A deep learning model for incorporating temporal information in haze removal

Xiaofeng Ma [a], Qunming Wang [a,*], Xiaohua Tong [a], Peter M. Atkinson [b,c]

[a] College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China

[b] Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK

[c] Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK

*Corresponding author. E-mail: wqm11111@126.com.

Abstract: Haze contamination is a very common issue in remote sensing images, which inevitably limits data usability and further applications. Several methods have been developed for haze removal, which is an ill-posed problem. However, most of these methods involve various strong assumptions coupled with manually-determined parameters, which limit their generalization to different scenarios. Moreover, temporal information amongst time-series images has rarely been considered in haze removal. In this paper, the temporal information is proposed to be incorporated for more reliable haze removal, and guided by this general idea, a temporal information injection network (TIIN) is developed. The proposed TIIN solution for haze removal extracts the useful information in the temporally neighboring images provided by the regular revisit of satellite sensors. The TIIN method is suitable for images with various haze levels. Moreover, TIIN is also applicable for temporal neighbors with inherent haze or land cover changes due to a long-time interval between images. The proposed method was validated through experiments on both simulated and real haze images as well as comparison with five state-of-the-art benchmark methods. This research provides a new paradigm for enhancing haze removal by incorporating temporally neighboring images.

Keywords: Haze removal; Remote sensing images; Temporal information; Deep learning; Convolutional neural network.

1. Introduction

Haze arises mostly from atmospheric constituents of water droplets, dust, fog/smog and other particles that dim the clarity of the scene (Cai et al., 2016), and leads to inaccurate measurement of radiance and information loss in remote sensing images, which is similar to thin cloud contamination (i.e., part of (rather than all) the information is lost in thin cloud). Therefore, for images acquired with haze contamination, the visibility, contrast and intensity may be affected greatly (Jiang et al., 2018). The commonly used haze degradation model (Narasimhan and Nayar, 2003) is as follows:

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

where $\mathbf{x}$, $\mathbf{I}$, and $\mathbf{J}$ indicate the location of a pixel in the image, the haze contaminated image, and clear image, respectively. $\mathbf{A}$ is the global atmospheric light, and $t(\mathbf{x})$ indicates the haze transmission map. Therefore, the haze removal process can be formulated as follows:

$$\mathbf{J}(\mathbf{x})=\frac{\mathbf{I}(\mathbf{x})\text{-}\mathbf{A}}{t(\mathbf{x})}+\mathbf{A}. \tag{2}$$

That is, each hazy image can be recovered with global atmospheric light **A** and haze transmission map $t(\mathbf{x})$. For haze removal, only the acquired hazy image is known, and both the global atmospheric light and haze transmission map need to be estimated in advance, which is an obvious ill-posed problem.

Over the last decade, several dehazing methods have been proposed. Tan (2008) addressed dehazing by using Markov random field to maximize the local contrast. Fattal (2008) calculated the albedo of the scene and medium transmission maps based on the assumption of uncorrelation between transmission and surface shading. A dark-channel prior was applied to reconstruct the haze contamination of outdoor images (He et al., 2011). Moreover, a color attenuation prior was employed as a constraint in predicting haze-free images (Zhu et al., 2015). Although the aforementioned methods produced acceptable results in some cases, they are based on various strong assumptions such that their general applicability is affected. Besides, haze removal can also be considered as an image enhancement goal (Cho et al., 2018). Specifically, the input hazy images can be decomposed to produce ambient maps and transmission maps for further refinement based on the Laplacian module, which does not require any prior information. Recently, an enhanced atmospheric scattering model was developed for haze removal (Ju et al., 2021). Most of these methods can be categorized as traditional model-based methods with various assumptions which inevitably limit the generalization ability of the methods for different scenarios.

Deep neural networks have been utilized in many computer vision tasks owing to their strong non-linear modeling ability. For dehazing regular images, Cai et al. (2016) proposed an end-to-end DehazeNet to estimate transmission maps. A multi-scale convolutional neural network (CNN) was applied to explore transmission features in both coarse and fine domains (Ren et al., 2016, 2020).

Both methods use the model in Eq. (2) for haze removal, and the outputs are predicted transmission maps. Thus, post-processing steps are required to produce the haze-free images and the uncertainty in the interim deep neural network-based is propagated to the final dehazed results (Zhu et al., 2018). To reduce such uncertainty, Li et al. (2017) built a network to bridge hazy images and clear images directly which can produce haze-free images without any post-processing steps. A symmetric encoder-decoder structure was employed for dehazing (Ren et al., 2018). Generative adversarial networks (GANs) were applied to handle haze removal without any manually-set prior (Li et al., 2018b; Qu et al., 2019). Recently, Li et al. (2020a) combined Retinex Theory (Land, 1978) with neural networks to remove the haze of regular images.

Haze removal of remote sensing images is more difficult compared with that for regular images due to the sophisticated atmosphere, complex spatial textures, and abundant spectral information of remote sensing images. Zhang et al. (2002) employed a haze optimized transformation algorithm for hazy Landsat images. Wavelet analysis was also applied to remove the haze of fine spatial resolution remote sensing images (Du et al., 2002). Makarau et al. (2014) calculated a haze thickness map via dark-object subtraction to dehaze both calibrated and uncalibrated multispectral images. The correlation between the visible (or infrared) band and the cirrus band was also utilized for haze removal in Xu et al. (2014). Moreover, a cloud removal noise-adjusted principal components transform (CR-NAPCT) method (Xu et al., 2019) was employed for Landsat-8 images with additional cloud detection operators, such as Fmask developed in Zhu and Woodcock (2012), which may cause intermedium uncertainty for post-cloud removal. Based on the dark image prior of regular image dehazing (He et al., 2011), a deformed haze imaging model was introduced to dehaze remote sensing images (Pan et al., 2015). However, this method can handle only the RGB bands of remote sensing images. Furthermore, the sphere model (Li et al., 2018a) and elliptical boundary prior (Guo et al., 2019) can also be employed for haze removal. Similarly, an empirical method was developed for visible bands 1-to-4 of Landsat-8 (Lv et al. 2016). Guo et al. (2020a) utilized the haze degradation model in Eq. (2) to remove haze bands with different wavelengths, which takes both haze particle size and concentration into account during haze removal. However, these model-based methods make various assumptions between the ideal clear and hazy images. Moreover, they need fine-tuned model parameters for different haze condition scenarios.

In recent years, various learning-based (e.g., CNN-based) dehaze methods have been developed with state-of-the-art performance for remote sensing images. Jiang and Lu (2018) applied a multi-scale residual CNN to estimate the transmission maps before dehazing based on the model in Eq. (2). Compared with Ren et al. (2016), Jiang and Lu (2018) employed dilation convolution for feature extraction at different scales. In contrast, several convolutional layers were employed to predict haze-free remote sensing images directly and considered the haze variation of different wavelengths (Qin et al., 2018). However, this method requires a large number of haze levels when constructing the training data, which inevitably increases the computational burden. In Guo et al. (2020b), the haze variation between different bands was considered in global residual learning with channel attention for dehazing Landsat-8 OLI images. GANs can also be employed for haze removal (Li et al., 2020b).

The commonly used satellite sensors, such as the Landsat series and the Terra/Aqua MODerate resolution Imaging Spectroradiometer (MODIS), can provide a large number of remote sensing images of the same region at different times, due to their regular revisit capabilities (Wang et al., 2020). Theoretically, the temporally neighboring images can provide complementary information to tackle the haze removal issue as formulated in Eq. (2). Specifically, the temporally neighboring images can provide a spatial distribution prior for the target hazy images because of the temporal correlation between observations (i.e., images acquired on two proximate days tend to resemble each other, especially when the time interval is small). Therefore, the uncertainty of this ill-posed issue can be reduced potentially. Although temporal information has been considered in other image restoration issues, such as thick cloud removal (Chen et al., 2020; Ji et al., 2021) and Landsat ETM+ SLC-off gap filling (Wang et al., 2021), it has been neglected in existing haze removal studies. The use of temporal information in haze removal is quite different from that for thick cloud removal. Haze contamination is highly correlated to spectral wavelength. Generally, longer wavelength bands are more robust to haze. Moreover, hazy pixels are usually a mixture of haze and the original signal of land covers, such that part of the original signal is retained. This is different from the case of completely dead pixel through all bands caused by thick cloud contamination or a SLC-off gap. Thus, the scheme of using temporally neighboring images in thick cloud removal (Shen et al., 2015) (i.e., completely neglecting the information under thick cloud as it contains no information) or gap filling (Wang et al., 2021) is not appropriate for haze removal, as it would waste the potentially valuable information in the hazy pixels, especially for longer wavelength bands.

In this research, temporal information is considered for haze removal and a novel solution incorporating temporally neighboring images is proposed. It should be stressed that two issues arise when temporally neighboring images are used. First, driven by natural evolution and human activities, the land cover type of the Earth surface usually changes in the temporal domain. Therefore, the spatial distribution prior in temporally neighboring images can be different to the target hazy image. Second, the temporally neighboring images may also be contaminated by haze. That is, although the observed temporally neighboring images may be abundant, the effective spatial distribution prior in the temporal domain is reduced due to land cover changes and haze contamination. These two challenges also hamper the development of using temporal information in haze removal, especially for model-based methods. Generally, model-based methods involve various assumptions and the solutions for incorporating temporal information may not be straightforward especially for solutions that are universal for different scenarios. Specifically, model-based methods cannot select automatically the usable spatial distribution prior in the temporal images. Since deep learning is fully data-driven and can automatically transform the feature representation into a higher and more abstract level (LeCun et al., 2015; Li et al., 2020c; Shao et al., 2019; Wu et al., 2021). Thus, by deep learning, various input images can be automatically distilled into abstract levels without any specific assumptions. Therefore, this paper investigates this type of method for incorporating temporally neighboring images that can be affected by haze and land cover changes. Accordingly, to distill the effective spatial distribution prior of temporal information for dehazing, a temporal information injection network (TIIN) is proposed. The TIIN method convolves both hazy images and temporally neighboring images in a parallel manner with stacking layers for feature extension, and emphasizes the useful temporal features for dehazing by using different attention modules.

The remainder of this paper is organized as follows. In Section 2, the mechanism of the proposed TIIN architecture is presented. The experimental results of both simulated and real hazy data are provided in Section 3. Section 4 discusses the findings and the problems to be investigated further. Section 5 summarizes the conclusion of this study.

2. Methods

2.1. Overview of the TIIN architecture

The proposed TIIN method is a parallel CNN-based architecture, as shown in Fig. 1. The architecture contains two branches (i.e., hazy branch and temporal branch) and several blocks. At the beginning, both hazy images and temporally neighboring images are input to the corresponding branches simultaneously. In each branch, two convolutional layers are applied for shallow feature extraction with 32 filters. Subsequently, the group convolution block, as shown in Fig. 2(a), is employed to extract land cover information. Next, three temporal information injection (TII) blocks are applied to transfer information from the temporal images. Then, a concatenation and fusion block is applied to integrate the extracted features of the branches, followed by a modified spatial attention (MSA) block to focus on hazy regions. At the tail of the architecture, based on the Retinex Theory (Land, 1978; Li et al., 2020a), a global multiply residual is used to alleviate the burden of network training. Finally, a 3×3 convolutional layer with seven filters is applied to transform the haze-free features to haze-free Landsat-8 OLI images. Note that all convolutional layers in TIIN, except for the final convolution layer, include 32 channels for convolution and the rectified linear unit (ReLU) is used for function activation.

2.2. The group convolution block

The group convolution block consisting of three layers is used for feature extension by extracting multiscale semantic and contextual information. Diverse filter sizes of the convolutional layers can make the model more suitable for reconstruction of scenarios with different sizes of land cover objects. Specifically, in this research, the filter sizes of the three layers are 1×1 pixel, 3×3 pixels and 5×5 pixels. Moreover, according to Guo et al. (2020b)'s study, the channel number of each group convolutional layer is set to 32, which can provide sufficient useful features. Subsequently, using a concatenation strategy, the extracted multiscale features of both the hazy and temporal branches are concatenated into unified features in the channel dimension. Finally, in each branch, the integrated features are fed into a fusion block for further feature integration.

2.3. The TII block

The TII block is used to bridge the gap between hazy and temporal branches. As shown in Fig. 2, the output features of each basic block of the temporal branch are injected into a channel attention (CA) module, which are then concatenated with the output of corresponding basic block of the hazy branch. The CA block can utilize fully the correlation of different bands between hazy images and temporally neighboring images. Moreover, the CA block bridges the gap between the hazy images and temporally neighboring images. That is, the spatial distribution prior in the temporal branch is transferred into the hazy branch via the CA block. Specifically, the CA module includes squeeze, excitation and recalibration (Hu et al., 2020). First, a global average pooling (GAP) layer is used to provide a global spatial information squeeze. Then, for excitation, a 1×1 filter is used for channel-wise feature reduction with filter number $\frac{C}{r}$ (C is the number of feature channels and r denotes the reduction ratio). In this research, the reduction ratio was determined as 4. Next, a convolutional layer with C filters (each with a size of 1×1) is employed to increase the dimensionality for further excitation, followed by a sigmoid activation function. After excitation, a residual multiplication layer is used for channel-wise feature recalibration. Afterwards, a fusion block is applied to integrate the CA-derived features. The fused features consider the spatial information in both the hazy images and the spatial distribution prior from the temporally neighboring images.

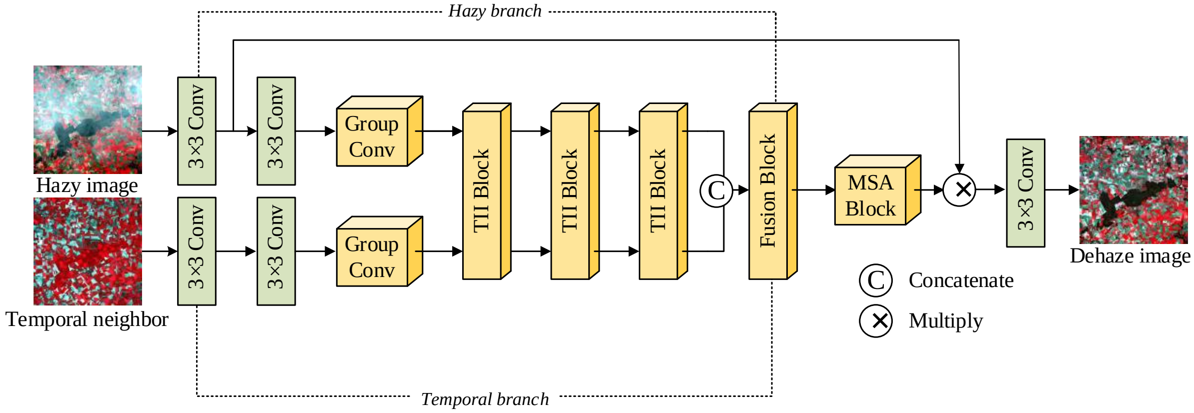

Fig. 1. The whole architecture of the proposed TIIN.

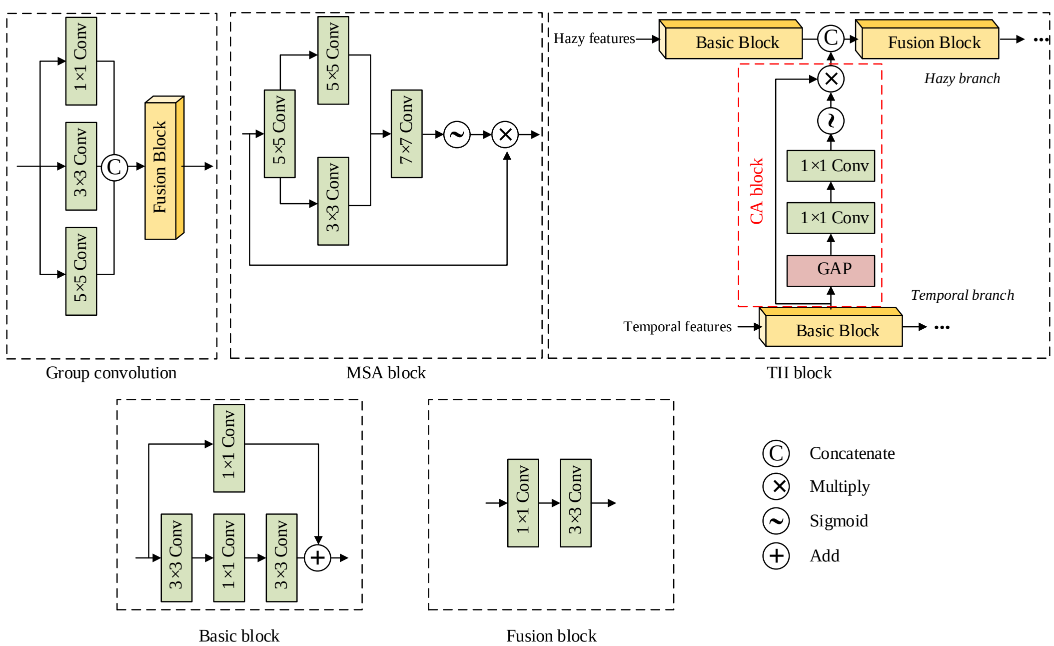

Fig. 2. Different components of the proposed architecture in Fig. 1.

2.4. The MSA block

After the TII block, the extracted features from both the hazy and temporal branches are integrated via a concatenation and fusion block. To focus on the hazy regions in the spatial domain, the spatial attention block is applied. However, spatial attention usually utilizes global average pooling and maximal pooling to capture global common and distinctive information (Li et al., 2020a), which is non-learnable. For robustness, the pooling operators of spatial attention are replaced by convolutional layers in this paper. The MSA strategy is applied to adjust automatically the spatial-wise weights. As shown in Fig. 2, the pooling operator in spatial attention consists of multiscale convolutional layers with filter sizes of 3×3 pixels and 5×5 pixels, followed by a 7×7 pixels convolutional layer to extract more multiscale contextual information. It should be noted that only one channel is considered consistently for these convolutional layers to achieve spatial attention, with only one sigmoid function for activation in the final layer.

2.5. The basic and fusion blocks in TIIN

Since remote sensing images always contain complicated spatial context and semantic information, hierarchical CNNs are considered to represent the spatial features. Therefore, basic blocks with residual learning are applied to enhance the performance of feature extraction. The detail of the proposed basic block is portrayed in Fig. 2. The basic block is a derivative of a residual model. Specifically, as shown in Fig. 2, three convolutional layers are applied for feature extraction with the ReLU as activation function. Meanwhile, the input of the basic block is convolved via a 1×1 filter and then skip-connected with the output of the last convolutional layer for residual learning.

To integrate the various feature maps derived from two branches of the network, a fusion block is employed, as depicted in Fig. 2. Two convolution layers are applied to aggregate different feature maps and the filter sizes are 1 and 3. These layers also apply ReLU for activation.

2.6. Residual learning and the loss function in TIIN

In the Retinex Theory (Land, 1978; Li et al., 2020a), a hazy image can be considered as a dehazed image multiplied by the residual illumination map. Residual learning is a popular strategy to reduce the requirement for network training (He et al., 2016), as the residual learning strategy can estimate the residual between the network input and the reference. Hence, the expected results are the summation of residual (network outputs) and the network inputs. Based on the theory, a global product residual operator is proposed for haze removal. Specifically, the summation operator of residual learning is modified to an elementwise product operator. Therefore, the expected dehazed result is the elementwise product of $f(\mathbf{J})$ and $F(\mathbf{J})$ in Eq. (3):

$$\mathbf{I}' = f(\mathbf{J}) \cdot F(\mathbf{J}) \tag{3}$$

where $\mathbf{I}'$ denotes the haze-free features, $\mathbf{J}$ is the hazy image, $f(\mathbf{J})$ is the output of the first convolutional layer of the shallow feature extraction, and $F(\mathbf{J})$ is the output of the MSA block. Based on the Retinex Theory, $f(\mathbf{J})$ can be seen as hazy features and the reciprocal of $F(\mathbf{J})$ is the residual illumination map. The final 3×3 pixels filter can transform the haze-free feature $\mathbf{I}'$ to the dehaze result $\mathbf{I}$.

In terms of the loss function for the network, the widely-used mean square error (MSE) loss function is considered to guide the network training iteratively.

3. Experiments

Comprehensive experiments on simulated data and real-data were carried out to examine the robustness and applicability of the proposed TIIN method. Specifically, Landsat-8 OLI data of the L1 level Collection 2 product (30 m spatial resolution and 16-day temporal resolution) were used in this study. Moreover, hazy images with different land cover scenarios and temporally neighboring images with long-time interval were also considered.

Table 1. Details of the training images used in the experiments

No.		Date	Centre position	Size
Pair 1	Original haze-clear images	2019.06.02	48°40’38.58’’N, 2°20’0.30’’E	3741×5023
	Temporally neighboring images	2019.07.04	48°40’38.58’’N, 2°20’0.30’’E	3741×5023
Pair 2	Original haze-clear images	2018.10.05	48°52’42.90’’N, 2°21’18.49’’E	5094×4995
	Temporally neighboring images	2018.10.21	48°52’42.90’’N, 2°21’18.49’’E	5094×4995

3.1. Data preparation

The available data were organized into training data and testing data. To provide training images for the proposed TIIN architecture, several haze-clear images and temporally neighboring images of Landsat-8 OLI with seven bands (including coastal, blue, green, red, NIR, SWIR1, and SWIR 2 bands) covering Paris, France (path is 199, and row is 26) were acquired. Note that the output of the TIIN architecture is an image with the same size (i.e., the number of bands is also the same as the input). Moreover, the training data are spatially neighboring to the testing data. More detailed information on the acquired images is depicted in Table 1. It should be noted that the temporal neighboring images mean the images that cover the same location as the original image but were acquired at different times. The max-min scaling normalization was applied to each band of the Landsat-8 OLI images. Moreover, the uncertainty in image registration between the original images and temporal neighbors was ignored due to the high confidence in the registration of the Landsat-8 product (Irons et al., 2012).

The CNN is a supervised learning strategy, which needs massive labelled data for training. For image dehazing, however, it is unrealistic to acquire both haze-clear and haze-contaminated conditions for the same scene at the same time. Therefore, a haze simulation strategy was implemented based on Guo et al. (2020b) to generate sufficient training data for more reliable fitting. First, the haze transmission map $t(\mathbf{x})$ in Eq. (1) can be estimated:

$$t(\mathbf{x})=e^{-\beta(\lambda,\ \gamma(\mathbf{x}))d} \tag{4}$$

where λ indicates the wavelength of a certain band of the hazy image, β is the scattering coefficient, d is the distance between satellite sensors and surface objects, and $\gamma(\mathbf{x})$ is a spatial-based function to determine the haze spatial distribution at pixel $\mathbf{x}$ (Guo et al., 2020b). In general, γ ranges from 0 to 4 (Chavez, 1988; Guo et al., 2020b). The scattering coefficient can be estimated as:

$$\beta(\lambda,\ \gamma(\mathbf{x}))=T\lambda^{-\gamma(\mathbf{x})} \tag{5}$$

where T is a constant. Therefore, the haze imaging model can be transformed as follows:

$$\mathbf{I}(\mathbf{x})=\mathbf{J}(\mathbf{x})e^{-\beta(\lambda,\ \gamma(\mathbf{x}))d}+\mathbf{A}(1-e^{-\beta(\lambda,\ \gamma(\mathbf{x}))d}) \tag{6}$$

Moreover, the first band of the Landsat-8 OLI image was selected as the reference band. Then, a natural logarithm was implemented on both sides of Eq. (4) to further derive:

$$\ln t(\mathbf{x})=-\beta(\lambda,\ \gamma(\mathbf{x}))d \tag{7}$$

The ratio between the first band and the other bands can be calculated as follows:

$$\frac{\ln t_1(\mathbf{x})}{\ln t_i(\mathbf{x})}=\frac{\beta_1(\lambda_1,\ \gamma(\mathbf{x}))}{\beta_i(\lambda_i,\ \gamma(\mathbf{x}))} \tag{8}$$

Based on Eq. (8), the transmission map $t(\mathbf{x})$ of each band of the hazy Landsat-8 OLI image can be estimated. Therefore, each hazy band of the Landsat-8 OLI image can be expressed as:

$$\mathbf{I}_i(\mathbf{x})=\mathbf{J}_i(\mathbf{x})e^{(\frac{\lambda_1}{\lambda_i})^{\gamma(\mathbf{x})}\ln t_1(\mathbf{x})}+\mathbf{A}(1\text{-}e^{(\frac{\lambda_1}{\lambda_i})^{\gamma(\mathbf{x})}\ln t_1(\mathbf{x})}) \tag{9}$$

In Eq. (9), $t_1(\mathbf{x})$ is a reference transmission map. To be close to the real haze conditions, a cirrus band from a cloudy region was employed as reference in this paper to simulate nonuniform haze cover, as in Guo et al. (2020b). Hence, the reference transmission map can be formulated as:

$$t_1(\mathbf{x})=1\text{-}\omega c(\mathbf{x}) \tag{10}$$

where $c(\mathbf{x})$ is the selected cirrus band of the Landsat-8 OLI image, and the weight coefficient ω ranges from 0 to 1, controlling the haze contamination level. Normally, larger ω indicates heavier haze.

For haze simulation, 35 cirrus maps of Landsat-8 OLI from different cloudy regions were acquired as reference transmission maps during training samples preparation. In Eqs. (9) and (10), the unknown parameters include the global atmospheric light $\mathbf{A}$, spatial-based function $\gamma(\mathbf{x})$, weight coefficient ω, and the wavelength of the i-th band λ_i. For global atmospheric light $\mathbf{A}$, the strategy of Guo et al. (2020b) was used, and $\gamma(\mathbf{x})$ was set to 1 in training. To be applicable for various haze conditions, the weight coefficient ω was parameterized randomly from 0 to 1, with an interval of 0.1 for each image to simulate different haze contamination. The central wavelength of the Landsat-8 OLI images is used to set λ_i during haze simulation. Finally, the simulated haze images and temporally neighboring images were fed into the proposed architecture under the supervision of the corresponding reference haze-clear images.

Training data were prepared by cropping the simulated hazy images, temporally neighboring images, and original haze-clear images with 33635 mini-patches. To ensure sufficient patches, each mini-patch was cropped with a spatial size of 32×32 pixels. Using rotation and flipping for data augmentation, 100905 mini-patches were eventually produced as training data. Moreover, 10% of the set of training mini-patches were deployed for validation. For network training, the Adam optimization operator (Kingma, 2015) was applied with momentum parameters set to 0.9, 0.999, and 10^{-8}, respectively. The entire architecture was trained iteratively via 100 epochs with a learning rate of 10^{-4}. The whole model was implemented in Windows 10 equipped with an NVIDIA GTX 2080 Ti graphics processing unit. The training time was about 4 h. In addition, the architecture was implemented in the Keras framework via TensorFlow as back-end.

Table 2. Details of the images used in the experiments

No.	Regions		Date	Centre Position	Size
Case 1	Region 1	Original image	2019. 06. 02	49°5’27.62’’N,	300×300
		Temporal neighbor	2019. 07. 04	3°39’51.73’’E	
	Region 2	Original image	2020. 04. 01	48°38’42.14’’N,	500×500
		Temporal neighbor	2020. 05. 19	3°11’24.46’’E	
	Region 3	Original image	2020. 04. 01	49°25’30.81’’N,	400×400
		Temporal neighbor	2020. 05. 19	3°38’26.67’’E	
	Region 4	Original image	2013. 12. 10	49°1’59.74’’N,	400×400
		Temporal neighbor	2014. 03. 16	1°43’33.38’’E	
Case 2	Region 5	Original image	2019. 02. 26	48°37’28.80’’N,	400×400
		Temporal neighbor	2019. 06. 18	2°56’19.92’’E	
Case 3	Region 6	Original image	2020. 04. 01	46°58’29.69’’N,	500×500
		Temporal neighbor	2020. 05. 19	0°37’44.52’’E	
Real haze	Region 7	Original image	2020. 07. 22	49°1’32.41’’N,	800×800
		Temporal neighbor	2020. 08. 07	2°29’20.48’’E	
Real haze	Region 8	Original image	2020. 07. 22	48°57’38.73’’N,	1000×1000
		Temporal neighbor	2020. 08. 07	2°18’7.49’’E	
Real haze	Region 9	Original image	2021. 11. 23	31°53’51.86’’N,	900×900
		Temporal neighbor	2021. 04. 29	121°53’18.32’’E	
Real haze	Region 10	Original image	2019. 04. 15	49°6’13.18’’N,	2000×2000
		Temporal neighbor	2019. 02. 26	1°43’13.09’’E	
Real haze	Region 11	Original image	2019. 04. 15	48°23’19.43’’N,	2000×2000
		Temporal neighbor	2019. 02. 26	2°34’5.97’’E	

To facilitate the validation of the proposed TIIN solution for haze removal objectively, several Landsat-8 OLI images located near 48°N and 2°E were collected for haze simulation to test the different haze removal methods. Specifically, three groups of haze-clear Landsat-8 OLI images with different spatial extents were contaminated based on the haze simulation model in Eq. (10). Moreover, four real hazy images were used to examine the practicability of the proposed method. More details of the used images are depicted in Table 2.

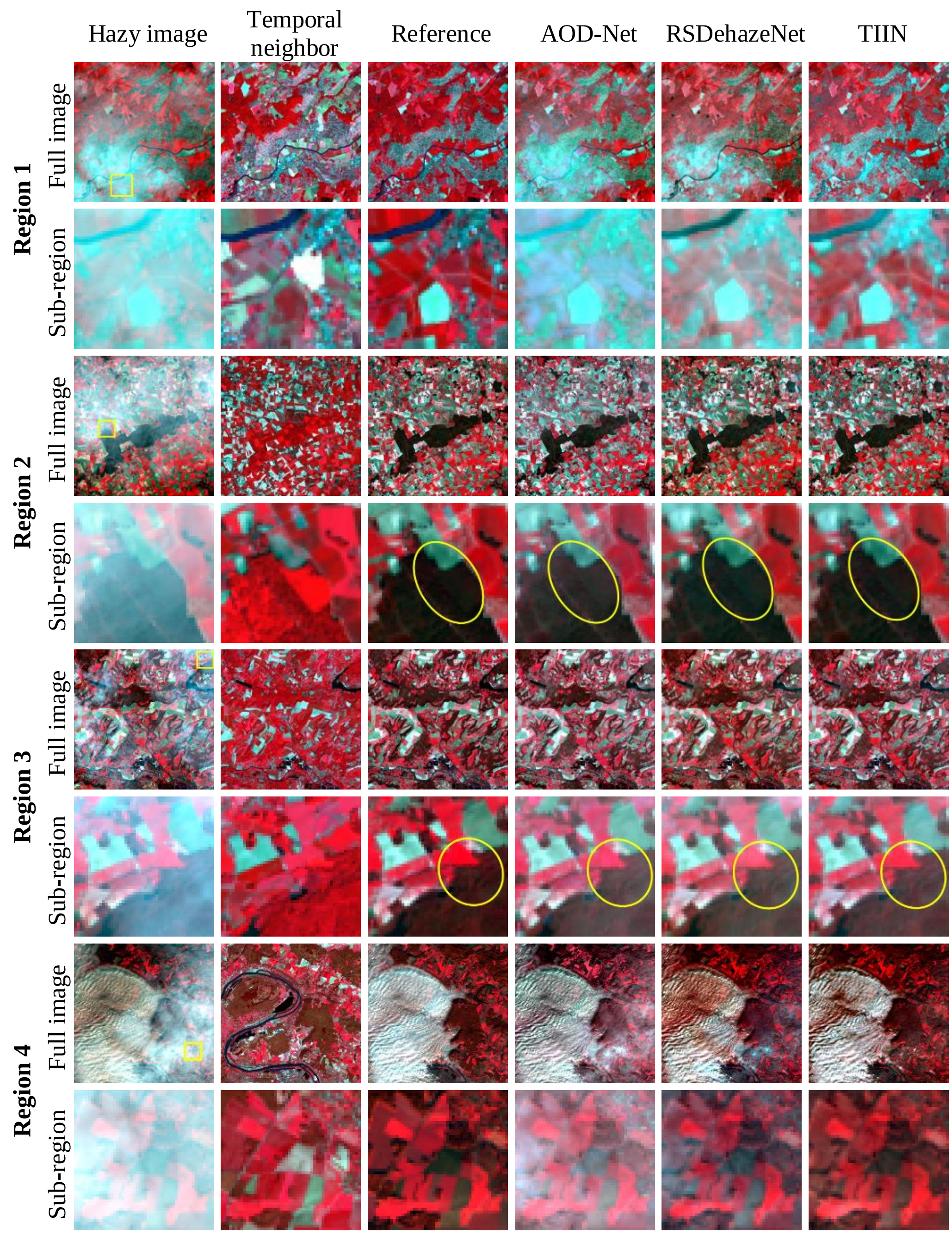

Fig. 3. Results of different methods for the four regions (Region 1: 300×300 pixels, Region 2: 500×500 pixels, Region 3: 400×400, Region 4: 400×400 pixels) and the corresponding enlarged sub-regions (yellow rectangle region in the corresponding full images) in Case 1 (NIR, red, and green as RGB).

3.2. Experiments on simulated haze data

In this section, the feasibility of the proposed TIIN solution for different scenarios was validated with both visual and quantitative assessment, as the reference data representing the haze-clear images are known perfectly. Specifically, three different cases were implemented, including various land cover scenarios with different levels of haze (Case 1), temporally neighboring images with long-time interval (Case 2), and haze image spatially distant to the training images (Case 3). Moreover, the proposed method was compared with four state-of-the-art dehaze methods existing in both the computer vision and remote sensing communities, including two learning-based methods (i.e., AOD-Net (Li et al., 2017) and RSDehazNet (Guo et al., 2020b)) and three model-based methods (i.e., the method in Cho et al. (2018), the automatic cloud removal method (ACRM) (Xu et al., 2014)) and CR-NAPCT (Xu et al., 2019). The correlation coefficient (CC), universal image quality index (UIQI) and root mean square error (RMSE) were employed to quantitatively evaluate the accuracy of the different methods.

3.2.1. Case 1 (various land cover scenarios with different levels of haze)

In this case, four images covering different land cover scenarios and simulated with different levels of haze (i.e., heavy haze for Regions 1 and 4; and moderate haze for Regions 2 and 3) were considered. The dehaze results of the four regions are displayed in Fig. 3. The sub-regions indicate the enlarged region of the yellow area in the original images. In Region 1, both AOD-Net and RSDehazeNet cannot recover the heavy haze region satisfactorily, where color distortion can be observed clearly. The proposed TIIN method can produce a more accurate result, although the temporally neighboring image contains apparent land cover changes (such as the decrease in vegetation cover). The same advantage of TIIN is seen in the results of Region 2. In Region 3, even with a 49-day temporal distance, the proposed method can still fully utilize the spatial distribution prior of the temporally neighboring image and produce more acceptable dehaze results than AOD-Net and RSDehazeNet. This advantage is attributed to the parallel stacking convolutional layers and different attention modules of the proposed network. Moreover, the proposed TIIN method can still produce visually pleasant results under large areas of snow cover in Region 4. The same conclusion can also be drawn from Fig. 4, where the error maps are provided for clearer visualization of the difference between the various methods. Generally, in each band, the proposed method can produce dehaze results with less error than the other methods.

Quantitative assessment results for Case 1 are presented in Table 3. It is seen that the proposed TIIN method can generally produce the most accurate dehaze results. Specifically, the proposed method can produce larger mean CC and UIQI and smaller mean RMSE for almost all bands of the Landsat-8 OLI images in the four regions. For example, in Region 1, compared with AOD-Net and RSDehazeNet, the mean CC of TIIN are 0.1698 and 0.1677 larger, respectively. Correspondingly, the mean RMSE of TIIN is 0.0102 and 0.0201 smaller than the two methods.

3.2.2. Case 2 (temporally neighboring images with long-time interval)

In this case, to evaluate the robustness of the TIIN method in relation to the long-time interval of temporally neighboring images, a Landsat-8 OLI image with a 112-day temporal distance was applied in TIIN. As seen from the results for Case 2 in Fig. 5, AOD-Net cannot remove the haze thoroughly. Both RSDehazeNet and TIIN can produce cleaner images. As depicted in the corresponding enlarged regions, however, RSDehazeNet fails to reconstruct the spectral information precisely, presenting spectral distortion. On the contrary, the proposed TIIN solution recovers the haze region more accurately, as it takes full advantage of the available temporal information and also can take land cover changes into account. Quantitative evaluation for Case 2 is depicted in Table 3. Both mean CC and UIQI of the proposed TIIN method are larger than for the other two methods, and the mean RMSE is also smaller.

Table 3. Quantitative assessment of different dehaze results in the simulated experiment

		Methods	**CC**	**UIQI**	**RMSE**
Case 1	Region 1	AOD-Net	0.7159	0.6530	0.0642
		RSDehazeNet	0.7180	0.6172	0.0741
		TIIN	**0.8857**	**0.8133**	**0.0540**
	Region 2	AOD-Net	0.9596	0.9147	0.0417
		RSDehazeNet	0.9385	0.8922	0.0381
		TIIN	**0.9696**	**0.9489**	**0.0289**
	Region 3	AOD-Net	0.9500	0.9164	0.0390
		RSDehazeNet	0.9280	0.8797	0.0420
		TIIN	**0.9585**	**0.9275**	**0.0310**
	Region 4	AOD-Net	0.7908	0.5540	0.1403
		RSDehazeNet	0.8808	0.6949	0.1330
		TIIN	**0.9453**	**0.8449**	**0.1044**
Case 2		AOD-Net	0.8703	0.8529	0.0192
		RSDehazeNet	0.9571	0.9447	**0.0119**
		TIIN	**0.9740**	**0.9585**	0.0138
Case 3		AOD-Net	0.8919	0.8312	0.0458
		RSDehazeNet	0.8324	0.7840	0.0472
		TIIN	**0.9158**	**0.8857**	**0.0326**

Table 4. Classification accuracy of the land cover maps derived from different dehaze results

	Region 1		**Region 3**	
	OA	Kappa	OA	Kappa
AOD-Net	0.7594	0.5083	0.9821	0.0267
RSDehazeNet	0.8048	0.5710	0.9887	0.5425
TIIN	0.9269	0.8167	0.9970	0.9077

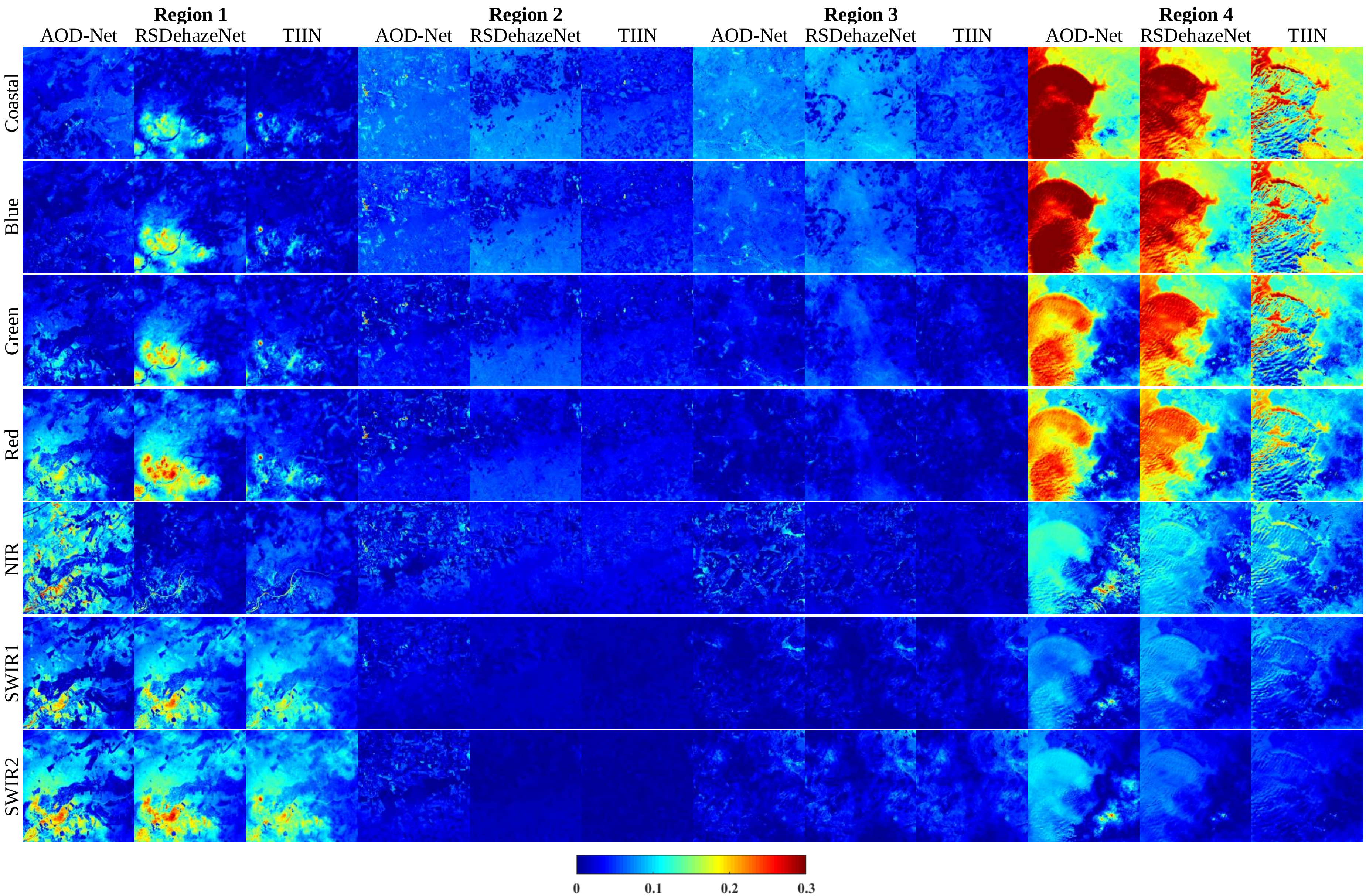

Fig. 4. Error maps (in absolute value) of different dehaze results of the three regions in Case 1.

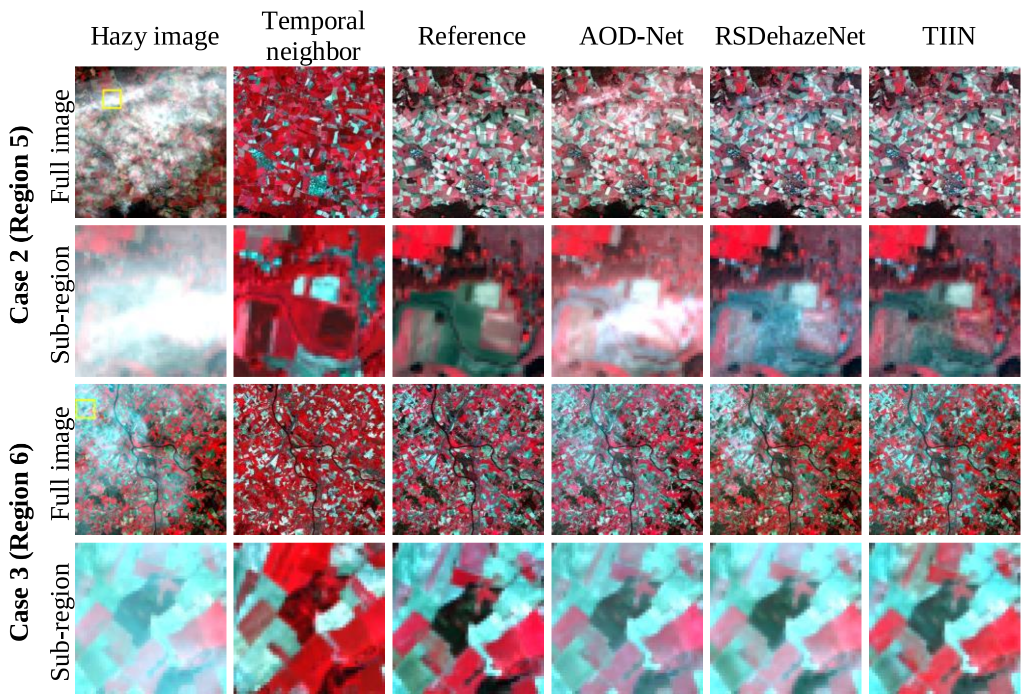

Fig. 5. Results of Cases 2 (400×400 pixels) and 3 (500×500 pixels) (NIR, red and green as RGB).

3.2.3. Case 3 (haze image spatially distant to the training images)

In the aforementioned experiments, the images for prediction were spatially in the same tile (path 199, and row 26, with a size of 7911×8011 pixels) as the training images (but acquired at different times). In this section, a scene (path 199, row 27) spatially distant to the training images was used and cropped as Region 5 to examine the proposed solution. The dehaze results of TIIN and the benchmark methods are presented in Fig. 5. It is clear that AOD-Net fails to remove the haze fully. Likewise, there is still haze remaining in the RSDehazeNet result. The proposed TIIN method can produce a dehaze result that is visually clearer and spectrally closer to the reference. The accuracy indices in Table 3 also suggest that the proposed method is more accurate. The mean CC and UIQI of TIIN are generally the largest.

3.3. Application examples

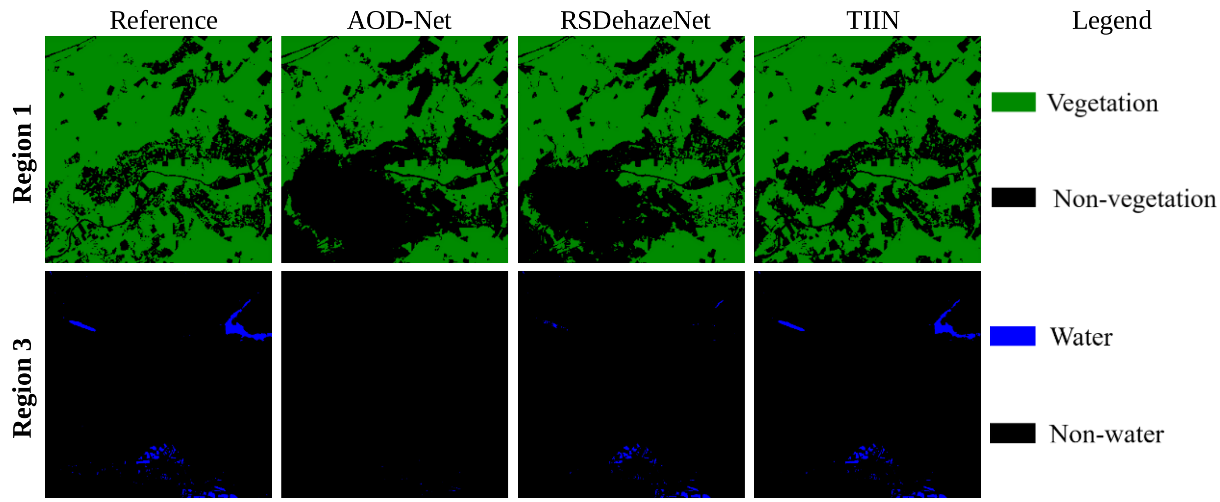

Fig. 6. Classification results based on the different haze removal methods (Region 1: 300×300 pixels, Region 3: 400×400 pixels).

To further analyze the application capability of the methods, post-processing of the dehaze results was considered, including normalized difference vegetation index (NDVI) image and normalized difference water index (NDWI) image. The dehaze results and corresponding haze-clear images in Regions 1 and 3 were acquired for the experiment. NDVI was applied to represent vegetation cover from both the original haze-clear image and the dehaze results in Region 1. Specifically, based on Guo et al. (2020b), the NDVI images were classified with a threshold of 0.5, such that pixels with NDVI larger than 0.5 were determined as vegetation and *vice versa*. The classification results are shown in Fig. 6. The overall accuracy (OA) and Kappa index were presented in Table 4 for quantitative assessment. It is obvious that the classification result of the proposed method is closer to with the reference than for AOD-Net and RSDehazeNet.

Since several lakes exist in Region 3, NDWI was utilized to evaluate the spectral preservation of the methods. Specifically, the NDWI images of different dehaze results were classified, where a pixel was determined as the water class if its NDWI is larger than 0.1. The classification results are shown in Fig. 6. Compared with the reference image, the proposed method can produce the most similar classification map to the reference.

3.4. Comparison with model-based methods

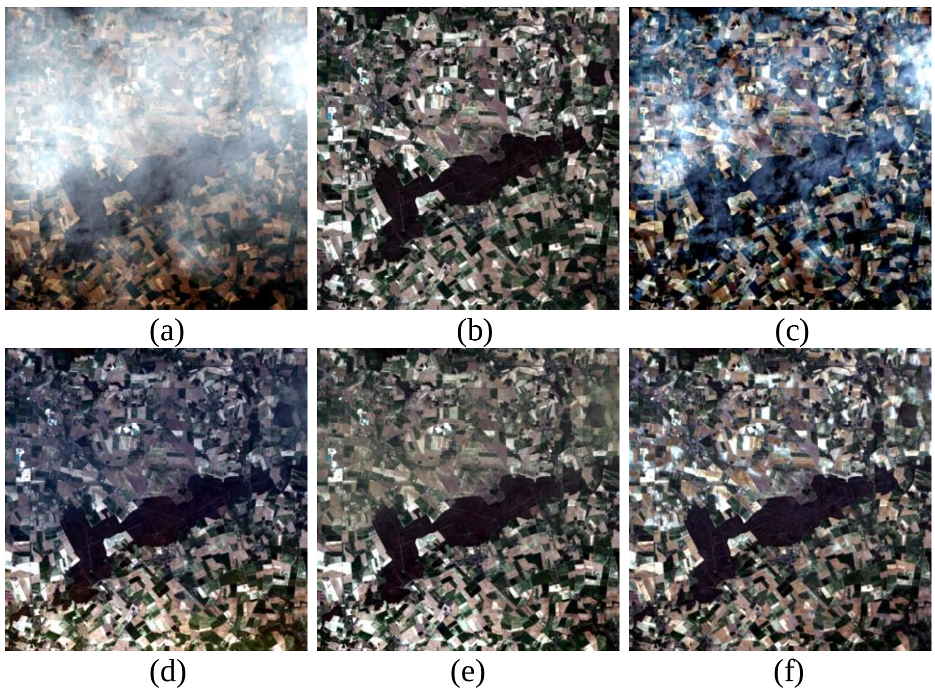

Fig. 7. Dehaze results of the model-based methods for Region 2 (true-color; Region 2: 500×500 pixels). (a) Hazy image. (b) Reference. (c) Cho et al. (2018). (d) ACRM. (e) CR-NAPCT. (f) TIIN.

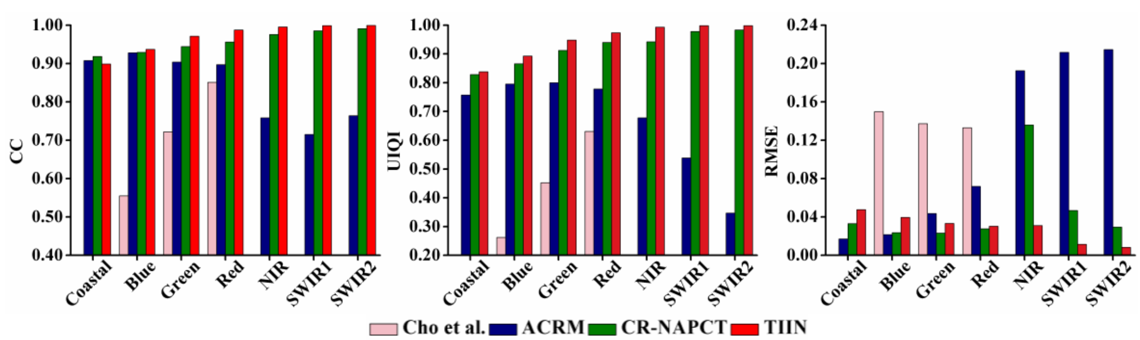

Fig. 8. Quantitative assessment of the model-based methods for Region 2. Note that only the RGB bands were considered as the Cho et al. (2018) method can only deal with these three bands.

To comprehensively validate the advantage of the proposed solution, model-based methods including Cho et al. (2018), ACRM (Xu et al., 2014) and CR-NAPCT (Xu et al., 2019) were employed. Specifically, the simulated haze image in Region 2 was used for validation. It should be noted that Cho et al.'s method was designed to handle the haze in regular images composed of only RGB bands. Therefore, the RGB bands of the dehaze results of both ACRM, CR-NAPCT and TIIN were extracted for visual comparison, as displayed in Fig. 7. Compared with the reference image in Fig. 7(b), obvious haze remains in the result of Cho et al. (2018). ACRM presents apparent color distortion. Moreover, slight color distortion exists in CR-NAPCT. On the contrary, the proposed TIIN solution produces a more accurate result than the Cho et al.'s, ACRM and CR-NAPCT methods.

Quantitative assessment of the dehaze results for all bands is shown in Fig. 8. Note that only the RGB bands were considered as the Cho et al. (2018) method can only deal with these three bands. As shown in Fig. 8, the proposed solution can produce larger CC and UIQI than the Cho et al.'s, ACRM and CR-NAPCT methods. For example, the UIQIs of the Cho et al. results are much smaller than ACRM, CR-NAPCT, and TIIN owing to the apparent haze remaining in Fig. 7. Moreover, the accuracy of CR-NAPCT is also smaller (with mean CC and RMSE of 0.9571 and 0.0457, respectively) than our method (with mean CC and RMSE of 0.9696 and 0.0289, respectively).

3.5. Effect of the haze level in hazy images

To analyze the applicability of the dehaze methods to tackle different haze levels, the weight coefficient ω in Eq. (10) was varied from 0.1 to 1 with an interval of 0.1. As ω increases, the haze contamination is heavier. The haze-clear images in two regions of Case 1 and the corresponding temporally neighboring images were collected for experiment.

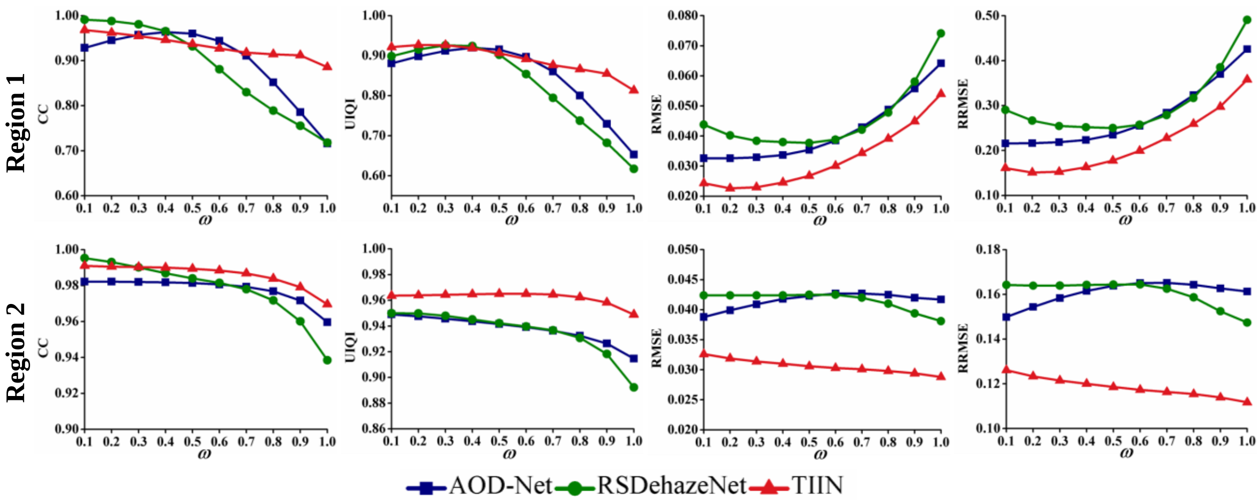

Fig. 9. Accuracy of different methods under various weight coefficients indicating different haze levels (larger weights indicate heavier haze).

The three learning-based methods (i.e., AOD-Net, RSDehazeNet and TIIN) were implemented for two regions in Case 1 and the accuracy indices of the results are displayed in Fig. 9. For all three methods, the accuracies vary apparently under different haze levels. Specifically, the CC and UIQI of AOD-Net and RSDehazeNet decrease with increasing haze, and the corresponding RMSE increases noticeably, especially for Region 1. To reduce the influence of the magnitude of reflectance, the relative RMSE (RRMSE) (Tang et al., 2020) was used. The variation in RRMSE is aligned with the RMSE. However, the decrease in RRMSE for Region 2 is not obvious. This may be attributed to the different spatial heterogeneity and haze contamination due to cirrus cloud in the regions. Furthermore, the proposed TIIN solution can produce more stable results and the advantage is greater when the haze is heavy.

3.6. Effect of the haze level in temporally neighboring images

In the proposed TIIN method, the spatial distribution prior provided by the temporally neighboring image plays an important role in haze removal. Generally, the temporally neighboring images may also be contaminated by haze. Therefore, to examine the robustness and applicability of the proposed solution, the haze contamination at different levels in the temporally neighboring images was also considered. Likewise, the haze-clear images and the corresponding temporally neighboring images of Regions 1 and 2 were assembled for the experiment. The results are shown in Fig. 10, where the weight coefficient ω=0 indicates the haze-clear temporal neighbor. For the two benchmark methods (i.e., AOD-Net and RSDehazeNet), they do not need the temporal information. Hence, their accuracies are invariant in relation to the haze level in the temporally neighboring images, as depicted in Fig. 10. It is seen that the proposed solution can produce larger CC and UIQI for most haze levels in the temporally neighboring image. The haze in the temporally neighboring images decreases the accuracy of haze removal because of the decreased amount of information in the spatial distribution prior. However, TIIN is still applicable to cases where haze also exists in the temporally neighboring images, but it is more advantageous when the haze is not heavy.

3.7. Effect of temporal distance

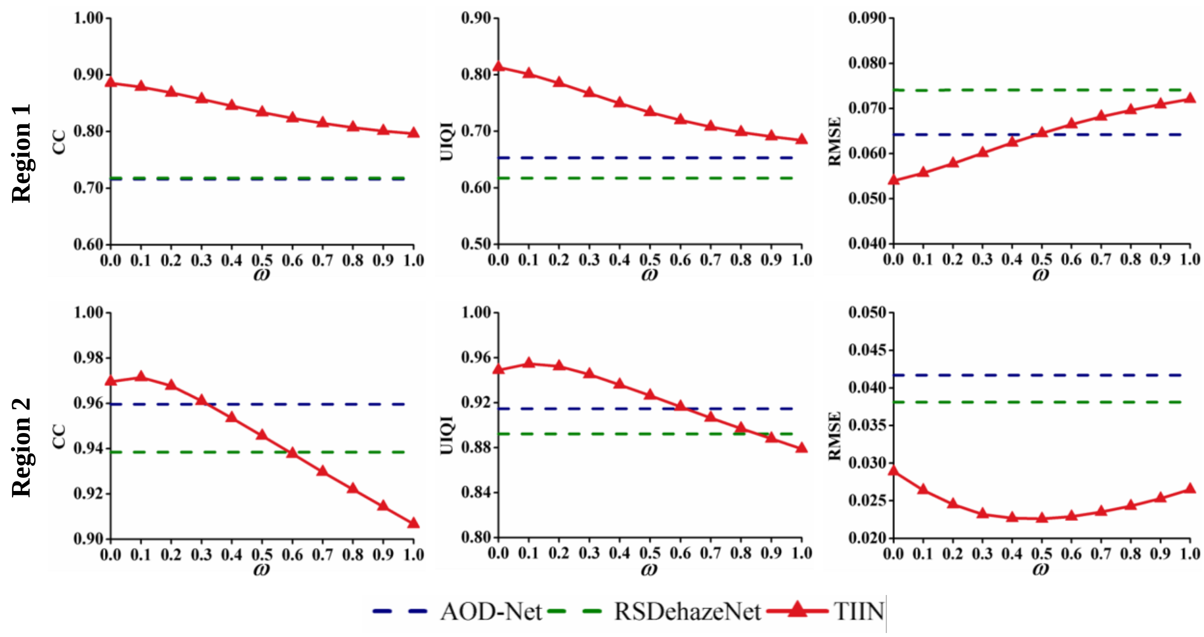

Fig. 10. Accuracy of TIIN under different haze contamination in temporally neighboring images. Note that the AOD-Net and RSDehazeNet methods do not use any temporal neighbors and, thus, their accuracies are just shown as dotted line for benchmark.

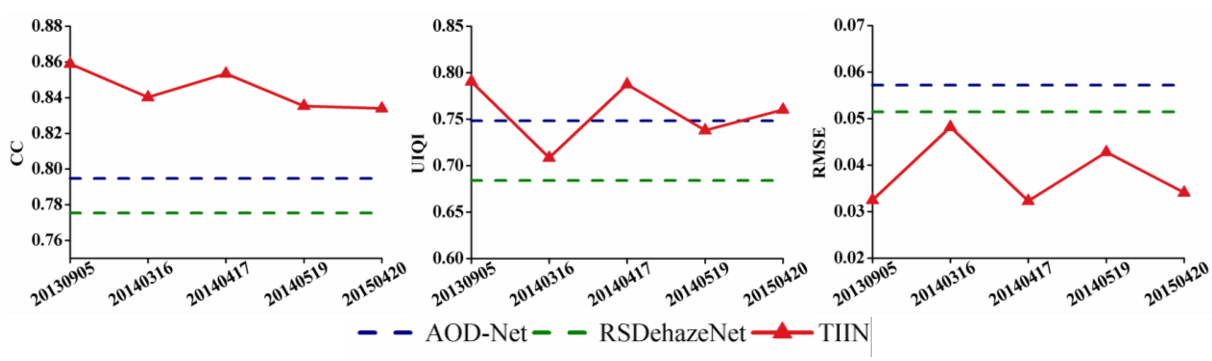

Fig. 11. Accuracy of TIIN under temporally neighboring images with different acquisition times.

There are usually abundant temporally neighboring images, due to the regular revisit capability of satellite sensors. However, useful spatial distribution prior may be limited due to land cover changes caused by a long-time interval. In this section, the influence of temporal distance between the hazy image and neighboring image was investigated. Specifically, a 300×300 pixels haze-clear image nearby the images in Case 3 was simulated with haze (ω=1). The haze-clear image was acquired on July 19, 2013. Another five haze-clear Landsat-8 OLI images, which were used as temporal neighbors, were acquired on September 5, 2013, March 16, 2014, April 17, 2014, May 19, 2014, and April 20, 2015. The accuracies based on the use of different temporally neighboring images are shown in Fig. 11. The accuracy of the proposed solution is greater than for AOD-Net and RSDehazeNet for almost all cases. Moreover, the accuracy of the proposed solution fluctuates and decreases in general as the time interval increases because of the decreased reliability of the spatial distribution prior in the temporal neighbors. However, the proposed solution can still produce more reliable dehaze results than AOD-Net and RSDehazeNet in the cases investigated here.

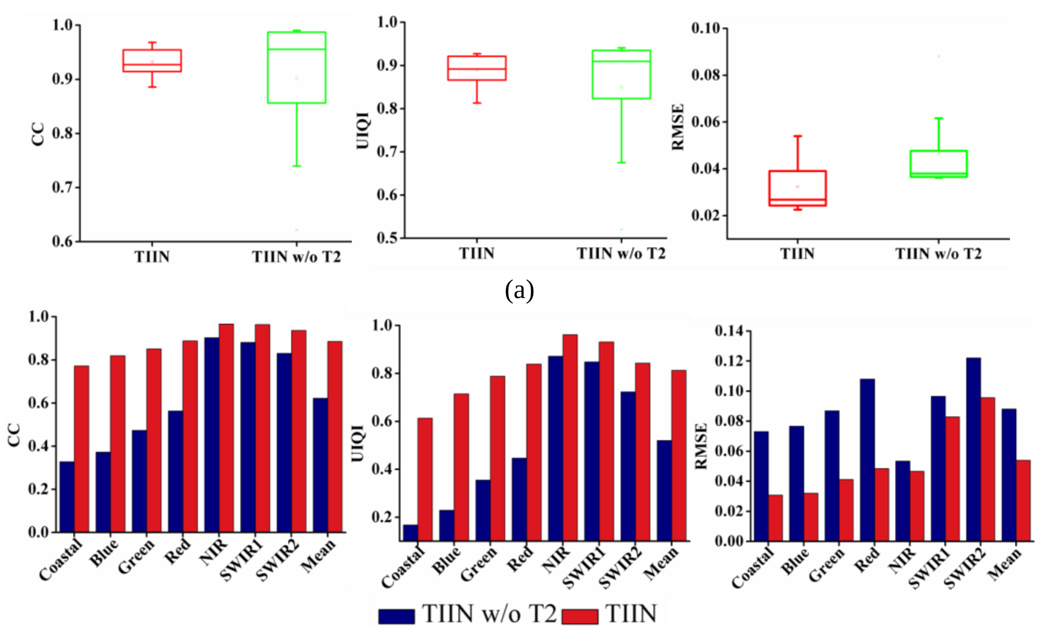

Fig. 12 Ablation study of temporally neighboring images (Region 1 as an example). (a) Box-plot of different haze levels. (b) Quantitative assessment for heavy haze (ω=1).

3.8. Ablation study

3.8.1. Ablation of different blocks in TIIN

Table 5. Ablation study for the three blocks (Region 1 as an example; w/o means without)

	CC	UIQI	RMSE
TIIN w/o TII	0.7594	0.6538	0.0789
TIIN w/o MSA	0.7608	0.6470	0.0785
TIIN w/o GroupConv	0.7806	0.6857	0.0726
TIIN	**0.8857**	**0.8133**	**0.0540**

To validate the effectiveness of the proposed TIIN architecture, several ablation studies were performed based on the simulated hazy image in Region 1. Specifically, the TII block, MSA block and group convolution block were considered. The results are displayed in Table 5, where the greatest accuracy in each case is marked in bold. The results indicate that the TII block, MSA block, and group convolution block are all effective for haze removal.

3.8.2. Ablation of temporally neighboring images

An ablation study of temporally neighboring images was carried out to validate the practicability of the proposed solution using the simulated hazy image in Region 1. Specifically, for comparison, a network named TIIN w/o T2, was considered. Its architecture is the same as for the original TIIN, and the only difference is that the former was trained with an absence of temporal neighbors. Different haze levels were employed for Region 1. The box-plot of dehazing accuracy under different haze levels is displayed in Fig. 12(a). The results indicate that the proposed solution is more stable for different haze levels than TIIN w/o T2. Moreover, quantitative assessment for the case of heavy haze contamination (ω=1) is presented in Fig. 12(b). It is obvious that the proposed TIIN method outperforms the method without utilizing temporally neighboring images.

3.9. Experiments on real haze images

Two large regions (Regions 7 and 8) with real haze contamination were applied to evaluate the applicability of the proposed dehaze method. The dehaze results for the two regions are displayed in Fig. 13. It can be observed that AOD-Net not only produces results with remaining haze, but also leads to apparent color distortion. Conversely, both RSDehazeNet and TIIN can alleviate the haze contamination more satisfactorily. However, focusing on the enlarged sub-regions of the coastal band, there remains noticeable haze in the RSDehazeNet results. Haze removal for Region 8 is more challenging. This is because this region is dominated by buildings with much more sophisticated spatial heterogeneity. Despite this, the proposed solution can still produce visually more pleasant results by taking the temporal information into account to deal with the spatial heterogeneity. Fig. 14 shows the dehaze results of the RGB bands of Cho et al. (2018), ACRM (Xu et al., 2014) and CR-NAPCT (Xu et al., 2019). Compared with the three benchmark methods, the proposed solution can simultaneously preserve the color and remove the hazy more satisfactorily.

To fully evaluate the applicability of the proposed TIIN method on regions spatially far from the training region, a water region (Region 9) located at Shanghai, China with real haze was considered. The haze removal results of different methods are shown in Fig. 15. Apparently, the proposed TIIN method can remove the haze more satisfactorily than the other methods.

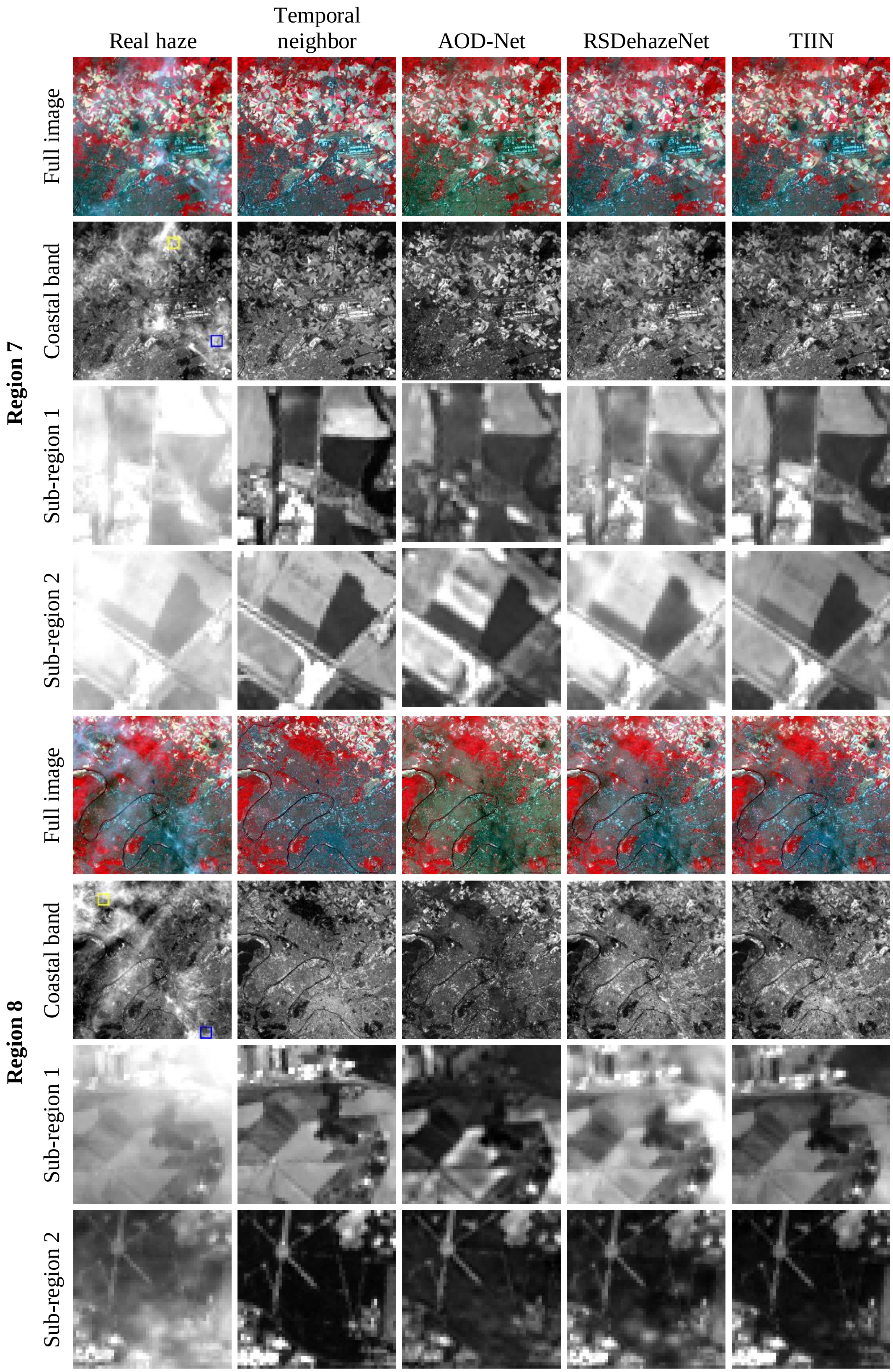
Real haze
Temporal neighbor
AOD-Net
RSDehazeNet
TIIN
Region 7
Full image
Coastal band
Sub-region 1
Sub-region 2
Region 8
Full image
Coastal band
Sub-region 1
Sub-region 2

Fig. 13. Results of real haze images in Regions 7 (800×800 pixels) and 8 (1000×1000 pixels) (pseudo-color: NIR, red and green as RGB; Sub-regions 1 and 2 are the yellow and blue rectangle regions in the coastal band).

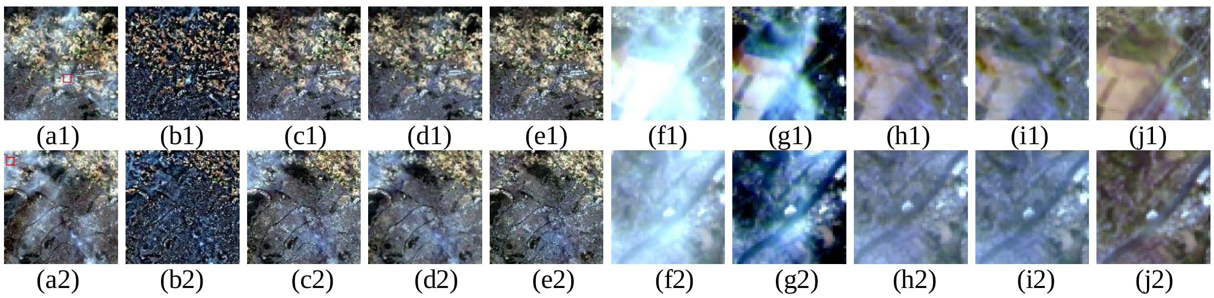

Fig. 14. Results of the model-based methods for the real haze images in Regions 7 (800×800 pixels; line 1) and 8 (1000×1000 pixels; line 2) (true-color). (a) Real haze image. (b) Cho et al. (2018). (c) ACRM. (d) CR-NAPCT. (e) TIIN. (f)-(j) are the corresponding zoom regions of the red box marked in (a).

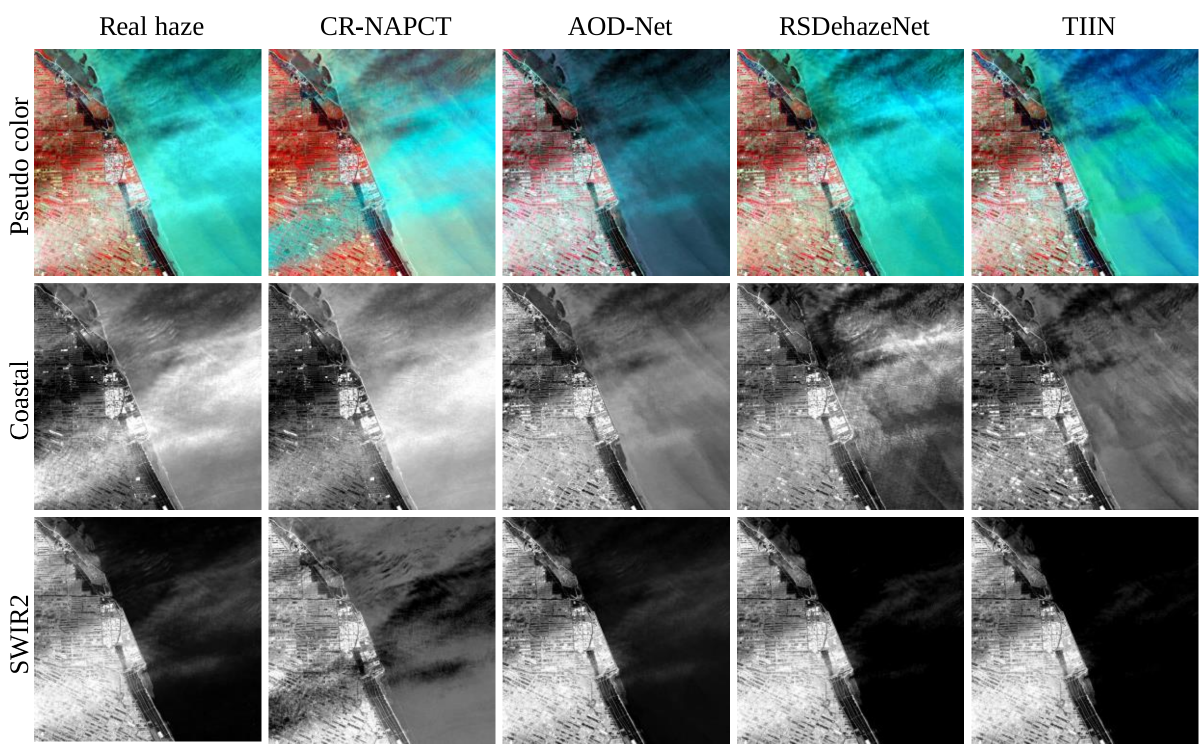

Fig. 15. Results of the real haze image covering a water region in Region 9 (900×900 pixels) (pseudo-color: NIR, red and green as RGB).

Two larger regions (i.e., Regions 10 and 11) with a spatial size of 2000 ×2000 Landsat-8 OLI pixels were also used to examine the proposed solution. The results for the entire area are displayed in Fig. 16. It is seen clearly that the proposed TIIN can remove the haze satisfactorily for the two larger regions.

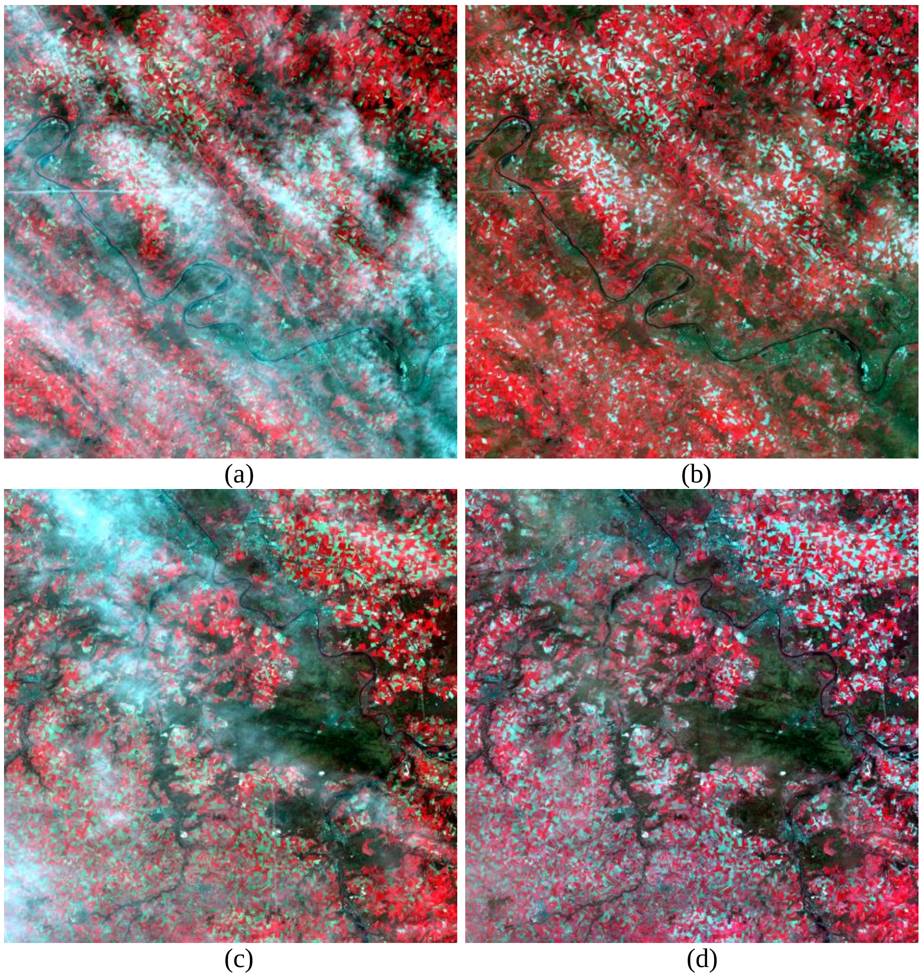

Fig. 16. Haze removal results for two larger regions (2000×2000 pixels). (a) Hazy image of Region 10. (b) TIIN prediction of Region 10. (c) Hazy image of Region 11. (d) TIIN prediction of Region 11.

4. Discussion

4.1. Rationale for utilizing temporal neighbors

Temporally neighboring images provided by the regular revisit of satellite sensors contain sufficient useful information, which can provide a spatial distribution prior to guide haze removal. It is noticed that the spatial distribution prior in the temporal neighbors is usually not the same as for the target hazy images due to land cover changes. However, the proportion of land cover changes is generally small, as most changes (e.g., in terms of hue of images) are driven by vegetation phenology and the condition of data acquisition. Thus, the spatial distribution in the temporal neighbors is undoubtedly a useful to guide for the haze removal process, reducing the uncertainty in this ill-posed problem.

In this research, we developed a TIIN architecture to incorporate temporal information for haze removal. In previous deep learning-based dehazing methods, the networks are trained to learn the relationship between the hazy images and haze-clear images directly. This would burden the training process since haze removal is an ill-posed issue. By incorporating temporally neighboring images, this burden can be alleviated in our method, as the temporal neighbors can provide auxiliary features for the input of the network. That is, the network is trained to learn the relationship between both hazy images and temporal neighbors with the prior. The use of the auxiliary variable (temporal neighbors) can, thus, reduce the uncertainty in the fitting process.

4.2. Applicability of the proposed TIIN method

In this research, for validation of the generalization ability of the proposed TIIN solution, several Landsat-8 OLI images with different land cover types or haze conditions were acquired in the experiments. It should be noted, however, that haze contamination exists widely in remote sensing images acquired by different optical sensors (such as MODIS, Sentinel-3 and -2 (Wang and Atkinson, 2018), Geofen series, etc.), and even in aerial images. For traditional haze removal methods, specific assumptions are made and parameters need to be determined manually for images acquired by different sensors, which is laborious and difficult to be generalized in various applications. This is not the case for TIIN, as the end-to-end learning strategy bypasses the complicated physical model of haze contamination, and can remove haze directly through parameter fitting in the network. As a result, TIIN is theoretically applicable for haze removal of images from different sensors, where the process is similar to that for the Landsat images investigated in this paper. The key requirement for TIIN is the need for temporally neighboring images, which may not be as straightforward to produce for aerial images.

4.3. Uncertainty in training data

The quality of training data is crucial to the reliability of neural network. In reality, however, it is impractical to collect hazy and haze-clear image pairs acquired at the same time for training. A practical strategy is to simulate haze contamination to create training data, as in this paper. However, the simulated haze is similar to but not equivalent to real haze. That is, real haze cannot be perfectly characterized by a simple mathematical model in most cases. In future research, it is necessary to develop a more comprehensive haze contamination strategy for greater approximation of real haze and, furthermore, to reduce the uncertainty introduced by training data.

With respect to the spatial content in the training data, it is critical to use images with an appropriate land cover distribution for training. That is, the spatial texture in the training images should be comprehensive to cover sufficient cases to deal with the haze images in the prediction stage of the network. The more representative the training images, the more generalized will be the network. It would be interesting to develop effective metrics (e.g., the similarity in semivariogram of longer wavelength bands between hazy images and training images) to identify useful training data from time-series data at the global scale.

4.4. Other choices of temporally neighboring images

In this research, temporally neighboring Landsat-8 OLI images were used, as the hazy images were also acquired by the same OLI sensor. It would be interesting to examine whether other choices of temporally neighboring images (e.g., images acquired by sensors that are different from those of the hazy images) are suitable for TIIN. This can be an important consideration when there are no effective temporal neighbors of the same sensors, due to cloud contamination in them. That is, the effective images of the same sensor may be temporally very far from the haze image and large land cover changes exist. On the other hand, many current satellite sensors provide temporally dense data including multispectral images (Gaofen series, Sentinel-2, etc.) and hyperspectral images (Gaofen-5, Zhuhai-1, etc.). These multi-source data can be temporally much closer to the haze image, even if they are acquired by different sensors. It is worthwhile to develop solutions to fill the gaps introduced by different platforms, and to take full advantage of these data and distill useful temporal information for TIIN.

4.5. Potential general solution for using temporal information in haze removal

Temporal neighbors can assist the proposed TIIN solution to produce much more reliable dehaze results than the four benchmark methods in the experiments. This demonstrated that the spatial distribution prior in temporally neighboring images is beneficial for the ill-posed problem of haze removal. In this research, a TII model was designed to extract this prior. It should be stressed, however, that the solution to incorporate temporal information is not limited to the specific CNN model proposed in this paper, but many other models might be potentially developed to enhance haze removal, such as, traditional physical models (atmospheric scattering model, etc.), and probabilistic models (maximum a *posteriori* probability, etc.). Thus, this paper provides an important guidance for considering temporal information in haze removal. In future research, it would be of great interest to develop the corresponding extended models and, furthermore, conduct a systematic comparison between them and identify the most advantageous type of methods for haze removal.

5. Conclusion

Haze contamination exists ubiquitously in remote sensing images. To remove the haze, we proposed to incorporate temporal information in this study. Following this general idea, the TIIN method was developed with parallel stacking layers and different attention modules to take full advantage of temporally neighboring images. As a result, the TIIN method can remove haze when the temporal neighbors contain inherent haze or land cover changes due to a long-time interval. To validate the proposed method, experiments on several groups of Landsat-8 OLI haze images were performed. The core conclusions are as follows.

1) Experiments on both simulated and real hazy images with various land cover types indicated that temporal information is beneficial to handle the ill-posed issue of haze

removal. The proposed TIIN method was found to be more accurate than two learning-based methods (i.e., AOD-Net and RSDehazeNet) and three model-based methods (i.e., the method in Cho et al. (2018), ACRM and CR-NAPCT).

2) TIIN was consistently more accurate than the benchmark methods under various haze levels, and the advantage was more obvious for heavy haze. Moreover, the TIIN-based dehaze results were also advantageous for further applications such as feature extraction.

3) The temporally neighboring images were still useful when they were also contaminated by haze. However, the haze in the temporal neighbors cannot be too heavy.

4) TIIN was still advantageous even when land cover changes exist between the hazy image and the temporal neighbor due to a long-time interval.

The code of the proposed TIIN method will be publicly available at https://qunmingwang.github.io.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grants 42171345 and 41971297 and Tongji University under Grant 02502350047. The authors would like to gratefully thank the editors and reviewers for their valuable comments, which have greatly improved this paper. The authors would also like to thank Dr. Meng Xu from Shenzhen University, China for providing the CR-NAPCT code and suggestions for implementation of CR-NAPCT.

References

Cai, B., Xiangmin, X., Kui, J., Chunmei, Q., & Dacheng, T. (2016). DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Trans Image Process, 25(11)*, 5187-5198.

Chavez, P.S. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment, 24*, 459-479.

Chen, Y., Weng, Q., Tang, L., Zhang, X., Bilal, M., & Li, Q. (2020). Thick Clouds Removing From Multitemporal Landsat Images Using Spatiotemporal Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing, Early Access*, 1-14.

Cho, Y., Jeong, J., & Kim, A. (2018). Model Assisted Multi-band Fusion for Single Image Enhancement and Applications to Robot Vision. *IEEE Robotics and Automation Letters, 3(4)*, 2822-2829.

Du, Y., Guindon, B., & Cihlar, J. (2002). Haze detection and removal in high resolution satellite image with wavelet analysis. *IEEE Transactions on Geoscience and Remote Sensing, 40(1)*, 210-217.

Fattal, R. (2008). Single image dehazing. *ACM Transactions on Graphics, 27(3)*, 1-9.

Guo, Q., Hu, H. M., and Li, B. (2019). Haze and thin cloud removal using elliptical boundary prior for remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11), 9114-9137.

Guo, J., Yang, J., Yue, H., Hou, C., & Li, K. (2020a). Landsat-8 OLI Multispectral Image Dehazing Based on Optimized Atmospheric Scattering Model. *IEEE Transactions on Geoscience and Remote Sensing*, Early Access, 1-11.

Guo, J., Yang, J., Yue, H., Tan, H., Hou, C., & Li, K. (2020b). RSDehazeNet: Dehazing Network With Channel Refinement for Multispectral Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, Early Access, 1-15.

He, K., Sun, J., & Tang, X. (2011). Single Image Haze Removal Using Dark Channel Prior. *IEEE Trans Pattern Anal Mach Intell, 33(12)*, 2341-2353.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). Las Vegas, NV, USA.

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Squeeze-and-Excitation Networks. *IEEE Trans Pattern Anal Mach Intell, 42(8)*, 2011-2023.

Irons, J., Dwyer, J., & Barsi, J. (2012). The next Landsat satellite: The Landsat Data Continuity Mission. *Remote Sensing of Environment*, 122, 11-21.

Ji, S., Dai, P., Lu, M., & Zhang, Y. (2021). Simultaneous Cloud Detection and Removal From Bitemporal Remote Sensing Images Using Cascade Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing, 59(1)*, 732-748.

Jiang, H., & Lu, N. (2018). Multi-Scale Residual Convolutional Neural Network for Haze Removal of Remote Sensing Images. *Remote Sensing, 10(6)*, 945.

Jiang, H., Lu, N., Yao, L., & Zhang, X. (2018). Single image dehazing for visible remote sensing based on tagged haze thickness maps. *Remote Sensing Letters, 9(7)*, 627-635.

Ju, M., Ding, C., Ren, W., Yang. Y., Zhang, D., & Guo, Y. (2021). IDE: Image Dehazing and Exposure Using an Enhanced Atmospheric Scattering Model. *IEEE Transactions on Image Processing*, 30, 2180-2192.

Kingma, D., & Ba, J., (2015). Adam: A method for stochastic optimization. In, *International Conference on Learning Representations (ICLR2015)* (pp. 1-15). San Diego, CA, USA.

Land, E. (1978). The Retinex Theory of Color Vision. *Scientific American, 237(6)*, 108-128.

LeCun, Y., Bengio, Y., & Hinton, G., (2015). Deep learning. *Nature*, 521, 436.

Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). AOD-Net: All-in-One Dehazing Network. In, *IEEE International Conference on Computer Vision* (pp. 4780-4788). Venice, Italy.

Li, J., Hu, Q., and Ai, M. (2018a). Haze and thin cloud removal via shpere model improved dark prior channel prior. *IEEE Geoscience and Remote Sensing Letters*, 16(3), 472-476.

Li, P., Tian, J., Tang, Y., Wang, G., & Wu, C. (2020a). Deep Retinex Network for Single Image Dehazing. *IEEE Trans Image Process, 30*, 1100-1115.

Li, R., Pan, J., Li, Z., & Tang, J. (2018b). Single Image Dehazing via Conditional Generative Adversarial Network. In, *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8202-8211). Salt Lake City, Utah, USA.

Li, J., Wu, Z., Zhang, J., Li, M., Mo, L., and Molinier, M. (2020b). Thin cloud removal in optical remote sensing images based on generative adversarial networks and physical model of cloud distortion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166, 373-389.

Li, W., Zou, Z., & Shi, Z., (2020c). Deep matting for cloud detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 58(12)*, 8490-8502.

Lv, H., Wang, Y., & Shen, Y. (2016). An empirical and radiative transfer model based algorithm to remove thin clouds in visible bands. *Remote Sensing of Environment, 176*, 183-195.

Makarau, A., Richter, R., Muller, R., & Reinartz, P. (2014). Haze Detection and Removal in Remotely Sensed Multispectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing, 52(9)*, 5895-5905.

Narasimhan, S.G., & Nayar, S.K. (2003). Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(6)*, 713–724.

Pan, X., Xie, F., Jiang, Z., & Yin, J. (2015). Haze Removal for a Single Remote Sensing Image Based on Deformed Haze Imaging Model. *IEEE Signal Processing Letters, 22(10)*, 1806-1810.

Qin, M., Xie, F., Li, W., Shi, Z., & Zhang, H. (2018). Dehazing for Multispectral Remote Sensing Images Based on a Convolutional Neural Network With the Residual Architecture. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(5)*, 1645-1655.

Qu, Y., Chen, Y., Huang, J., & Xie, Y. (2019). Enhanced Pix2pix Dehazing Network. In, *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 8152-8160). Long Beach, CA, USA.

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., & Yang, M.-H. (2016). Single Image Dehazing via Multi-scale Convolutional Neural Networks. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *European Conference on Computer Vision* (pp. 154-169). Amsterdam, The Netherlands: Springer International Publishing.

Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., & Yang, M. H. (2018). Gated fusion network for single image dehazing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3253-3261). Salt Lake City, UT, USA.

Ren, W., Pan, J., Zhang, H., Cao, X., & Yang, M., (2020). Single Image Dehazing via Multi-scale Convolutional Neural Networks with Holistic Edges. *International Journal of Computer Vision, 128(8)*, 240-259.

Shao, Z., Cai, J., Fu, P., Hu, L., & Liu, T. (2019). Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sensing of Environment, 235*, 111425.

Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li H., & Zhang, L., (2015). Missing Information Reconstruction of Remote Sensing Data: A Technical Review. *IEEE Geoscience and Remote Sensing Magazine*, *3(3)*, pp. 61-85.

Tan, R.T. (2008). Visibility in Bad Weather from a Single Image. In, *IEEE Conference on CVPR* (pp. 1-8). Anchorage, Alaska, USA

Tang, Y., Wang, Q., Zhang, K., & Atkinson, P.M. (2020). Quantifying the Effect of Registration Error on Spatio-Temporal Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13,* 487-503.

Wang, Q., & Atkinson, P.M., (2018). Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sensing of Environment, 204,* 31-42.

Wang, Q., Wang, L., Wei, C., Jin, Y., Li, Z., Tong, X., & Atkinson, P.M. (2021). Filling gaps in Landsat ETM+SLC-off images with Sentinel-2 images. *International Journal of Applied Earth Observation and Geoinformation, 101*, 102365.

Wang, Q., Tang, Y., Tong, X., & Atkinson, P.M. (2020). Virtual image pair-based spatio-temporal fusion. *Remote Sensing of Environment, 249*, 112009.

Wu, X., Shi, Z., & Zou, Z., (2021). A geographic information-driven method and a new large scale dataset for remote sensing cloud/snow detection. *ISPRS Journal of Photogrammetry and Remote Sensing, 174,* 87-104.

Xu, M., Jia, X., & Pickering, M. (2014). Automatic cloud removal for Landsat 8 OLI images using cirrus band. In, *IEEE Internatioanl Geoscience and Remote Sensing Symposium* (pp. 2511-2514). Quebec City, Canada .

Xu, M., Jia, X., Pickering, M., & Jia, S. (2019). Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform. *ISPRS Journal of Photogrammetry and Remote Sensing, 149,* 215-225.

Zhang, Y., Guindon, B. & Cihlar, J. (2002). An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. *Remote Sensing of Environment, 82,* 173-187.

Zhu, H., Peng, X., Chandrasekhar, V., Li, L., & Lim, J.-H. (2018). DehazeGAN: When Image Dehazing Meets Differential Programming. In, *27th International Joint Conference on Artificial Intelligence* (pp. 1234-1240). Palo Alto, USA.

Zhu, Q., Mai, J., & Shao, L. (2015). A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE Trans Image Process, 24(11),* 3522-3533.

Zhu, Z., Woodcock, C.E., (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment, 118,* 83-94.