Article

Deep Dehazing Network for Remote Sensing Image with Non-Uniform Haze

Bo Jiang [ID]**, Guanting Chen, Jinshuai Wang, Hang Ma, Lin Wang** [ID]**, Yuxuan Wang and Xiaoxuan Chen ***

School of Information Science and Technology, Northwest University, Xi'an 710127, China; jiangbo@nwu.edu.cn (B.J.); chenguanting@stumail.nwu.edu.cn (G.C.); wangjinshuai@stumail.nwu.edu.cn (J.W.); mahang1@stumail.nwu.edu.cn (H.M.); wanglin@nwu.edu.cn (L.W.); 201932091@stumail.nwu.edu.cn (Y.W.)
* Correspondence: chenxx@nwu.edu.cn

Abstract: The haze in remote sensing images can cause the decline of image quality and bring many obstacles to the applications of remote sensing images. Considering the non-uniform distribution of haze in remote sensing images, we propose a single remote sensing image dehazing method based on the encoder–decoder architecture, which combines both wavelet transform and deep learning technology. To address the clarity issue of remote sensing images with non-uniform haze, we preliminary process the input image by the dehazing method based on the atmospheric scattering model, and extract the first-order low-frequency sub-band information of its 2D stationary wavelet transform as an additional channel. Meanwhile, we establish a large-scale hazy remote sensing image dataset to train and test the proposed method. Extensive experiments show that the proposed method obtains greater advantages over typical traditional methods and deep learning methods qualitatively. For the quantitative aspects, we take the average of four typical deep learning methods with superior performance as a comparison object using 500 random test images, and the peak-signal-to-noise ratio (PSNR) value using the proposed method is improved by 3.5029 dB, and the structural similarity (SSIM) value is improved by 0.0295, respectively. Based on the above, the effectiveness of the proposed method for the problem of remote sensing non-uniform dehazing is verified comprehensively.

Keywords: remote sensing images; non-uniform haze; deep learning; image dehazing

1. Introduction

Remote sensing (RS) observations can be divided into two categories: the satellite RS and the aerial RS, according to the platforms they rely on. We mainly focus on the research of the aerial RS images in this paper. RS images taken by the aerial platforms benefit from rich information, high spatial resolutions, and stable geometric locations and they have already been widely used in meteorology, agriculture, the military, and other fields. However, RS images are particularly vulnerable to weather factors. The particles that suspend in the air, e.g., water vapor, clouds, and haze, easily weaken the light reflected from an object's surface. This attenuation may result in image degradation phenomena, such as contrast reduction, color distortion, and unclear detail information in the observed RS images [1]. It brings many negative impacts on the ground objects classification, recognition, tracking, and other advanced applications based on the RS images. Effective dehazing for RS images can decrease the impact of hazy weather on the RS imaging system, which is vital to the later advanced applications of RS images [2].

RS images are different from the ground images. On the one hand, the quality of the RS images that are obtained even under the haze-free conditions is equivalent to the quality of images that are shot on the ground under the hazy conditions because the light is largely attenuated, which is caused by the air particles due to the long imaging paths. Therefore, the contrast, saturation, and color fidelity of the captured RS images are not good under normal circumstances. On the other hand, the distribution of haze in RS images is also

quite different from that in the ground images. In severe weather conditions, e.g., haze, since the ground imaging is obtained in the short imaging distance, the imaging field of view is very limited. Although the ground imaging also has the non-uniform distribution of haze, the haze on the captured image shows uniform distribution characteristics due to the limited imaging field of view. There is a great difference between the RS images and ground images. Given the large imaging distances, the obtained RS images contain a large range of ground coordinates, and it is difficult to ensure uniform distribution of haze in such a large ground range. The RS images under the hazy weather have a typical non-uniform haze distribution. In summary, the dehazing for non-uniform haze remote sensing (NHRS) images has more practical significance for the research on the clarity of RS images.

Most of the traditional dehazing methods are based on the atmospheric scattering model that was first proposed by McCartney [3]. A detailed derivation and description are carried out by Narasimhan and Nayar [4,5]. This model can be formulized as:

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

where $I(x)$ denotes the observed hazy image, $J(x)$ denotes the haze-free image, A refers to the global atmospheric light, $t(x)$ refers to the transmission, and x represents the corresponding pixel. The dehazing process is to estimate A and $t(x)$ from $I(x)$, and then recover the ultimate haze-free image $J(x)$. In Equation (1), $I(x)$ is a known parameter, while the other parameters are unknown. It is critical to estimate A and $t(x)$. The problem of solving multiple unknowns from one equation is a typical ill-conditioned problem in mathematics. Therefore, in the actual solving process, the statistical prior knowledge of physical quantities is usually fully utilized to achieve the purpose of solving multiple physical quantities in the equation.

Before the advent of deep neural networks, dehazing methods were mostly based on the above-mentioned atmospheric scattering model. With the development of deep learning technology, some dehazing methods based on deep learning have been proposed for the task of ground image dehazing. However, since RS images correspond to large imaging fields of view in most cases, the haze presents in the form of non-uniform distribution in the RS images. Therefore, most of the existing dehazing methods, including deep learning networks, are unsuitable for the non-uniform dehazing tasks, that is, the problem of RS image dehazing cannot be well solved by these methods due to the difference in the imaging characteristics between ground images and RS images.

Compared with the ground imaging for a natural scene, the imaging distance of RS image is greatly increased, and the imaging optical path could encounter the non-uniform haze distribution. As a result, the light with different wavelengths no longer has the same or nearly the same scattering coefficient, which leads to the typical dehazing method encountering the issue that the atmospheric scattering model is difficult to solve. Moreover, since atmospheric scattering model only takes the short-distance single scattering into account, this model has certain limitations on long-distance imaging. In recent years, the deep learning method has achieved great success in the field of image processing as it can solve complex issues. However, directly applying it to the problem of non-uniform dehazing for RS images will mainly face the following challenges:

Firstly, in the field of image dehazing, the commonly used "haze and haze-free" datasets include the NYU2 dataset, RESIDE dataset, Middlebury Stereo dataset, etc., all of which contain datasets of ground natural scenes with uniform haze distribution. For the task of RS dehazing, the NHRS dataset is relatively scarce, which brings a challenge to the data-dependent dehazing method of deep learning. Secondly, RS images have a long imaging distance, which leads to the general fuzziness of the texture and edge details in the collected images. In addition, compared with ground imaging, the imaging field of view of RS images is greatly increased, resulting in a significant decrease in the correlation between pixels in RS images, which makes deep feature learning in deep learning networks encounter challenges.

In this paper, we research the aforementioned challenges, and the main contributions of this paper are as follows:

- Firstly, a single RS image dehazing method, which combines both wavelet transform and deep learning technology, is proposed. We employ the atmospheric scattering model and 2D stationary wavelet transform (SWT) to process a hazy image, and extract the low-frequency sub-band information of the processed image as the enhanced features to further strengthen the learning ability of the deep network for low-frequency smooth information in RS images.
- Secondly, our dehazing method is based on the encoder–decoder architecture. The inception structure in the encoder can increase the multi-scale information and learn the abundant image features for our network. As the hybrid convolution in the encoder combines standard convolution with dilated convolution, it expands the receptive field to better improve the ability of detecting the non-uniform haze in RS images. The decoder fusions the shallow feature information of the network through multiple residual blocks to recover the detailed information of the RS images.
- Thirdly, a special design in the aspect of loss function is made for the non-uniform dehazing task of RS images. As the scene structure edges of an RS image itself are usually weak, the structure pixels are weakened more seriously after dehazing. Therefore, on the basis of the L1 loss function, we employ the multi-scale structural similarity index (MS-SSIM) and Sobel edge detection as the loss function to make the dehazed image more natural and improve the edge of the dehazed RS images.
- Lastly, aiming at the problem that a deep learning network depends on the support of high-quality datasets, we propose a non-uniform haze-adding algorithm to establish a large-scale hazy RS image dataset. We employ the transmission of the real hazy image and the atmospheric scattering model in the RGB color space to obtain the RGB synthetic hazy image. The haze in a hazy image is mainly distributed on the Y channel component of the YCbCr color space. Based on this distribution characteristic of haze, the RGB synthetic hazy image and the haze-free image are jointly corrected to obtain the final synthetic NHRS image in the YCbCr color space.

The remainder of the paper is organized as follows: Firstly, we review the related research in the field of image dehazing in Section 2, then introduce the proposed deep learning dehazing method detailedly in Section 3. The performance evaluation of the proposed network for image dehazing is conducted in Section 4, which also includes a description of the dataset and the training procedure. Finally, Section 5 is the conclusion.

2. Related Work

At present, the methods of image dehazing can be roughly divided into two categories. The first one is the dehazing method based on the atmospheric scattering model, which is mainly based on the traditional methods of mathematical models. These methods have been developed for a long time and are relatively mature. The second one is the modern dehazing method based on deep learning in recent years.

2.1. Traditional Dehazing Methods

The basic principle of the traditional dehazing methods is to rely on the artificial extraction of features associated with haze, such as dark channel, tonal difference, and local contrast, etc. Specifically, the dehazing method of the atmospheric scattering model based on Equation (1) mostly uses statistical prior knowledge to estimate physical quantities , such as the transmission, to implement dehazing. Typical methods are as follows: He et al. [6] proposed the dark channel prior (DCP) dehazing method and found that the darkest pixel value in the RGB three-channel image block of the haze-free image is close to zero (excluding sky area). The DCP is based on the statistics of outdoor haze-free images, and it makes image dehazing simple and effective. Fattal et al. [7] proposed a method to estimate the transmission by using the locally uncorrelated characteristic of the transmission and surface shading, and the scattered light is eliminated to increase the scene visibility and

recover haze-free scene contrasts. Berman et al. [8] introduced a non-local prior dehazing (NLD) method to identify haze-lines and estimated the per-pixel transmission in the hazy image space coordinate system. They found that the colors of the haze-free image formed a tight cluster in the RGB space after being approximated to a few hundred distinct colors. Zhu et al. [9] proposed a color attenuation prior (CAP) to the single image dehazing method, which created a linear model of the scene depth of hazy images and combined with a supervised learning method to recover the depth information. With the help of the depth maps of hazy images, the transmission can be easily estimated, and the radiation of the scene can be recovered by the atmospheric scattering model, so as to effectively remove the haze from a single hazy image.

The above-mentioned traditional dehazing methods have preferable dehazing results on uniform distribution hazy images of the ground. Due to the complex structures of RS imaging scenes and the large imaging fields of view, the traditional dehazing methods are difficult to apply directly to the restoration of NHRS images.

2.2. Dehazing Methods Based on Deep Learning

Although single image dehazing is a challenging task, the human brain can quickly distinguish the haze concentration in the natural scenes without any additional information. Based on this, the deep learning methods that simulate the mechanism of the human brain are used in the research of image dehazing. Deep neural networks have been successfully applied to many computer vision and signal processing tasks, such as image clarity, object/action recognition [10], emotion recognition [11], etc.

The typical methods are as follows: Mei et al. [12] proposed a U-net encoder–decoder deep network called the progressive feature fusion network (PFF-Net) using progressive feature fusion. It completed the restoration of a hazy image by learning a highly non-linear transformation function from the observed hazy image to the haze-free ground-truth. Yang et al. [13] established a two-stage and end-to-end network, and we name it as Wavelet U-net (W-U-Net), which contained two characteristics. On the one hand, discrete wavelet transform and discrete wavelet inverse transform were used to replace the downsampling and upsampling operations, respectively, in the network. On the other hand, a chromatic adaptation transform is employed to adjust the luminance and color, which makes the generated image closer to the ground-truth. Qu et al. [14] proposed an enhanced pix2pix dehazing network (EPDN) , which was independent of the atmospheric scattering model. A GAN is embedded in the network to transform the image dehazing problem into an image-to-image conversion problem. The original purpose of the above-mentioned methods is to dehaze the ground images. For the hazy RS images, Li et al. [15] proposed a first-coarse-then-fine two-stage dehazing network named FCTF-Net. In the first stage, the encoder–decoder architecture is used to extract multi-scale features. The second stage is used to refine the results of the first stage.

In general, due to the difference in the imaging mechanisms of ground images and RS images, there are still many unsolved problems in RS images dehazing research, especially the more realistic problem that NHRS images dehazing is less involved. This paper has conducted special research on this issue. The deep neural network can extract both shallow features and deep spatial semantic information at the same time. This advantage of the deep neural network is effectively utilized in this paper. The neural network module is used to replace the traditional mathematical model with a large amount of calculation and optimization estimation to build an NHRS image dehazing network.

3. Proposed Method

In this section, we introduce our proposed method in three sub-sections, namely, the network architecture, the loss function design, and the non-uniform haze-adding algorithm.

3.1. Network Architecture

To use the powerful image representation capabilities of deep learning to improve the dehazing effect of RS images, an image dehazing network based on the encoder–decoder architecture is proposed. In the following section, the proposed RS image dehazing network is introduced in detail. As shown in Figure 1, the network is divided into three parts: an encoder, a decoder, and a wavelet channel.
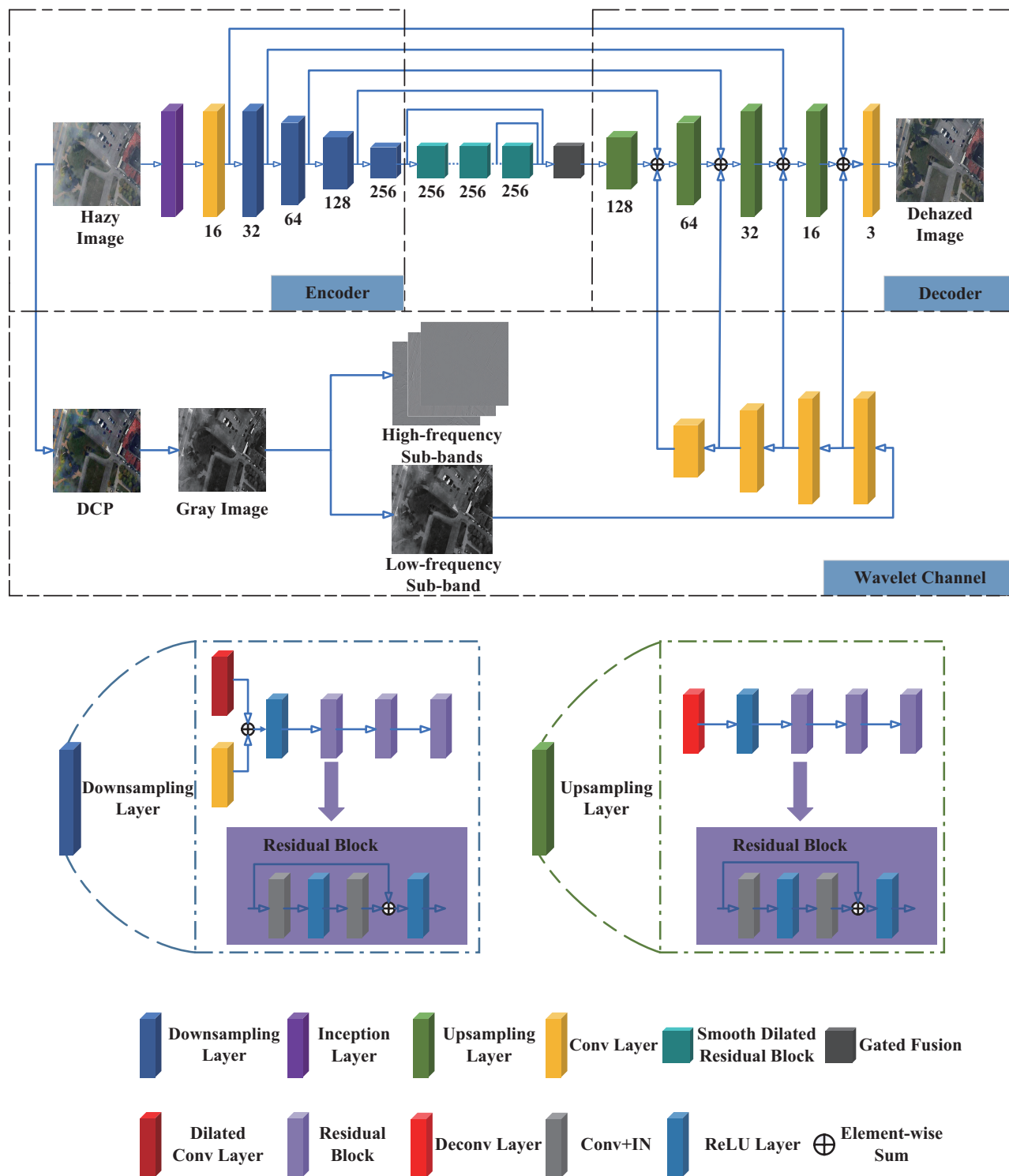
Figure 1. The architecture of the proposed network. Our network mainly consists of three parts (encoder, decoder, and wavelet channel). The encoder and decoder are connected by smooth dilated residual blocks and a gated fusion sub-network.

Encoder: The encoder consists of an inception module, a convolutional layer, and four downsampling layers. The scene scales in the RS images vary greatly, which means that there are big scene buildings and small scene cars. During the initial input phase of the network model, a structure similar to the inception [16] is introduced, namely using different scales of convolution layers to extract scene features of different scales in order to achieve the effect of multi-scale information extraction. It is very beneficial to the special image feature structures of RS images. In addition, the inception structure can reduce the number of parameters while increasing the depth and width of our network, which can better deal with the phenomenon of non-uniform distribution of haze in RS images. In the downsampling layer, a hybrid convolution composed of a standard convolution and a dilated convolution [17] is employed to extract the shallow features of the image more effectively, expand the receptive field to improve the ability of detecting the non-uniform haze in RS images, and make the dehazing results of the NHRS images more clear and natural. The size of convolution kernel for the standard convolution and dilated convolution is 3, and the dilation rate of dilated convolution is 2. After the hybrid convolution, three residual blocks [18] are used to further extract the image features. Each residual block is composed of two 3×3 convolutional layers, two instance normalization layers [19], and two rectified linear units (ReLU) functions. The residual block deepens the network structure. It can effectively extract features, strengthen feature propagation ability, prevent a vanishing gradient caused by the very deep network structure, generate natural and real images more easily, and improve the dehazing results of the NHRS images. The detailed encoder structure is given in Tabel 1. In this paper, the inception module also includes the pooling layer, but for the sake of brevity, the pooling layer is omitted in the figure and table.

Table 1. Details of the encoder. It includes the local structure, output size, and output channel of each layer in the encoder.

Layer	Local Structure	Output Size	Output Channel
Input		320×320	3
Inception	$(1 \times 1$ Conv$) \times 5$, $(3 \times 3$ Conv$) \times 3$	320×320	3
Convolution	11×11 Conv	320×320	16
Downsampling Layer1	$(3 \times 3$ Conv, s = 2, d = 1$)$, $(3 \times 3$ Conv, s = 2, d = 2$)$ ReLU $((3 \times 3$ Conv, IN, ReLU$) \times 2) \times 3$	160×160 160×160 160×160	32 32 32
Downsampling Layer2	$(3 \times 3$ Conv, s = 2, d = 1$)$, $(3 \times 3$ Conv, s = 2, d = 2$)$ ReLU $((3 \times 3$ Conv, IN, ReLU$) \times 2) \times 3$	80×80 80×80 80×80	64 64 64
Downsampling Layer3	$(3 \times 3$ Conv, s = 2, d = 1$)$, $(3 \times 3$ Conv, s = 2, d = 2$)$ ReLU $((3 \times 3$ Conv, IN, ReLU$) \times 2) \times 3$	40×40 40×40 40×40	128 128 128
Downsampling Layer4	$(3 \times 3$ Conv, s = 2, d = 1$)$, $(3 \times 3$ Conv, s = 2, d = 2$)$ ReLU $((3 \times 3$ Conv, IN, ReLU$) \times 2) \times 3$	20×20 20×20 20×20	256 256 256

Decoder: The decoder consists of four upsampling layers and a convolutional layer. The upsampling layer is composed of a deconvolution layer and three residual blocks. The deconvolution kernel size is 3 and the stride is 2. The structure of the residual block is the same as that of the residual block in the encoder. In order to obtain more high-resolution information and better recover the details in the original image, extra information is added at the skip connection. This image information is derived from the first-order low-frequency sub-band information of the 2D SWT after dark channel processing of the hazy RS image, which will be introduced in detail in the wavelet channel module. Between the encoder and decoder refers to the structure of the smoothed dilated convolution residual block and the gated fusion sub-network proposed by Chen et al. [20]. In this specific design, we use the

smoothed dilated convolution residual block and the gated fusion sub-network to connect the encoder and decoder, and set the number of channels of the residual block to 256 to meet the propagation requirements of the network. Compared with the standard residual block, the smoothed dilated convolution [21] residual block has a larger receptive field and can effectively aggregate contextual semantic information. This makes the network more effective in processing a non-uniform hazy image. The kernel size of the smoothed dilated convolution is set to 3. A gated fusion sub-network is added after the smoothed dilated convolution residual blocks. The purpose is to fuse the feature information from different levels to further improve the results of NHRS image dehazing. The detailed decoder structure is given in Tabel 2.

In Tables 1 and 2, 'Conv' and 'Deconv' are short for 'Convolution' and 'Deconvolution', respectively. 'ReLU' and 'IN' are short for 'Rectified Liner Funtion' and 'Instance Normalization Layer', respectively. Furthermore,'s' and 'd' are short for 'stride' and 'dilation rate', respectively.

Table 2. Details of the decoder. It includes the local structure, output size, and output channel of each layer in the decoder.

Layer	Local Structure	Output Size	Output Channel
	3×3 Deconv, s = 2	40×40	128
Upsampling Layer1	ReLU	40×40	128
	$((3 \times 3$ Deconv, IN, ReLU$) \times 2) \times 3$	40×40	128
	3×3 Deconv, s = 2	80×80	64
Upsampling Layer2	ReLU	80×80	64
	$((3 \times 3$ Deconv, IN, ReLU$) \times 2) \times 3$	80×80	64
	3×3 Deconv, s = 2	160×160	32
Upsampling Layer3	ReLU	160×160	32
	$((3 \times 3$ Deconv, IN, ReLU$) \times 2) \times 3$	160×160	32
	3×3 Deconv, s = 2	320×320	16
Upsampling Layer4	ReLU	320×320	16
	$((3 \times 3$ Deconv, IN, ReLU$) \times 2) \times 3$	320×320	16
Convolution	11×11 Conv	320×320	3
Dehazed output		320×320	3

Wavelet Channel: The 2D SWT can decompose an image into multiple sub-bands, including one low-frequency sub-band and several high-frequency sub-bands [22]. The low-frequency sub-band has a significant impact on the objective quality of the image, while the high-frequency sub-band has a significant impact on the perceived quality. In the dehazing task of an NHRS image, the low-frequency sub-band is selected as an additional image feature for the proposed dehazing network. This low-frequency information can provide more image features to improve the objective quality of the image. Compared with features of a heavy hazy image, the features of a thin hazy image can improve the image dehazing effect more obviously, and improve the objective quality and visual effect of the image more effectively. Therefore, the traditional dehazing method based on the atmospheric scattering model is first used for the initial image processing. The improved DCP method slightly changes in extracting global atmospheric light value *A*. It extracts the *A* values of the RGB three channels, respectively, and performs the subsequent calculations of the haze-free image restoration. After the initial processing, the image is converted to a gray one. Then, 2D SWT decomposition is performed to extract the information of the first-order low-frequency sub-band, which acts as an input into the network as the fourth channel and is downsampled with a standard convolution with the stride of 2 and the kernel size of 3. Before downsampling, we use standard convolution with the kernel of 1 to increase the dimension here. Then, the feature fusion is performed at the upsampled skip connection as the enhanced feature information of the image to further strengthen

the learning ability of deep network for low-frequency smooth information in RS images, thereby improving the non-uniform images dehazing results of the network.

3.2. Loss Function Design

Because of the particularity of NHRS images, a special design is made on the loss function, which plays a key role in the deep learning network training. This special design uses a combination of an L_1 loss, an MS-SSIM loss, and a Sobel edge detection loss.

L_1 Loss: The difference in pixels between the dehazed image $\hat{J}$ and the ground-truth J is calculated:

$$L_1 = ||J - \hat{J}||_1 \tag{2}$$

MS-SSIM Loss: The structural similarity index (SSIM) [23] is an index to measure the similarity of two images. It considers the luminance, contrast, and structure indicators, which is more suitable with the human visual experience. The sensitivity of the human visual system to noise depends on the local luminance, contrast, and structure. We use MS-SSIM [24], an SSIM loss function based on multiple layers (e.g., 5 layers) [25]:

$$L_{MS\text{-}SSIM} = 1 - MS\text{-}SSIM(J, \hat{J}) \tag{3}$$

Sobel Edge Loss: The Sobel operator is usually used to obtain the first-level gradients of an image. It further detect the edge by weighting the gray value difference of the up, down, left, and right areas of each pixel in the image. Therefore, the Sobel operator can be used to extract the edges of one image and suppress noise smoothly:

$$L_{Sobel} = ||E_{Sobel}(\hat{J}) - E_{Sobel}(J)||_1 \tag{4}$$

where E_{Sobel} denotes Sobel edge detection.

Total Loss: Based on the aforementioned loss functions, the total loss for training the proposed network is defined as:

$$L = \lambda_1 L_1 + \lambda_2 L_{MS\text{-}SSIM} + \lambda_3 L_{Sobel} \tag{5}$$

where λ_1, λ_2, and λ_3 are used to adjust the weights between three loss components.
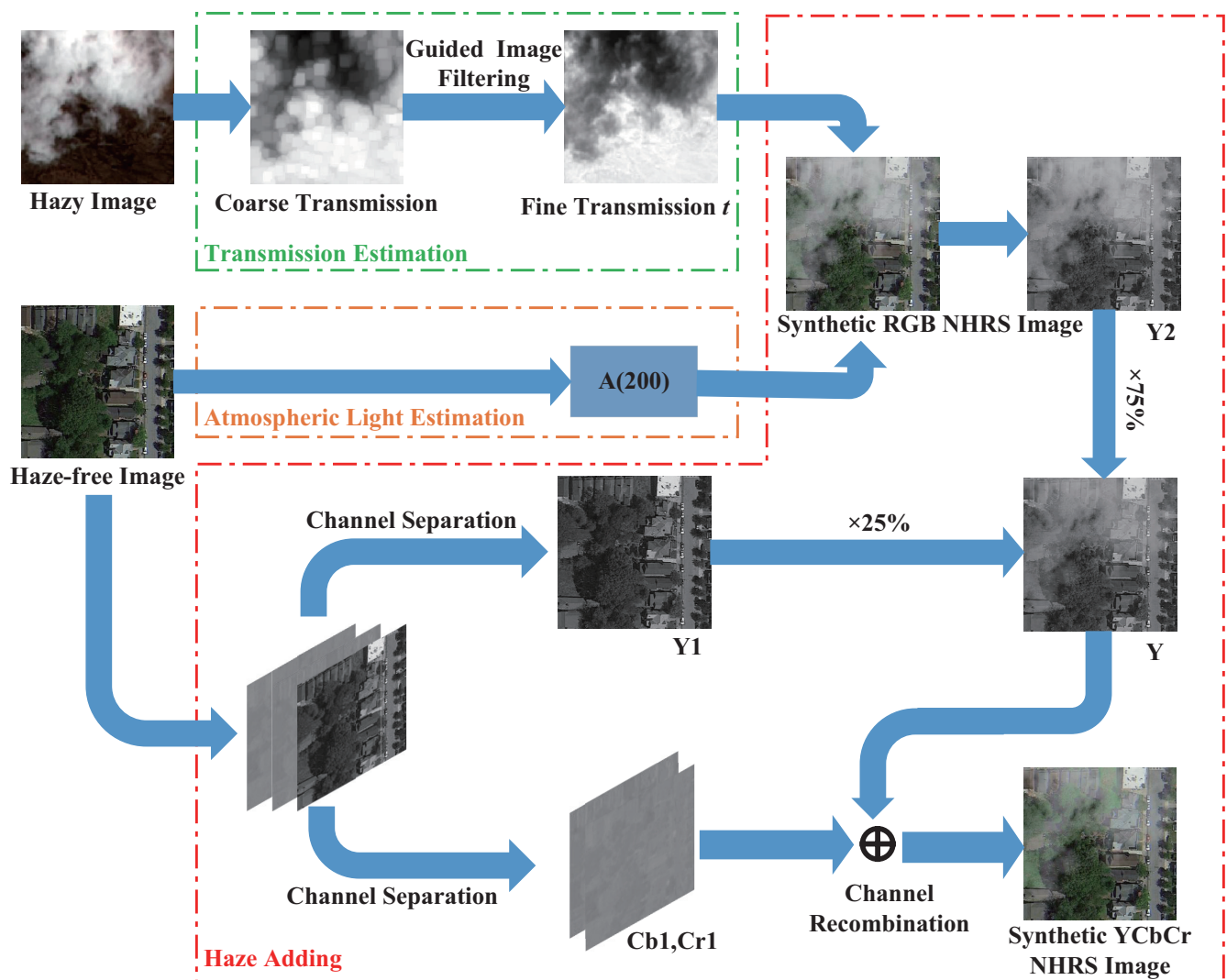
3.3. Non-Uniform Haze Adding Algorithm

To train the dehazing network, both haze and haze-free images are essential. However, in practice, it is difficult to obtain the matching data directly, especially for the NHRS images. The haze of real hazy images is used in the process of non-uniform haze adding, so that the effect of non-uniform haze adding is close to real hazy images. The haze itself is not purely low-frequency information or high-frequency information. Based on this characteristic, it is difficult to extract the haze from an image using the time domain or frequency domain filtering. In most dehazing research based on deep learning, the atmospheric scattering model is mostly used to directly add haze to obtain the training hazy images, and many of these methods directly extract both the transmission $t(x)$ and the global atmospheric light value A randomly. In most cases, there is a great gap between the synthetic hazy images and the real hazy images. This is also the most difficult problem in the research of deep learning dehazing based on training data.

We have conducted more in-depth research on NHRS image training data. The proposed non-uniform haze-adding algorithm mainly includes three parts: transmission estimation, atmospheric light estimation, and haze adding. As for transmission estimation, the coarse transmission is extracted by the dark channel prior and minimum filtering method in real hazy images. The block effect and halo effect [26] in the coarse transmission will affect the effect of the non-uniform haze adding. Thus, the resulting coarse transmission requires an additional step of refinement. Guided image filtering [26] is used to refine the coarse transmission and obtain the fine transmission t. For atmospheric light estimation, we use an improved DCP method to estimate the global atmospheric light value. It is

worth noting that the global atmospheric light value of the haze-free image is used in the non-uniform haze-adding algorithm. This method makes synthetic images close to real hazy images. With regard to haze adding, the haze is found to distribute mainly in the Y channel component in the YCbCr color space, according to the hazy distribution characteristic proposed by Dudhane et al. [27] in the process of the ground image dehazing. Through a broad range of experiments, we have found that the NHRS images also have the same characteristic. Therefore, the channel separation and recombination in the YCbCr color space are applied to the non-uniform haze-adding algorithm. First, the synthetic RGB NHRS images are obtained by using the atmospheric scattering model. The parameters $t(x)$ and A come from the two parts mentioned above. Then, the Y channel components (Y2) of the synthetic RGB NHRS images and the Y channel components (Y1) of haze-free images are fused into a new one Y. Finally, the Y is used to generate the final synthetic YCbCr NHRS images that are close to the real hazy images and retain the haze-free images' details. The flow chart of thw non-uniform haze-adding algorithm is shown in Figure 2. The experimental results of the non-uniform haze-adding algorithm are shown in Section 4.

Figure 2. Flow chart of non-uniform haze-adding algorithm. Non-uniform haze-adding algorithm consists of three parts (transmission estimation, atmospheric light estimation, and haze adding).

4. Experiment and Discussion

This section is divided into two sub-sections, namely, the experiment on non-uniform haze-adding algorithm and the experiment on proposed method on NHRS images de-

hazing, which is sufficiently compared with different typical dehazing methods from the perspective of quantitative and qualitative evaluations.

4.1. Experiment of Non-Uniform Haze-Adding Algorithm

Here we describe the experiment of non-uniform haze-adding algorithm in two parts, including the implementation details and dataset establishment.

4.1.1. Implementation Details

To obtain a better hazy image dataset, we have made precise settings for the parameters in the non-uniform haze-adding algorithm. As shown in Figure 2, in the part of transmission extraction, a minimum filtering is used to estimate the coarse transmission from the real hazy images. Then, the guided image filtering [26] is applied to refine the coarse transmission into a fine transmission t. The minimum filtering window size is three when synthesizing the training set and validation set, and four when synthesizing the test set. The filter radius of the guided image filtering is four times larger than the minimum filtering window size, and the adjustment parameter is 10^{-6}. In the part of atmospheric light extraction, the parameter $A0$ is set to the maximum global atmospheric light value that is equal to 200. If the calculated value $A1$ is greater than or equal to $A0$, the global atmospheric light value A equals $A0$. Moreover, we conduct massive experiments on haze adding. In conclusion, 75% of the Y channel (Y2) of the synthetic RGB NHRS images and 25% of the Y channel (Y1) of haze-free images are selected and fused into a new one, Y. The final synthetic YCbCr NHRS images are composed of Y, newly obtained, and Cb1, Cr1 of the haze-free images.

Lots of experiments have been carried out according to the non-uniform haze-adding algorithm mentioned in Figure 2. Some representative experimental results of the non-uniform haze-adding algorithm are shown in Figure 3.

As shown in Figure 3, the proposed non-uniform haze-adding algorithm retains some detailed information of haze-free images and alleviates obscuring the textures caused by thick haze. In addition, the proposed algorithm only adds haze on the Y channel component, thus, the synthesized images are close to the real NHRS images.

4.1.2. Dataset Establishment

To meet the training requirements of the dehazing network, the dataset establishment is essential. The training dataset used in this paper is established according to the proposed non-uniform haze-adding algorithm described in Section 3.3. We utilize three typical RS haze-free images dataset: AID30 [28], RSSCN7 [29], and BH [30] dataset as basic images.

The AID30 dataset is a large aerial RS images dataset released by Wuhan University in 2017. It has a spatial resolution of 0.5-0.8 meters and the size of 600×600. 660 haze-free images are selected for synthesizing the non-uniform haze-adding training set. The RSSCN7 dataset is a 400×400 aerial RS image dataset published by Qin Zou [29] of Wuhan University in 2015. From RSSCN7 dataset, 164 haze-free images are selected for synthesizing the non-uniform haze-adding validation set. The BH Dataset is a 3840×2160 aerial RS image dataset released by the Federal University of Minas Gerais in 2020. From BH dataset, 36 haze-free images are randomly cropped for synthesizing the non-uniform haze-adding validation set. In addition, 22 haze-free images from the AID30 dataset, and 18 haze-free images from the Internet are selected for synthesizing the non-uniform haze-adding test set. The 22 haze-free images are selected from the AID30 dataset, where we ensure that these images are not included in the 660 haze-free images from the AID30 dataset.
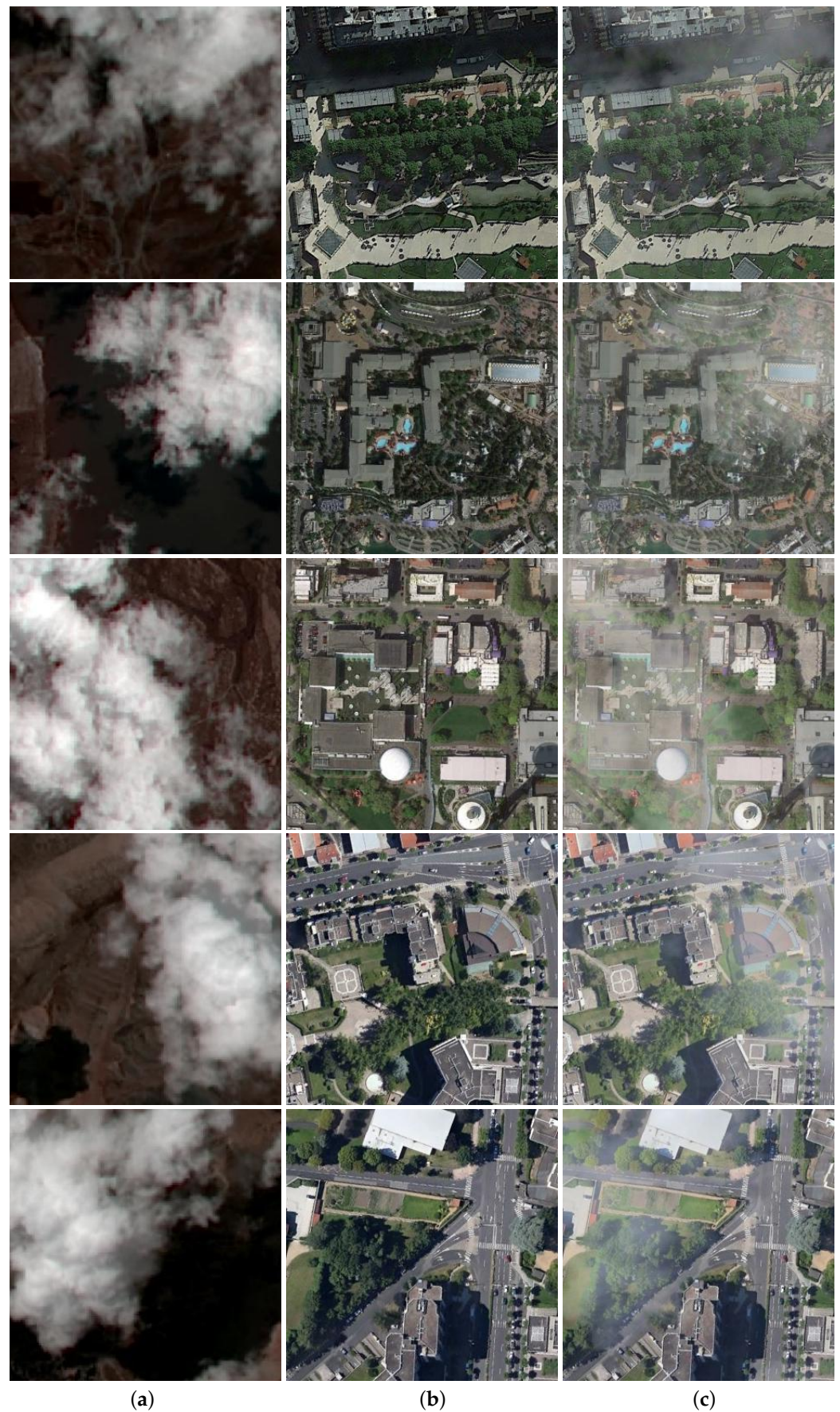
Figure 3. Experimental results of non-uniform haze-adding algorithm. (**a**) Real NHRS image. (**b**) Haze-free RS image. (**c**) Synthetic NHRS image.

Similar to the selection of haze-free images, the non-uniform haze extraction images used in the haze-adding algorithm are used to be different to ensure the synthetic training set, validation set, and test set are different. The imaging mechanism of satellite NHRS and aerial NHRS is basically the same, and the former is easier to obtain in the public dataset. Therefore, the satellite NHRS images are used to replace the aerial NHRS images as the non-uniform haze extraction images. In this paper, 82 non-uniform haze extraction images are randomly cropped from three satellite NHRS images, which are obtained from Landsat 8 OLI-TIRS satellite digital products on the website of Geospatial Data Cloud [31]. Among them, there are 40, 31, and 11 NHRS images for the training set, validation set, and test set, respectively.

Before adding the haze, we augment the training images by randomly selecting and flipping the real non-uniform hazy images and haze-free images followed by rotating one of 0, 90, 180, and 270 degrees. Then, the real non-uniform hazy images and the haze-free images are cropped to a fixed size, which is 320×320 for the training set and validation set, and 480×320 for the test set. To reduce the probability of identical synthetic NHRS images, the transmission is also randomly expanded by 0.7–0.9, such that the haze density changes together. Finally, we use 660 haze-free images from the AID30 dataset and 40 non-uniform haze extraction images to obtain 10,000 synthetic NHRS images as the training set. Likewise, the validation set contains 2000 synthetic NHRS images, which are obtained by 164 haze-free images from the RSSCN7 dataset, 36 haze-free images from the BH dataset, and 31 non-uniform haze extraction images. The test set contains 500 synthetic NHRS images, which are obtained by 22 haze-free images from the AID30 dataset, 18 haze-free images from the Internet, and 11 non-uniform haze extraction images.

4.2. Experiment of Proposed Dehazing Method

4.2.1. Training Details

In the proposed dehazing method, the PyTorch framework is used for the training and test, and the NVIDIA 2080Ti GPU training model in Ubuntu 18.04.1 LTS is used. The initial learning rate is set to 0.0001, and the Adam optimizer is used to accelerate the training. We train the network for 100 epochs with a batch size of 8. Through a large number of experiments, the hyper-parameters λ_1, λ_2, and λ_3 in the loss function of Equation (5) are selected as 0.075, 0.04, and 0.0002, respectively. In the experiment, we try our best to make fair comparisons using unified 100 epochs for different methods.

4.2.2. Result Evaluation

In this section, we compare our method with seven classical dehazing methods, including some deep learning methods, and evalute them overall. The comparative typical methods including DCP [6,26], NLD [8,32], CAP [9], PFF-Net [12], W-U-Net [13], EPDN [14], and FCTF-Net [15]. Here, the guided image filtering [26] is applied to DCP to further improve the estimation accuracy of transmission. For the sake of fairness, all of the methods based on deep learning adopt our established synthetic non-uniform haze dataset to train.

Qualitative Evaluation

To further qualitatively evaluate and compare the dehazing results of several different methods of RS images, Figures 4–7 show the several representative dehazing results of synthetic NHRS images and real NHRS images. Figures 4 and 5 show the dehazing results of synthetic NHRS images. The first row is synthetic NHRS images, the last row is ground-truth, and the middle rows are several different dehazing results of NHRS images.

It can be seen that DCP, CAP, and NLD all have excessive enhancement when dealing with NHRS images. When faced with haze with different intensities, the processing effect is poor, and the region with a higher density of haze cannot be effectively processed, which shows that most of the haze in the image can be removed based on the prior dehazing method, while the scene and target details can be generally recovered. However,

the dehazing results have obvious color distortion. There are obvious limitations when dealing with NHRS images. PFF-Net and W-U-Net based on deep learning are effective in processing non-uniform hazy regions in NHRS images, and there is no color distortion, but it has the obvious haze residual in thick haze regions in the images. EPDN has better results when processing NHRS images, and there will be no large amount of haze residue, but there is still a small area of haze residue. FCTF-Net that is proposed for non-uniform RS images has a better effect in dealing with non-uniform haze regions, but there will still be a certain amount of haze residue. Compared with other comparative methods, our proposed method has the least haze residue and achieves results that are the closest to the ground-truth.
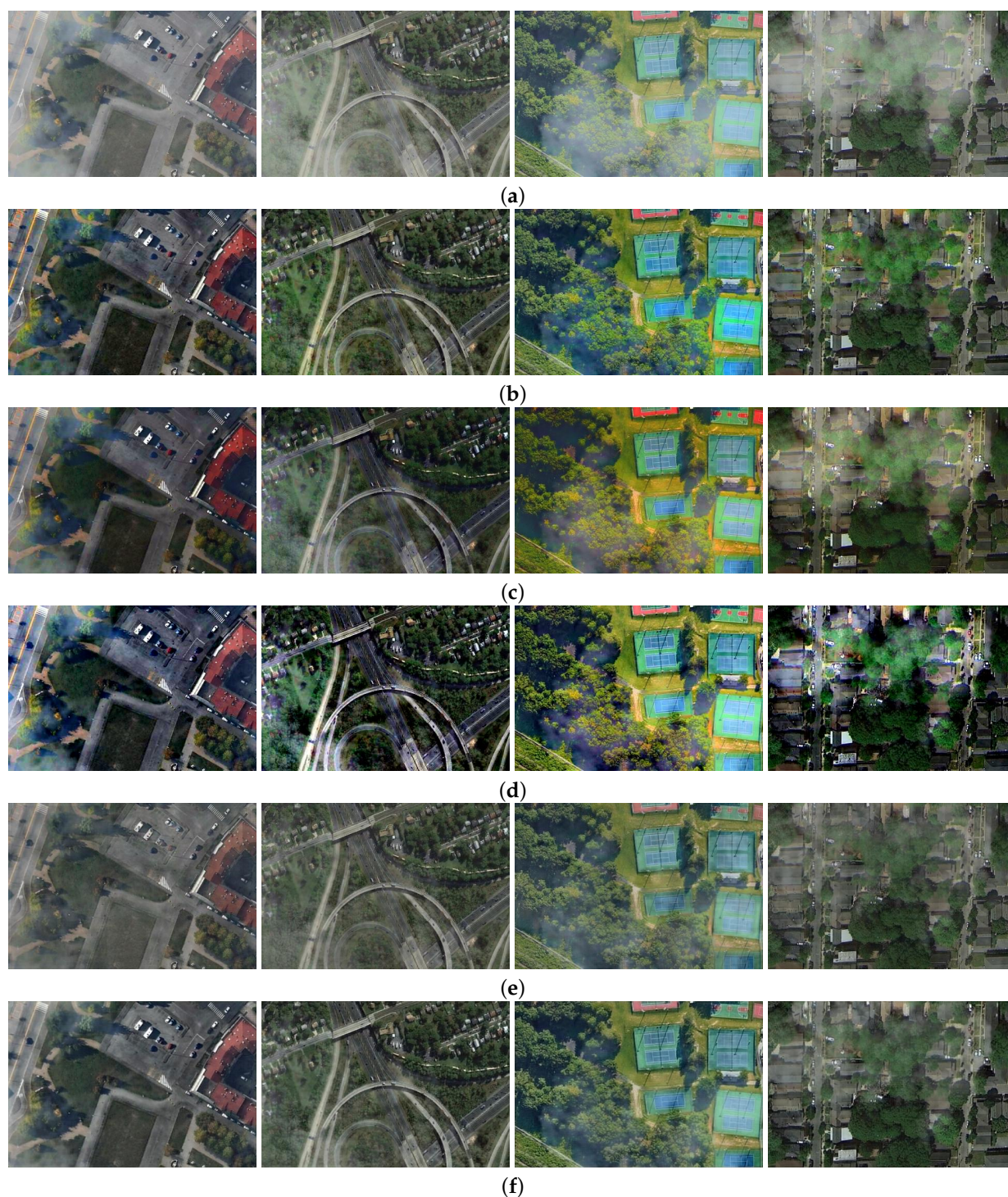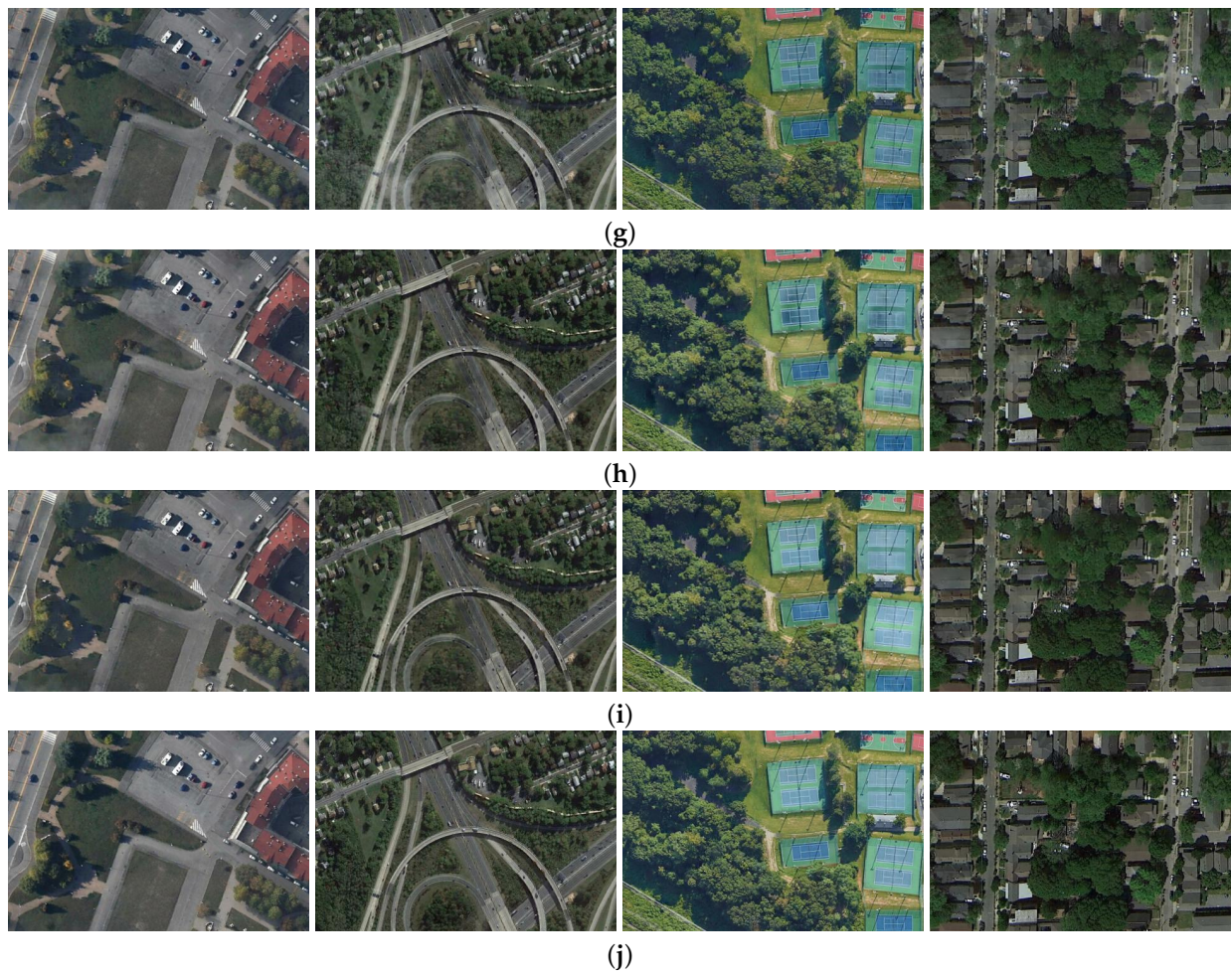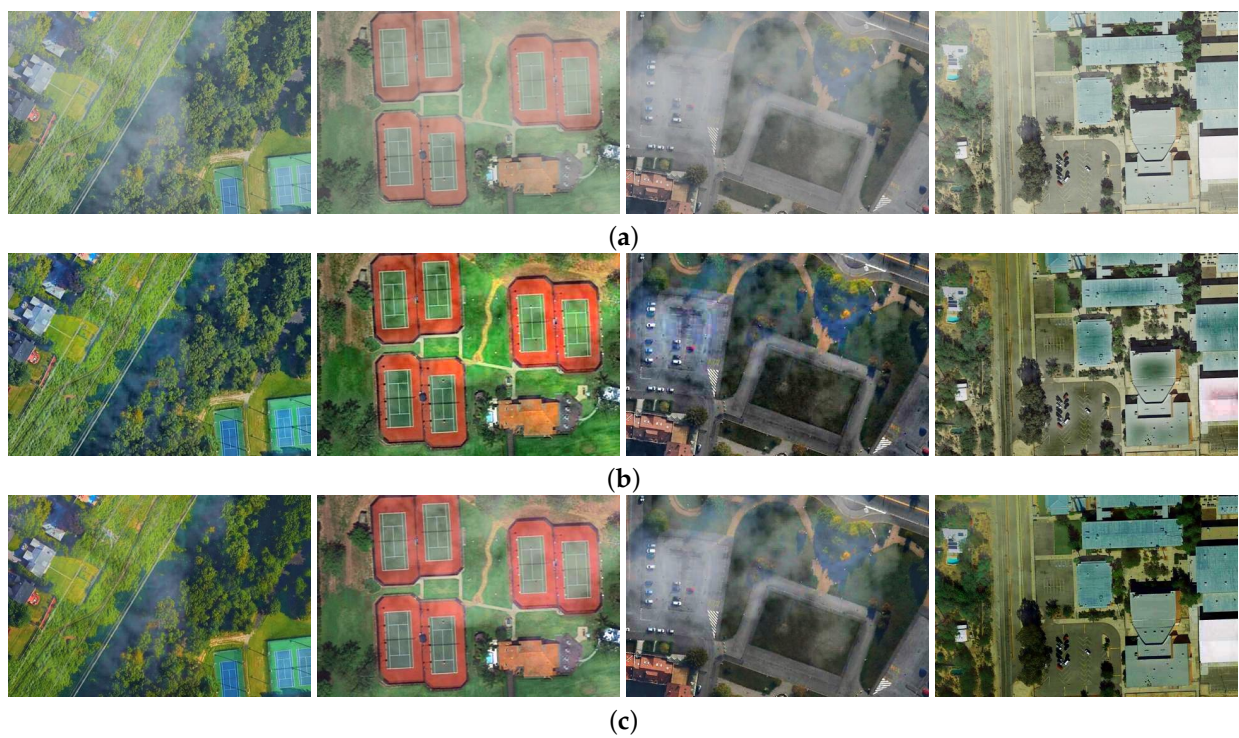

(a)

(b)

(c)

(d)

(e)

(f)

Figure 4. *Cont.*

(g)

(h)

(i)

(j)

Figure 4. Dehazing results I on synthetic images. (**a**) Synthetic NHRS images. (**b**) DCP. (**c**) CAP. (**d**) NLD. (**e**) PFF-Net. (**f**) W-U-Net. (**g**) EPDN. (**h**) FCTF-Net. (**i**) Ours. (**j**) Ground-truth.

(a)

(b)

(c)

Figure 5. *Cont.*

Figure 5. Dehazing results II on synthetic images. (**a**) Synthetic NHRS images. (**b**) DCP. (**c**) CAP. (**d**) NLD. (**e**) PFF-Net. (**f**) W-U-Net. (**g**) EPDN. (**h**) FCTF-Net. (**i**) Ours. (**j**) Ground-truth.

To better verify the effectiveness of our network to deal with NHRS images, we test the network on real NHRS images. Figures 6 and 7 show the comparison results of eight real NHRS images. DCP removes most of the haze in the image, but color distortion appears after the haze removal. NLD causes an over-brightness phenomenon, therefore, the result of this method is not natural. CAP also has some haze residue and color distortion. The dehazing results of PFF-Net and W-U-Net still leave a certain amount of haze, but the color distortion in the prior method is well solved. EPDN and FCTF-Net remove most haze, but they cannot deal with it effectively in the regions with thick haze. Our proposed method has achieved a good balance between image dehazing and keeping the natural color of image dehazing results. The scene of dehazed image by the proposed method is naturally clear. The proposed dehazing method is obviously superior to other contrast methods in processing NHRS images; the dehazed images are complete in information, rich in edge features and details, and the method has good dehazing results.

As discussed in previous works, the deep learning method proposed in this paper achieves a good compromise in terms of the degree of removal of NHRS images and the preservation of the naturalness of the restored images. For the thick haze region in the NHRS images, this method has the highest degree of image information restoration and moderate overall brightness. Generally speaking, the colors are more realistic and the scenery is more lifelike.

There exists only a subjective evaluation for the test of the real NHRS images according to no corresponding reference image without haze. Although the synthetic hazy dataset is used for training, Figures 6 and 7 show our method has also achieved satisfactory dehazing effects on real NHRS images.
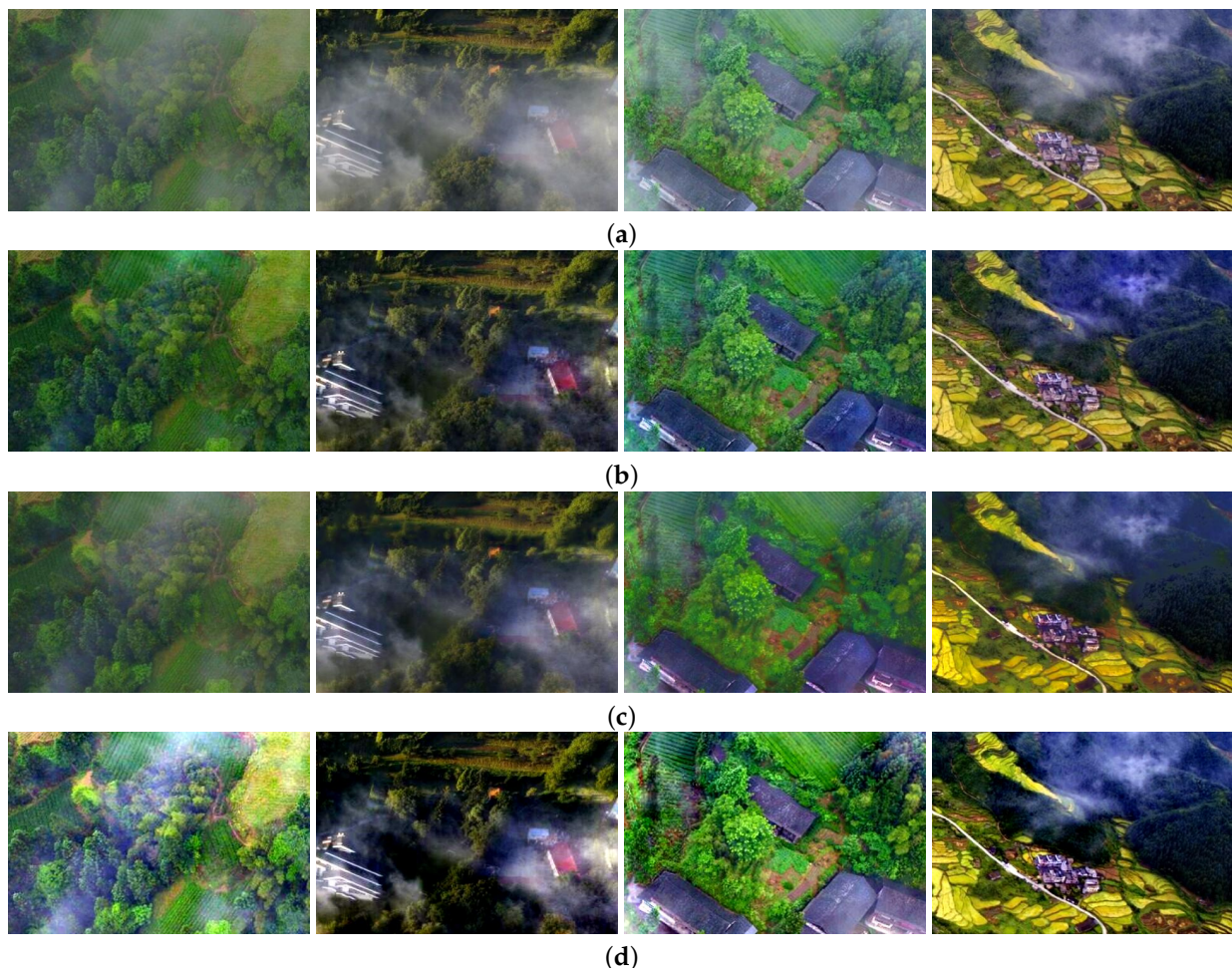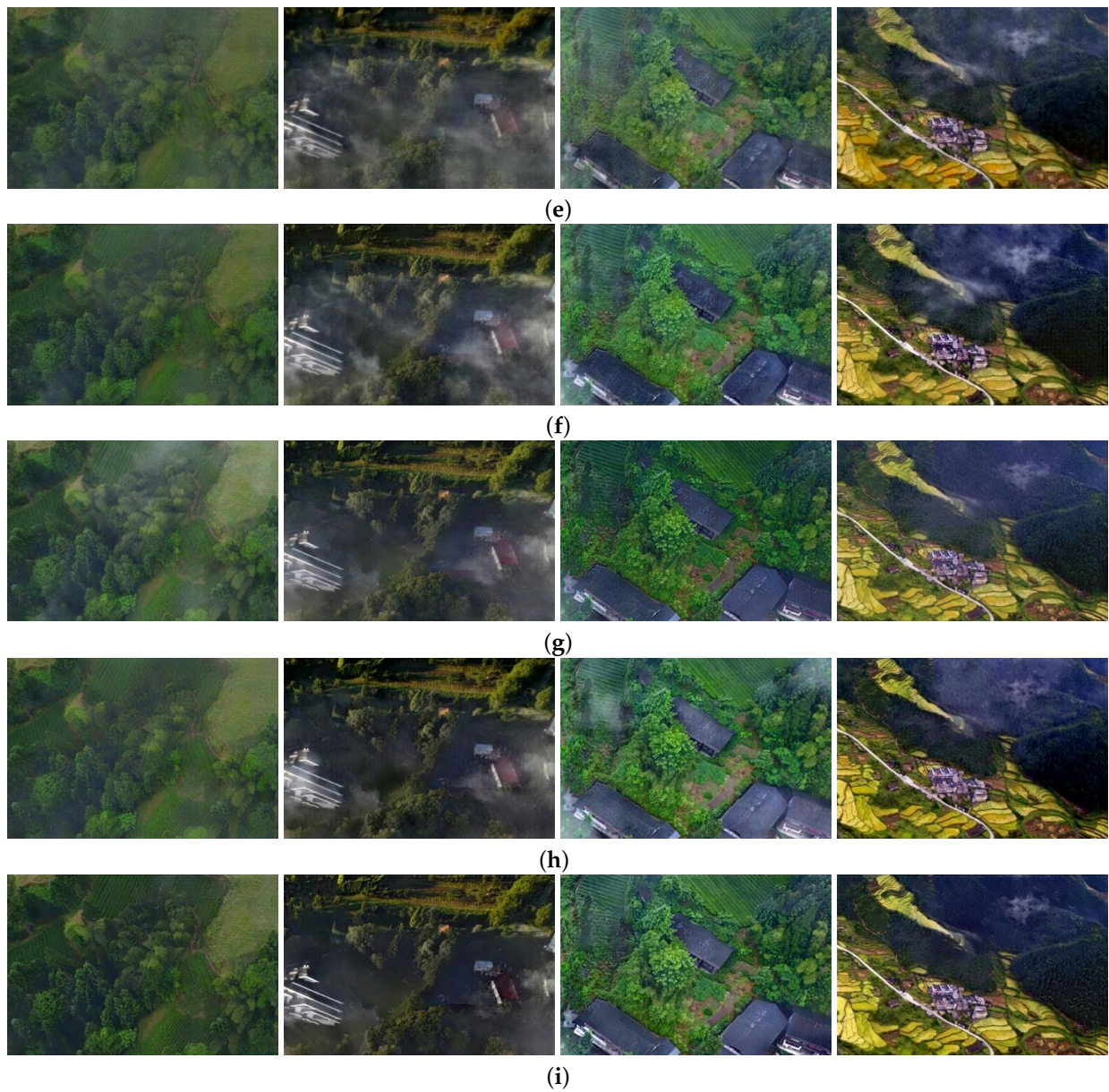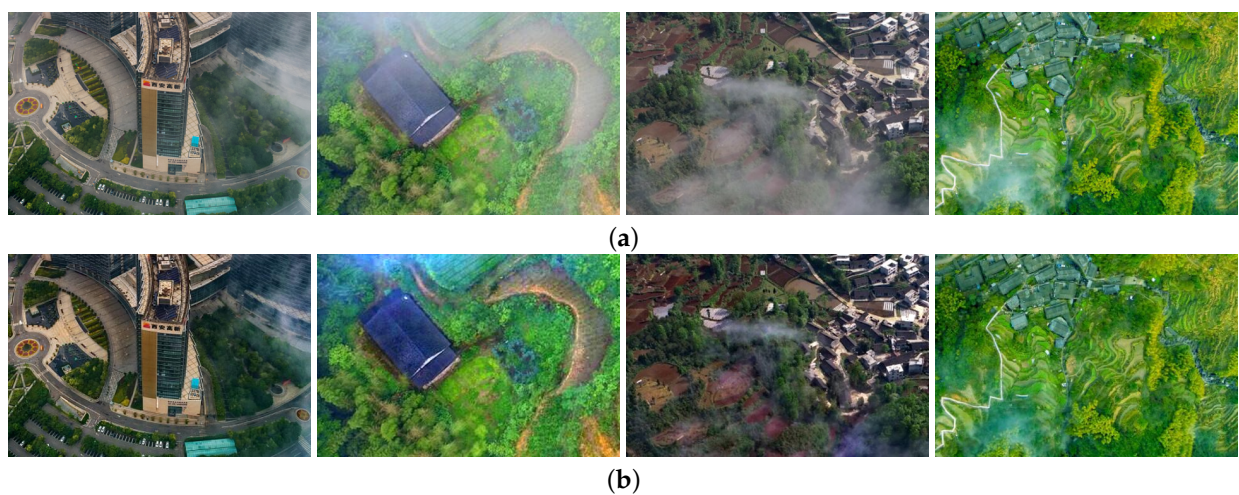

(a)

(b)

(c)

(d)

Figure 6. *Cont.*

Figure 6. Dehazing results I on real hazy images. (**a**) Real NHRS images. (**b**) DCP. (**c**) CAP. (**d**) NLD. (**e**) PFF-Net. (**f**) W-U-Net. (**g**) EPDN. (**h**) FCTF-Net. (**i**) Ours.

Figure 7. *Cont.*

Figure 7. Dehazing results II on real hazy images. (**a**) Real NHRS images. (**b**) DCP. (**c**) CAP. (**d**) NLD. (**e**) PFF-Net. (**f**) W-U-Net. (**g**) EPDN. (**h**) FCTF-Net. (**i**) Ours.

Quantitative Evaluation

To quantitatively evaluate the dehazing effects of different methods, we use two commonly used evaluation indicators in the dehazing task on the synthetic dataset, the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). PSNR is based on the error between the corresponding pixels, which is an error-sensitive image quality evaluation index. A higher PSNR value means higher similarity between the dehazing results and the ground-truth, and smaller image distortion. SSIM measures image similarity from three aspects of luminance, contrast, and structure. A higher SSIM value means that the results after dehazing are closer to the ground-truth.

Tables 3 and 4 show the PSNR and SSIM values of the eight synthetic NHRS images in Figures 4 and 5. In these eight images, the PSNR and SSIM values of the proposed method exceed the other seven methods. In comparison with the FCTF-Net for NHRS images, the proposed method also achieved better dehazing results, which proves that the proposed method has certain advantages in processing NHRS images.

It can be seen from Tables 3 and 4 that for the seven comparison methods, the PSNR and SSIM values obtained by DCP, CAP, and NLD are lower, indicating that although they are very successful image dehazing methods, there are limitations on the restoration of NHRS images, which is consistent with the subjective effects in Figures 4 and 5, and there is a certain gap compared with the deep learning method. PFF-Net, W-U-Net, EPDN, and FCTF-Net obtain intermediate PSNR values and SSIM values. It can also be seen from Figures 4 and 5 that these four methods have a better dehazing performance, which is better than the three dehazing methods based on the atmospheric scattering model, but slightly worse than the method proposed by us. In a word, from the PSNR and SSIM values in Tables 3 and 4 and the corresponding dehazing results in Figures 4 and 5, it validates that the proposed method has superiority in image dehazing compared with the seven methods.

Table 3. Quantitative evaluation results of PSNR for 8 images in Figures 4 and 5. Bold index indicates the best performance.

Image	DCP	CAP	NLD	PFF-Net	W-U-Net	EPDN	FCTF-Net	Ours
Figure 4 1st column	18.1757	16.8021	16.0314	22.1598	20.6270	25.1673	26.6991	**26.9116**
Figure 4 2nd column	23.0893	18.9500	16.8124	21.4356	21.7374	23.7852	27.1162	**28.0056**
Figure 4 3rd column	20.6492	17.9509	18.6025	21.2522	21.9921	23.7839	26.3346	**28.1458**
Figure 4 4th column	20.6405	16.2992	16.1343	19.4756	19.7069	25.5164	27.9014	**28.6765**
Figure 5 1st column	22.0578	19.9599	17.0074	22.5066	23.3277	23.6046	24.4548	**27.4342**
Figure 5 2nd column	17.7801	19.9324	16.7586	23.2878	22.2942	22.7151	27.2257	**30.3904**
Figure 5 3rd column	18.4220	17.4235	14.4382	21.3188	20.9592	27.1209	25.0667	**28.2225**
Figure 5 4th column	16.9078	13.5731	14.8070	21.7622	22.5752	23.5890	25.3069	**28.4855**

Table 4. Quantitative evaluation results of SSIM for 8 images in Figures 4 and 5. Bold index indicates the best performance.

Image	DCP	CAP	NLD	PFF-Net	W-U-Net	EPDN	FCTF-Net	Ours
Figure 4 1st column	0.7791	0.7836	0.6383	0.8516	0.8706	0.9163	0.9229	**0.9261**
Figure 4 2nd column	0.8519	0.7744	0.6459	0.8315	0.8363	0.8743	0.8949	**0.8991**
Figure 4 3rd column	0.8570	0.7286	0.8036	0.8603	0.8874	0.9325	0.9415	**0.9443**
Figure 4 4th column	0.8231	0.7204	0.6052	0.7797	0.8074	0.8718	0.8984	**0.9001**
Figure 5 1st column	0.8646	0.8203	0.8051	0.8548	0.8797	0.9127	0.9188	**0.9285**
Figure 5 2nd column	0.7258	0.8717	0.7364	0.9072	0.9261	0.9346	0.9539	**0.9617**
Figure 5 3rd column	0.7220	0.8033	0.5719	0.8350	0.8550	0.9006	0.9014	**0.9153**
Figure 5 4th column	0.8431	0.6970	0.6738	0.8860	0.8900	0.9264	0.9306	**0.9419**

To quantitatively evaluate the dehazing effects of different dehazing methods, we tested the average PSNR value and SSIM value of 500 test images, and also tested the average feature similarity index (FSIM) value [33]. FSIM is a novel low-level feature-based

image quality assessment (IQA) metric, and FSIM can achieve much higher consistency with subjective evaluations than state-of-the-art IQA metrics. A higher FSIM value denotes that the results after dehazing are closer to the ground-truth. The PSNR, SSIM, and FSIM results are listed in Table 5.

Table 5. Quantitative evaluation results obtained with the synthetic NHRS images (average of 500 images). Bold index indicates the best performance.

Metrics	DCP	CAP	NLD	PFF-Net	W-U-Net	EPDN	FCTF-Net	Ours
PSNR	20.7352	17.5927	17.4697	22.8827	23.2384	24.6138	26.9153	**27.9154**
SSIM	0.8633	0.7734	0.7505	0.8812	0.8966	0.9178	0.9342	**0.9369**
FSIM	0.9384	0.8950	0.8838	0.9166	0.9325	0.9502	0.9513	**0.9529**

The results in Table 5 show that our method obtains certain advantages in comparison with the results of the seven comparison methods. It has better NHRS images dehazing performance, with a PSNR value of 27.9154 dB, SSIM value of 0.9369, and FSIM value of 0.9529. The PSNR value, SSIM value, and FSIM value obtained by our method are all higher than the other seven comparison methods. Among the seven comparison methods, FCTF-Net obtains the second highest PSNR value, SSIM value, and FSIM value, while the PSNR value, SSIM value, and FSIM value obtained by DCP, CAP, and NLD are lower, and the dehazing performance is poor. PFF-Net, W-U-Net, and EPDN obtain intermediate PSNR, SSIM, and FSIM values. In order to see the advantages of our method more intuitively, we display the experimental results in the form of bar charts in Figure 8. From Figure 8, it can also be seen that our method achieves better performance in the PSNR value, SSIM value, and FSIM value. Compared with the dehazing methods based on the atmospheric scattering model, it has been greatly improved, and it also has some advantages compared with the dehazing methods based on deep learning. In short, the experimental results on the synthetic dataset validate that our proposed method achieves the best experimental results.

Under the qualitative and quantitative verification of a large number of experiments, the proposed method can effectively compensate for the image contrast reduction caused by haze, reconstruct the haze-free image with bright colors, and improve the visibility of RS images. The haze residue in our method is less than other methods, and there is no obvious color cast.
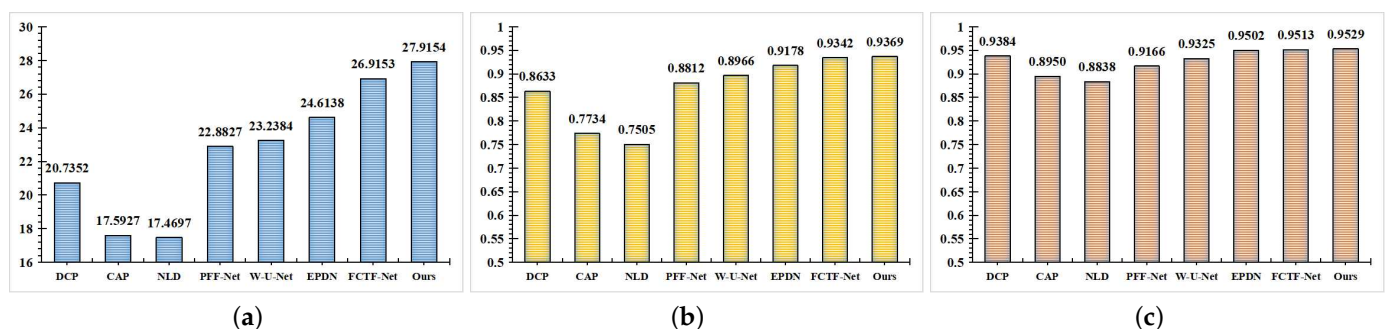
Figure 8. Visual display of the quantitative evaluation results in Table 5. (**a**) PSNR. (**b**) SSIM. (**c**) FSIM.

4.2.3. Discussion

In this paper, we propose a new advanced non-uniform dehazing network, which combines deep learning with the atmospheric scattering model. Figures 4 and 5 illustrate the results of seven typical dehazing methods (b–h) and our method (i) for dehazing synthetic non-uniform haze images. It can be seen from the comparison that those dehazing methods (b–d) based on the atmospheric scattering model have certain limitations to the task of non-uniform image dehazing. After dehazing, those images have an obvious color

distortion or excessive enhancement, especially the third column in Figure 4c and the second columns in Figure 5b,d. In dehazing methods (e–i) based on deep learning, all of EPDN, FCTF-Net, and our method show the better dehazing capability for synthetic non-uniform haze images. From the second columns in Figures 4 and 5, we can see that there is a slighter haze remaining in some regions of the dehazing result of our method over that of EPDN and FCTF-Net. It is proved that our method performs better than the other methods. Not only does our method achieve the ideal performance in a qualitative analysis, but it also has obvious competitiveness in a quantitative analysis. The average value of the quantitative evaluation of eight dehazing methods are shown in Table 5. These values are more clearly displayed in Figure 8 with bar charts, and it verifies that our method has achieved a relatively ideal performance on the three evaluation indexes. Compared with the best FCTF-Net among the seven comparison methods, our method improves the PSNR value by 1.0001 dB, SSIM value by 0.0027, and FSIM value by 0.0016.

The results of eight dehazing methods for real non-uniform haze images are shown in Figures 6 and 7, respectively. The proposed method has obvious advantages over seven typical dehazing methods in the dehazing work of real NHRS images. However, there are still some limitations to our method. As shown in Figure 7, in our dehazing result, there exists some residual haze in some regions of the images. Unlike the light with different wavelengths which travels through the thin haze region in the imaging process, when faced with a thick haze region, the scattering coefficients are no longer identical or nearly identical. Hence, the traditional methods usually encounter difficulties in solving the atmospheric scattering model, while the methods based on deep learning are difficult to construct an effective dataset. As a result, there is still a certain degree of haze residue in the thick haze region, and our future research will emphasize the thick haze removal in real NHRS images.

5. Conclusions

In this paper, we propose a dehazing method for RS images with non-uniform haze, which combines the wavelet transform and deep learning technology. Aiming at the non-uniform distribution of haze in RS images, we introduce the low-frequency sub-band of the 2D SWT for one-level decomposition into the network to further strengthen the learning ability of the deep network for low-frequency smooth information in RS images. The inception structure and the hybrid convolution combining the standard convolution with dilated convolution are embedded into the encoder module of the network. As a result, the network can learn more abundant image features and improve the overall detection capability for non-uniform haze in RS images.

The NHRS dataset is established through our proposed haze-adding algorithm. The experimental results on the established dataset demonstrated that the proposed method achieves better performance than the other typical dehazing methods based on the atmospheric scattering model and deep learning in the case of non-uniform haze distribution. However, due to a certain gap between the synthetic hazy dataset and the real hazy dataset, it is difficult to complete the precise design of the network model to achieve the effect of complete dehazing. When the proposed method is applied to real hazy remote sensing images, there will be a certain degree of haze residue in the thick hazy region in the images. In our future work, we will further exploit the relevant features of haze, improve our non-uniform haze-adding algorithm, and try to create an image dataset closer to the real NHRS images. Moreover, we also explore the deep learning network to better solve the problem of haze residue in real RS image dehazing.

Data Availability Statement: Te data of experimental images used to support the fndings of this study are available from the corresponding author upon the reasonable request.

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

Abbreviations

The following abbreviations are used in this manuscript:

RS	Remote sensing
NHRS	Non-uniform haze remote sensing
SWT	Stationary wavelet transform
PSNR	Peak signal-to-noise ratio
SSIM	Structural similarity
FSIM	Feature similarity

References

1. Wierzbicki, D.; Kedzierski, M.; Grochala, A. A Method for Dehazing Images Obtained from Low Altitudes during High-Pressure Fronts. *Remote Sens.* **2019**, *12*, 25. [CrossRef]
2. Zhu, Z.; Luo, Y.; Wei, H.; Li, Y.; Qi, G.; Mazur, N.; Li, Y.; Li, P. Atmospheric Light Estimation Based Remote Sensing Image Dehazing. *Remote Sens.* **2021**, *13*, 2432. [CrossRef]
3. Mccartney, E.J.; Hall, F.F. Optics of the Atmosphere: Scattering by Molecules and Particles. *Phys. Today* **1977**, *30*, 76–77. [CrossRef]
4. Narasimhan, S.G.; Nayar, S.K. Chromatic Framework for Vision in Bad Weather. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition—CVPR 2000, Hilton Head, SC, USA, 15 June 2000; pp. 598–605.
5. Narasimhan, S.G.; Nayar, S.K. Vision and the atmosphere. *Int. J. Comput. Vis.* **2002**, *48*, 233–254. [CrossRef]
6. He, K.; Sun, J.; Tang, X. Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353. [PubMed]
7. Fattal, R. Single Image Dehazing. *ACM Trans. Graph.* **2008**, *27*, 1–9. [CrossRef]
8. Berman, D.; Treibitz, T.; Avidan, S. Non-local Image Dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition—CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 1674–1682.
9. Zhu, Q.; Mai, J.; Shao, L. A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE Trans. Image Process.* **2015**, *24*, 3522–3533. [PubMed]
10. Muhammad, K.; Mustaqeem; Ullah, A.; Imran, A.S.; Sajjad, M.; Kiran, M.S.; Sannino, G.; de Albuquerque, V.H.C. Human Action Recognition Using Attention Based LSTM Network with Dilated CNN Features. *Future Gener. Comput. Syst.* **2021**, *125*, 820–830. [CrossRef]
11. Mustaqeem; Kwon, S. Optimal Feature Selection Based Speech Emotion Recognition Using Two-Stream Deep Convolutional Neural Network. *Int. J. Intell. Syst.* **2021**, *36*, 5116–5135. [CrossRef]
12. Mei, K.; Jiang, A.; Li, J.; Wang, M. Progressive Feature Fusion Network for Realistic Image Dehazing. In Proceedings of the Asian Conference on Computer Vision—ACCV 2018, Perth, Australia, 2–6 December 2018; pp. 203–215.
13. Yang, H.H.; Fu, Y. Wavelet U-Net and the Chromatic Adaptation Transform for Single Image Dehazing. In Proceedings of the IEEE International Conference on Image Processing—ICIP 2019, Taipei, Taiwan, 22–25 September 2019; pp. 2736–2740.
14. Qu, Y.; Chen, Y.; Huang, J.; Xie, Y. Enhanced Pix2pix Dehazing Network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition—CVPR 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 8152–8160.
15. Li, Y.; Chen, X. A Coarse-to-Fine Two-stage Attentive Network for Haze Removal of Remote Sensing Images. *IEEE Geosci. Remote Sens.* **2020**, *18*, 1751–1755. [CrossRef]
16. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition—CVPR 2015, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
17. Yu, F.; Koltun, V. Multi-scale Context Aggregation by Dilated Convolutions. In Proceedings of the International Conference on Learning Representations—ICLR 2016, San Juan, PR, USA, 2–4 May 2016; pp. 1–13.
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition—CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
19. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance Normalization: The Missing Ingredient for Fast Stylization. *arXiv* **2016**, arXiv:1607.08022.

20. Chen, D.; He, M.; Fan, Q.; Liao, J.; Zhang, L.; Hou, D.; Yuan, L.; Hua, G. Gated Context Aggregation Network for Image Dehazing and Deraining. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision—WACV 2019, Waikoloa, HI, USA, 7–11 January 2019; pp. 1375–1383.
21. Wang, Z.; Ji, S. Smoothed Dilated Convolutions for Improved Dense Prediction. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD 2018, London, UK, 19–23 August 2018; pp. 2486–2495.
22. Deng, X.; Yang, R.; Xu, M.; Dragotti, P.L. Wavelet Domain Style Transfer for an Effective Perception-Distortion Tradeoff in Single Image Super-Resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision—ICCV 2019, Seoul, Korea, 27 October–2 November 2019; pp. 3076–3085.
23. Zhou, W.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612.
24. Zhou, W.; Simoncelli, E.P.; Bovik, A.C. Multiscale Structural Similarity for Image Quality Assessment. In Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 9–12 November 2003; pp. 1398–1402.
25. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss Functions for Image Restoration with Neural Networks. *IEEE Trans. Comput. Imaging* **2016**, *3*, 47–57. [CrossRef]
26. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409. [CrossRef] [PubMed]
27. Dudhane, A.; Murala, S. RYF-Net: Deep Fusion Network for Single Image Haze Removal. *IEEE Trans. Image Process.* **2019**, *29*, 628–640. [CrossRef] [PubMed]
28. Xia, G.-S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3965–3981. [CrossRef]
29. Qin, Z.; Ni, L.; Tong, Z.; Qian, W. Deep Learning Based Feature Selection for Remote Sensing Scene Classification. *IEEE Geosci. Remote Sens.* **2015**, *12*, 2321–2325.
30. BH-Pools/Watertanks Datasets. Available online: http://patreo.dcc.ufmg.br/2020/07/29/bh-pools-watertanks-datasets/ (accessed on 30 August 2021).
31. Geospatial Data Cloud. Available online: http://www.gscloud.cn (accessed on 30 August 2021).
32. Berman, D.; Treibitz, T.; Avidan, S. Air-light Estimation Using Haze-lines. In Proceedings of the IEEE International Conference on Computational Photography—ICCP 2017, Stanford, CA, USA, 12–14 May 2017; pp. 1–9.
33. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [CrossRef] [PubMed]