPR'22)	13.940 ¹	0.646	14.63	0.4996	132.4M	- ¹
RRA-Net (Ours)	<u>19.813</u>	<u>0.765</u>	<u>15.78</u>	<u>0.5154</u>	0.3M	166.11

¹ For fair comparisons, we conducted the tests for all the methods on the same computer with 2 RTX 2080Ti GPUs by passing whole images into these models. For TBNN and Dehamer models, GPU "out of memory" issue was reported during the tests run on this specific computer since they have much higher model complexities. Thus, we used a patching strategy to quantitatively evaluate these two models. Note that this strategy may incur slight losses of performance. Accordingly, the runtimes of these two models are not reported.

Visual comparisons of the recovered results on nonhomogeneous hazy scenes are given in Figures 4 and 5. It can be seen that RRA-Net produces results with less residual haze and preserves more image details and contrast. For example, in Figure 4, the results of the FFA-Net and TBNN methods still have some residual haze; the result of MSBDN loses pixel color saturation to some extent and has a whitish appearance, while the result of DMPHN introduces visible artifacts in the sky area. In addition, the Dehamer method cannot effectively remove the nonhomogeneous haze from either sample given in Figures 4 and 5.

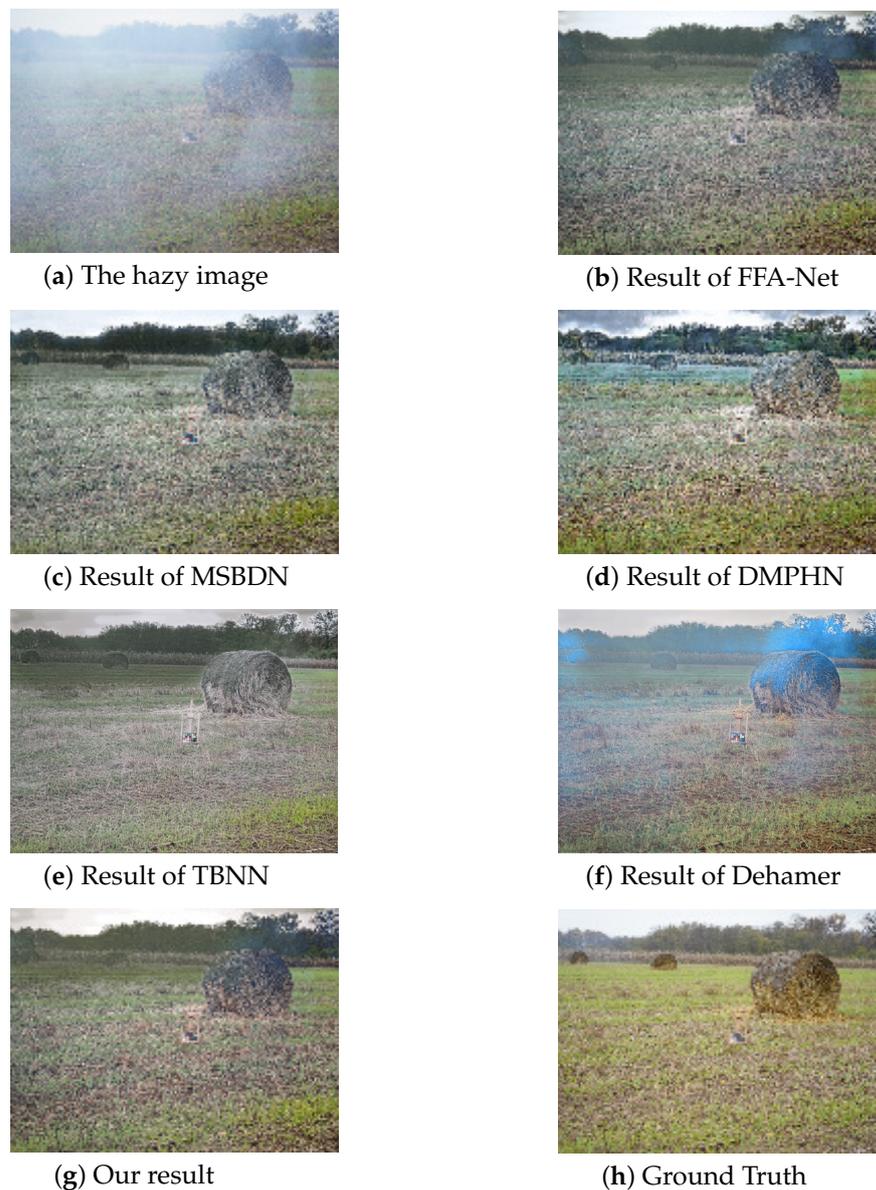


Figure 4. Qualitative comparison of dehazing results on one nonhomogeneous hazy image from NH-Haze (2021) dataset.

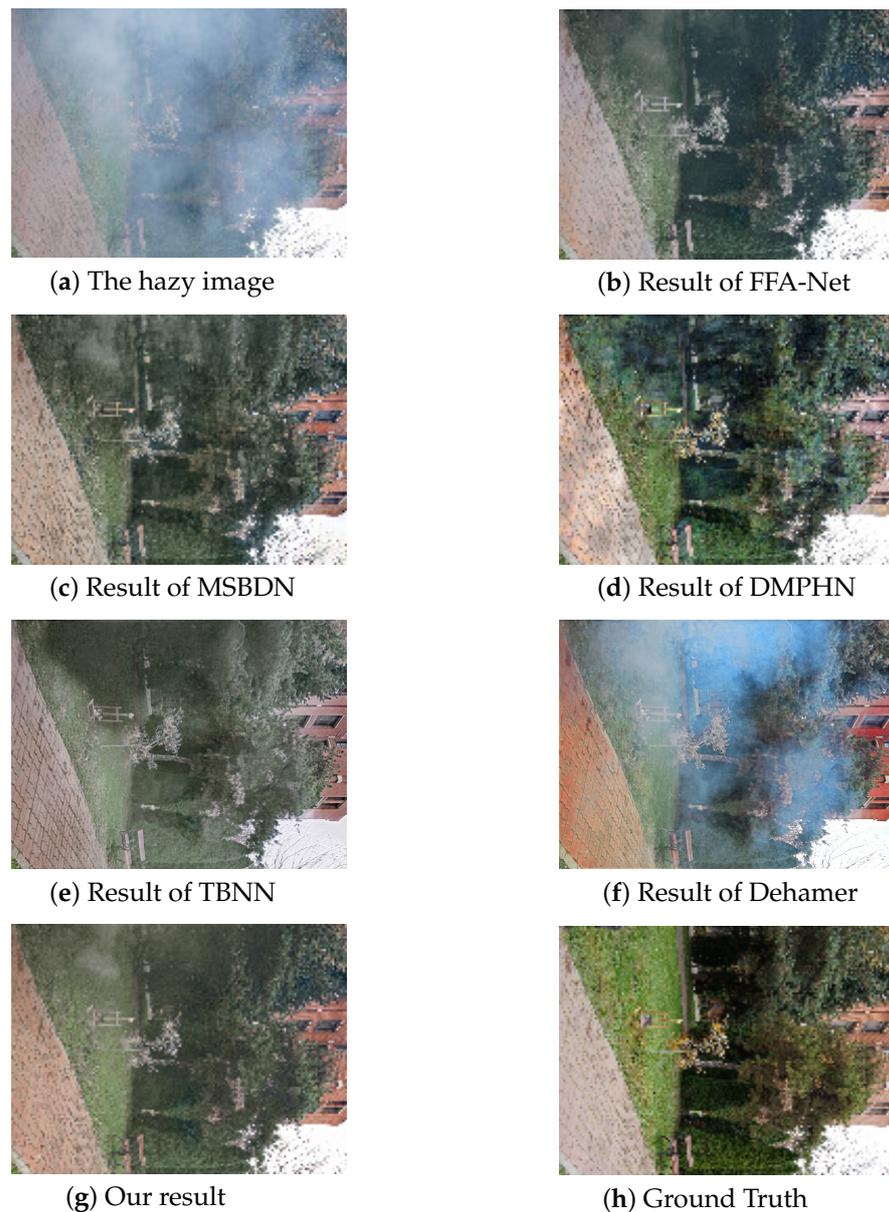


Figure 5. Qualitative comparison of dehazing results on another nonhomogeneous hazy image from NH-Haze (2021) dataset.

A visual comparison of the homogeneous dense hazy scene is given in Figure 6. It can be seen that in such a case, RRA-Net is also capable of producing a result with less residual haze.

Apart from having better dehazing results, RRA-Net is designed to be lightweight from the beginning. Under the same configuration, it has only about 0.3M parameters and runs at 166 FPS to dehaze a 1600×1200 image on RTX 2080Ti GPU. As can be seen from Table 1, RRA-Net runs faster than the other methods, except for AOD-Net, which falls behind in terms of the PSNR and SSIM metrics. The speed comparison is conducted under the same hardware setup, demonstrating the superior efficiency of the proposed network.

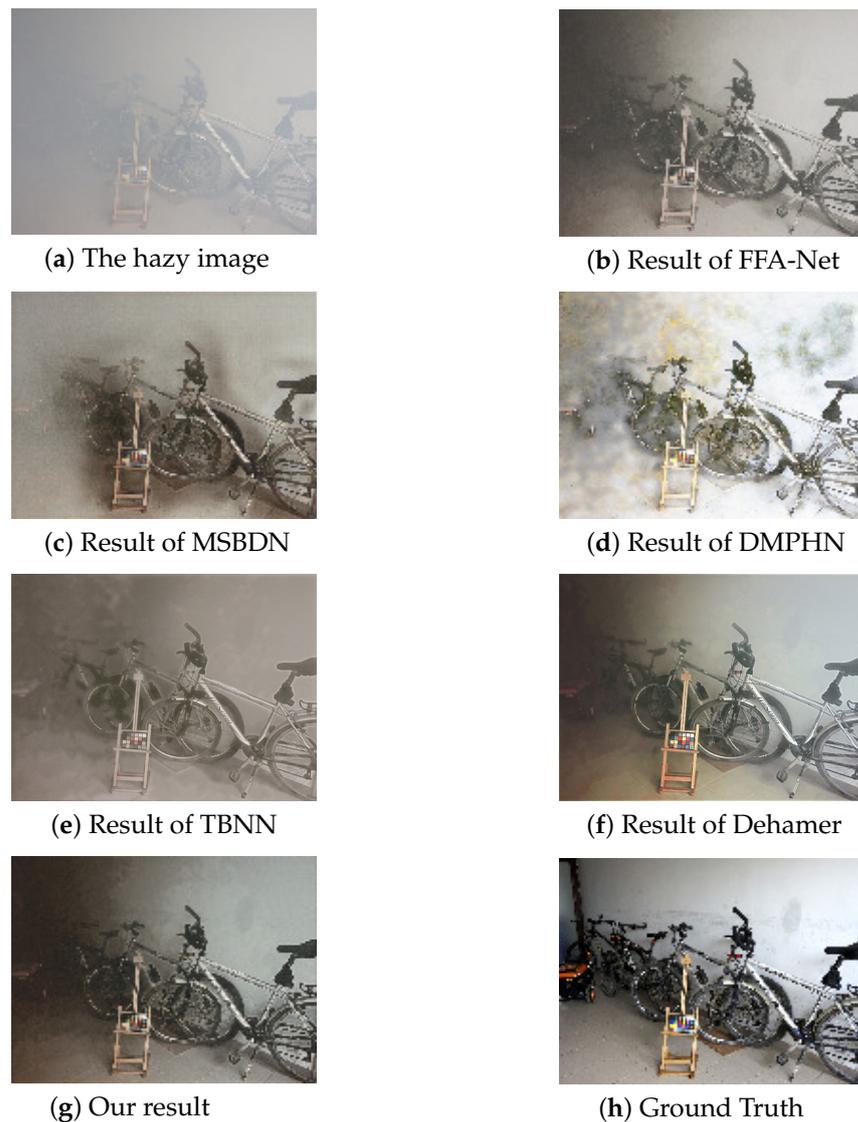


Figure 6. Qualitative comparison of dehazing results on an image with dense haze from the Dense-Haze dataset.

4.2. Results of Ablation Studies

The ablation studies are conducted to validate several key choices in the design of the proposed RRA-Net, whose results are described as follows.

4.2.1. Effect of MRAB Structure

We first verify the effectiveness of different modules of the basic MRAB via ablation experiments on the NH-Haze (2021) dataset.

The results are listed in the top rows of Table 2. In the S1 setting, the MRAB consists of only multi-branch convolutions and the LeakyReLU layer. On top of S1, we applied the other components to form different settings: S2 (apply the BN layer and Attention Module) and RRA-Net (apply the BN layer, attention module, and local residual connection). It can be seen that the inclusion of the BN layer and attention mechanism greatly improve the dehazing performance of the network, and the local residual connection is also beneficial.

Table 2. Ablation studies on main modules of MRAB (top rows) and on two new loss functions (bottom rows). *BN*, *AM*, and *LR* stand for batch normalization, attention module (spatial attention and channel attention), and local residual connection in MRAB, respectively. Best results are underlined.

Setting	Module/Loss						Metric	
	<i>BN</i>	<i>AM</i>	<i>LR</i>	L_1	L_{CA}	$L_{Laplace}$	PSNR↑	SSIM↑
S1				✓	✓	✓	18.797	0.733
S2	✓	✓		✓	✓	✓	19.532	0.754
RRA-Net	✓	✓	✓	✓	✓	✓	<u>19.813</u>	<u>0.765</u>
S3	✓	✓	✓	✓			19.21	0.746
S4	✓	✓	✓	✓	✓		19.34	0.741
RRA-Net	✓	✓	✓	✓	✓	✓	<u>19.813</u>	<u>0.765</u>

4.2.2. Effect of Two Novel Loss Functions

Secondly, the effectiveness of the two newly proposed loss functions is also verified by the ablation study, as listed in the bottom rows of Table 2. In the S3 setting, only the L_1 loss function is used during training. On top of that, the L_{CA} loss function is added to the S4 setting, which regulates the network to recover the saturation and brightness. It can be seen that S4 enjoys a moderate PSNR gain, while suffering a slight loss in terms of the SSIM. Since L_{CA} is calculated in a pixel-wise way, it does not focus on the recovery of the structural information, which is evaluated by the SSIM. Finally, in the RRA-Net setting, both the L_{CA} and $L_{Laplace}$ loss functions are used during training. With the help of $L_{Laplace}$, which is designed to protect high-frequency features, both the PSNR and SSIM metrics are lifted substantially from 19.21 dB and 0.746 to 19.813 dB and 0.765, respectively, as shown in the last row of Table 2. Through this ablation study, we find that using a combination of L_{CA} and $L_{Laplace}$ can boost the dehazing performance effectively since these two loss functions are designed from different perspectives.

4.2.3. Effect of the Batch Normalization Strategy

During training, we use a single BN layer after summing up the outputs of three convolution branches, rather than applying individual BN layers on each branch as applied in RepVGG. Table 3 shows the performance of these two different BN strategies. It can be seen that the BN strategy we propose results in much higher PSNR and SSIM scores, demonstrating that applying a single BN after multi-branch convolutions can stabilize the training more effectively and demonstrate improvement over the original strategy in RepVGG.

Table 3. Ablation study on different BN strategies. Best results are underlined.

BN Strategy	PSNR	SSIM
Individual BN on each branch	18.627	0.688
A single BN after multi-branch	<u>19.813</u>	<u>0.765</u>

5. Conclusions and Further Studies

This paper proposes an efficient RRA-Net to perform single-image dehazing. The contributions of the work are a training-time multi-branch residual attention block, an end-to-end lightweight network structure, and two new loss functions. Multi-branch convolutions in the MRAB are able to deal with non-uniformity of haze. Additionally, due to lightweight design, RRA-Net only has about 0.3M parameters, achieving a good balance between the dehazing performance and model complexity. The experimental results demonstrate that RRA-Net outperforms other SOTA methods on both realistic nonhomogeneous and homogeneous hazy image datasets.

With the help of its lightweight network structure, RRA-Net is able to dehaze 1600×1200 HD images at a speed of 166 fps on a computer with 2 RTX 2080Ti GPUs, demonstrating its application potential in real-time dehazing. RRA-Net can not only produce high-quality

clean images for better human perception but also contribute to the success of subsequent real-world vision tasks in hazy weather, such as object detection and autonomous driving.

Despite the superiority of RRA-Net, it is CNN-based and lacks ability for global modeling. As mentioned earlier, recently, there have been some efforts to take advantage of both visual transformers and CNNs. However, such hybrid networks usually have a much higher complexity. Thus, it is worth investigating how to bestow RRA-Net with the merits of transformers while keeping the network as efficient as possible at the same time.

Author Contributions: Conceptualization, E.C. and T.Y.; methodology, E.C. and T.Y.; software, T.Y. and J.J.; validation, J.J. and L.T.; formal analysis, Q.Y. and L.T.; investigation, E.C. and T.Y.; resources, Q.Y.; data curation, T.Y.; writing—original draft preparation, T.Y., J.J. and Q.Y.; project administration, Q.Y.; funding acquisition, E.C. and Q.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Fujian Province of China under grant no. 2021J01867, the Scientific Research Foundation of Jimei University under grant no. ZQ2018012, and the Xiamen Municipal Bureau of Ocean Development under grant no. 22CZB013HJ04.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: <https://data.vision.ee.ethz.ch/cvl/ntire18/o-haze/> (accessed on 15 January 2021), <https://data.vision.ee.ethz.ch/cvl/ntire18/i-haze/> (accessed on 15 January 2021), <https://data.vision.ee.ethz.ch/cvl/ntire20/nh-haze/> (accessed on 1 April 2021) and <https://data.vision.ee.ethz.ch/cvl/ntire19/dense-haze/> (accessed on 15 January 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AOD-Net	All-in-One Dehazing Network
BN	Batch Normalization
CA	Color Attenuation
CNN	Convolutional Neural Network
DCP	Dark Channel Prior
DMPHN	Deep Multi-Patch Hierarchical Network
FFA-Net	Feature Fusion Attention Network
GAN	Generative Adversarial Network
MRAB	Multi-branch Residual Attention Block
MSBDN	Multi-Scale Boosted Dehazing Network
PSNR	Peak Signal-to-Noise Ratio
RRA-Net	Re-parameterization Residual Attention Network
SSIM	Structural Similarity Index
TBNN	Two-Branch Neural Network
TCAM	Transformer-based Channel Attention Module

References

1. Agrawal, S.C.; Jalal, A.S. A Comprehensive Review on Analysis and Implementation of Recent Image Dehazing Methods. *Arch. Comput. Methods Eng.* **2022**, *29*, 4799–4850. [CrossRef]
2. Ancuti, C.O.; Ancuti, C.; Timofte, R. 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, Seattle, WA, USA, 14–19 June 2020; pp. 444–445. [CrossRef]
3. Gui, J.; Cong, X.; Cao, Y.; Ren, W.; Zhang, J.; Zhang, J.; Tao, D. A Comprehensive Survey on Image Dehazing Based on Deep Learning. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, Virtual, 19–26 August 2021; pp. 4426–4433. [CrossRef]

4. Singh, D.; Kumar, V. A comprehensive review of computational dehazing techniques. *Arch. Comput. Methods Eng.* **2019**, *26*, 1395–1413. [[CrossRef](#)]
5. Li, Y.; You, S.; Brown, M.S.; Tan, R.T. Haze visibility enhancement: A survey and quantitative benchmarking. *Comput. Vis. Image Underst.* **2017**, *165*, 3. [[CrossRef](#)]
6. Harish Babu, G.; Venkatram, N. A survey on analysis and implementation of state-of-the-art haze removal techniques. *J. Vis. Commun. Image Represent.* **2020**, *72*, 102912. [[CrossRef](#)]
7. Wang, W.; Yuan, X. Recent advances in image dehazing. *IEEE/CAA J. Autom. Sin.* **2017**, *4*, 410–436. [[CrossRef](#)]
8. Ye, T.; Zhang, Y.; Jiang, M.; Chen, L.; Liu, Y.; Chen, S.; Chen, E. Perceiving and Modeling Density for Image Dehazing. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2022; Springer: Cham, Switzerland, 2022; pp. 130–145. [[CrossRef](#)]
9. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353. [[CrossRef](#)] [[PubMed](#)]
10. Tan, R.T. Visibility in bad weather from a single image. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8. [[CrossRef](#)]
11. Zhu, Q.; Mai, J.; Shao, L. Single Image Dehazing Using Color Attenuation Prior. In Proceedings of the British Machine Vision Conference, Nottingham, UK, 1–5 September 2014; BMVA Press: Nottingham, UK, 2014. [[CrossRef](#)]
12. Liu, X.; Ma, Y.; Shi, Z.; Chen, J. Griddehazenet: Attention-based multi-scale network for image dehazing. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7314–7323. [[CrossRef](#)]
13. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11908–11915. [[CrossRef](#)]
14. Dong, H.; Pan, J.; Xiang, L.; Hu, Z.; Zhang, X.; Wang, F.; Yang, M.H. Multi-scale boosted dehazing network with dense feature fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 2157–2167. [[CrossRef](#)]
15. Das, S.D.; Dutta, S. Fast deep multi-patch hierarchical network for nonhomogeneous image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 482–483. [[CrossRef](#)]
16. Bu, Q.; Luo, J.; Ma, K.; Feng, H.; Feng, J. An enhanced pix2pix dehazing network with guided filter layer. *Appl. Sci.* **2020**, *10*, 5898. [[CrossRef](#)]
17. Chen, Z.; Wang, Y.; Yang, Y.; Liu, D. PSD: Principled Synthetic-to-Real Dehazing Guided by Physical Priors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 7180–7189. [[CrossRef](#)]
18. Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; Ma, L. Contrastive learning for compact single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10551–10560. [[CrossRef](#)]
19. Yu, Y.; Liu, H.; Fu, M.; Chen, J.; Wang, X.; Wang, K. A Two-Branch Neural Network for Non-Homogeneous Dehazing via Ensemble Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Nashville, TN, USA, 20–25 June 2021; pp. 193–202. [[CrossRef](#)]
20. Guo, C.L.; Yan, Q.; Anwar, S.; Cong, R.; Ren, W.; Li, C. Image Dehazing Transformer with Transmission-Aware 3D Position Embedding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 5812–5820. [[CrossRef](#)]
21. Berman, D.; Avidan, S.; Treibitz, T. Non-local image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1674–1682. [[CrossRef](#)]
22. Bui, T.M.; Kim, W. Single Image Dehazing Using Color Ellipsoid Prior. *IEEE Trans. Image Process.* **2018**, *27*, 999–1009. [[CrossRef](#)] [[PubMed](#)]
23. Yuan, F.; Zhou, Y.; Xia, X.; Qian, X.; Huang, J. A confidence prior for image dehazing. *Pattern Recognit.* **2021**, *119*, 108076. [[CrossRef](#)]
24. Cai, B.; Xu, X.; Jia, K.; Qing, C.; Tao, D. Dehazenet: An end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **2016**, *25*, 5187–5198. [[CrossRef](#)] [[PubMed](#)]
25. Li, B.; Peng, X.; Wang, Z.; Xu, J.; Feng, D. Aod-net: All-in-one dehazing network. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4770–4778. [[CrossRef](#)]
26. Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; Yang, M.H. Single image dehazing via multi-scale convolutional neural networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 154–169. [[CrossRef](#)]
27. Li, Z.; Zheng, C.; Shu, H.; Wu, S. Dual-Scale Single Image Dehazing via Neural Augmentation. *IEEE Trans. Image Process.* **2022**, *31*, 6213–6223. [[CrossRef](#)] [[PubMed](#)]
28. Xu, J.; Chen, Z.X.; Luo, H.; Lu, Z.M. An Efficient Dehazing Algorithm Based on the Fusion of Transformer and Convolutional Neural Network. *Sensors* **2023**, *23*, 43. [[CrossRef](#)] [[PubMed](#)]

29. Gao, G.; Cao, J.; Bao, C.; Hao, Q.; Ma, A.; Li, G. A Novel Transformer-Based Attention Network for Image Dehazing. *Sensors* **2022**, *22*, 3428. [[CrossRef](#)] [[PubMed](#)]
30. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. SwinIR: Image Restoration Using Swin Transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844. [[CrossRef](#)]
31. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 17683–17693. [[CrossRef](#)]
32. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. Repvgg: Making vgg-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13733–13742. [[CrossRef](#)]
33. Ancuti, C.O.; Ancuti, C.; Timofte, R.; Vleeschouwer, C.D. I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images. *arXiv* **2018**, arXiv:1804.05091v1.
34. Ancuti, C.O.; Ancuti, C.; Timofte, R.; De Vleeschouwer, C. 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, Salt Lake City, UT, USA, 18–22 June 2018; pp. 754–762. [[CrossRef](#)]
35. Ancuti, C.O.; Ancuti, C.; Vasluianu, F.A.; Timofte, R. NTIRE 2021 nonhomogeneous dehazing challenge report. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Nashville, TN, USA, 19–25 June 2021; pp. 627–646. [[CrossRef](#)]
36. Ancuti, C.O.; Ancuti, C.; Sbert, M.; Timofte, R. Dense haze: A benchmark for image dehazing with dense-haze and haze-free images. In Proceedings of the IEEE International Conference on Image Processing, (IEEE ICIP 2019), Taipei, Taiwan, 22–25 September 2019. [[CrossRef](#)]
37. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
38. Smith, L.N. Cyclical learning rates for training neural networks. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; IEEE: New York, NY, USA, 2017; pp. 464–472. [[CrossRef](#)]
39. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic Differentiation in PyTorch 2017. Available online: <https://openreview.net/pdf?id=BJjsrmfCZ> (accessed on 15 January 2020).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.