



Improving semantic segmentation accuracy in thin cloud interference scenarios by mixing simulated cloud-covered samples

Haoyu Wang^{a,b,c,d,e,f}, Junli Li^{a,d,*}, Zhanfeng Shen^{b,g,**}, Zihan Zhang^h, Linze Baiⁱ, Ruifeng Li^j, Chenghu Zhou^k, Philippe De Maeyer^{a,c,e,f}, Tim Van de Voorde^{a,c,e,f}

^a National Key Laboratory of Ecological Security and Sustainable Development in Arid Region, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China

^b University of Chinese Academy of Sciences, Beijing 100049, China

^c Department of Geography, Ghent University, 9000 Ghent, Belgium

^d Key Laboratory of GIS & RS Application Xinjiang Uygur Autonomous Region, Urumqi 830011, China

^e Sino-Belgian Joint Laboratory of Geo-Information, 9000 Ghent, Belgium

^f Sino-Belgian Joint Laboratory of Geo-Information, Urumqi 830011, China

^g National Engineering Research Center for Geomatics, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China

^h Beijing Key Lab of Spatial Information Integration and 3S Application, Institute of Remote Sensing and Geographic Information System, School of Earth and Space Science, Peking University, Beijing 100871, China

ⁱ School of Earth Sciences, Zhejiang University, Hangzhou 310058, China

^j Department of Information Technology, Ghent University, 9000 Ghent, Belgium

^k State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Keywords:

Semantic segmentation
Thin cloud interference
Data augmentation
Atmospheric scattering model
Cloud simulation

ABSTRACT

Thin cloud interference presents a significant challenge for the semantic segmentation of optical satellite imagery, which directly degrades the model accuracy and causes difficulties in sample selection. This paper generated a dataset named *Populus euphratica* and *Tamarix chinensis* discrimination (PTD), containing both cloudless and thin cloud scenarios. Based on this PTD dataset, an enhanced Atmospheric Scattering Model with Nonlinear Optimization (ASM_NL) was proposed to simulate high-fidelity thin clouds by incorporating two vital nonlinear terms: the point spread function and the Perlin noise. Additionally, we adopt a strategy of mixing simulated thin cloud-covered images (STCI) into the training set at a certain proportion to improve the semantic segmentation accuracy in thin cloud-covered scenarios. The conclusions are as follows: 1) ASM_NL can simulate high-fidelity clouds at an average Jensen-Shannon distance of 0.0699. 2) When dealing with medium- and high-cloud density datasets, mixing STCI proved to be more effective than cloud removal in mitigating thin cloud interference, resulting in average macro F1 score improvements of 0.164 and 0.094, respectively. 3) The semantic segmentation accuracy improved significantly by mixing STCI with a minimal proportion of 1/60, demonstrating the activation of model transfer capabilities. This study provides a concise and efficient methodology for effectively mitigating thin cloud interference in deep learning-based optical satellite imagery analysis.

1. Introduction

Cloud cover presents challenges for various optical remote sensing applications, including resource exploration, forest management, and environmental monitoring (Sun et al., 2021). Research indicates that

more than 50 % of satellite images are partially or fully obscured by clouds (Mahajan, 2020). Cloud absorption and scattering can easily distort the land surface spectral characteristics, and the extent of this distortion depends on the thickness of the clouds (Hassan et al., 2022). Thick clouds block the reflectance from land surfaces, making it hard to

* Corresponding author at: National Key Laboratory of Ecological Security and Sustainable Development in Arid Region, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China.

** Corresponding author at: University of Chinese Academy of Sciences, Beijing 100049, China.

E-mail address: lijl@ms.xjb.ac.cn (J. Li).

<https://doi.org/10.1016/j.jag.2024.104087>

Received 25 June 2024; Received in revised form 26 July 2024; Accepted 8 August 2024

Available online 13 August 2024

1569-8432/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

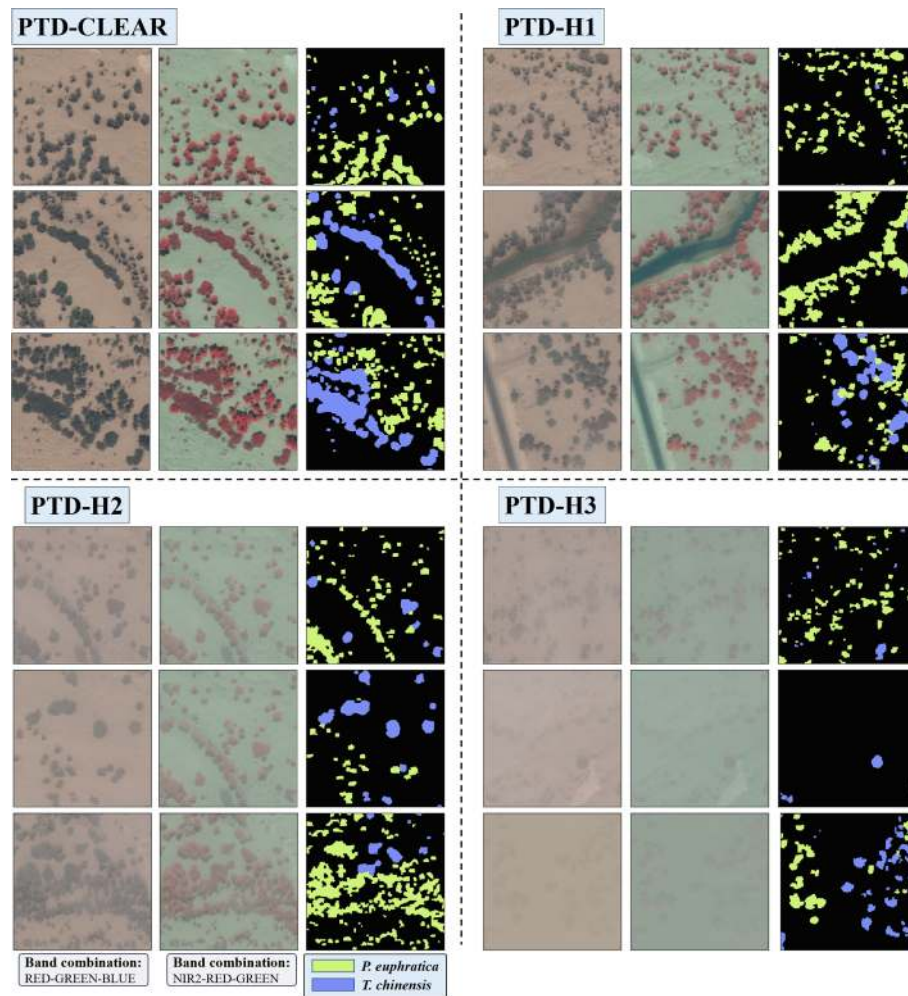


Fig. 1. Illustration of the PTD dataset: (a-d): PTD-CLEAR, PTD-H1, PTD-H2 and PTD-H3, respectively. For each subset, the first column contains the images with a Red-Green-Blue band combination, the second column shows the images with a NIR2-Red-Green band combination, and the third column stipulates the corresponding labels.

identify the ground objects. In comparison, thin clouds are transparent, and the contaminated ground features can be restored and mapped, while the mapping accuracy and reliability are easily affected by spectral distortion and contrast loss caused by cloud contamination (Li et al., 2020b).

Numerous methods have been proposed to improve the mapping accuracy under thin cloud interference conditions, which include traditional and deep-learning methods. The traditional methods are developed to reconstruct cloud-free images (Shen et al., 2014), which can be classified into multi-image and single-image methods. The multi-image methods restore images with thick clouds by replacing cloud pixels with other temporal images (Jia et al., 2023; Tseng et al., 2008; Zheng et al., 2023), and the single-image based methods usually enhance the cloud-contaminated images when the land features are visible (He et al., 2011; Singh and Komodakis, 2018; Xu et al., 2022). The traditional methods improve image interpretation by restoring image visual effects (Hassan et al., 2022), but they could not achieve the expected performance as the cloud removal process will produce spectral distortion (Pei et al., 2021). Hassan (Hassan et al., 2022) analyzed and identified two key reasons. Firstly, cloud removal enhances the existing image information without adding new information. Secondly, image restoration can hardly transfer the thin cloud images to their cloudless state, causing local spectral changes in the cloud areas of the satellite images. Therefore, the traditional methods mainly focus on restoring cloud-contaminated images to cloudless images, which

produces uncertainties in information extraction (Ma et al., 2022).

In contrast, the deep-learning methods selected the manually labelled training samples directly from the thin cloud images, and the cloud-contaminated land features can be delineated by models (Pei et al., 2021; Song et al., 2023). Labelling samples from thin cloud images is usually time-consuming and demands a high interpretation ability (Sun et al., 2021). In order to improve the robustness and performance, Zhang (Zhang et al., 2021) used simulated thin cloud samples for training, and these simulated samples were derived from the cloudless parts of the thin cloud image. As making the training samples with cloudless images is more effortless, many thin-cloud samples with different cloud scenarios can be generated through cloud simulation with several manually-labelled cloudless samples. However, deep learning models trained on simulated images often experience substantial performance degradation when applied to the real domain (Shao et al., 2020), primarily due to the domain gap (Zhou et al., 2023) between the simulated cloud images and the natural cloud images.

Numerous scholars have studied domain adaptation to minimize the domain gaps and improve the authenticity of simulated clouds (Atapour-Abarghouei and Breckon, 2018; Chen et al., 2018b; Li et al., 2023; Long et al., 2013; Yuan et al., 2022). Their research involves the application of adversarial learning strategies on feature space, image-to-image translation learning, and style transfer methods (Shao et al., 2020), which showed the effectiveness of producing high-fidelity simulated clouds. However, these methods are structurally complex and require

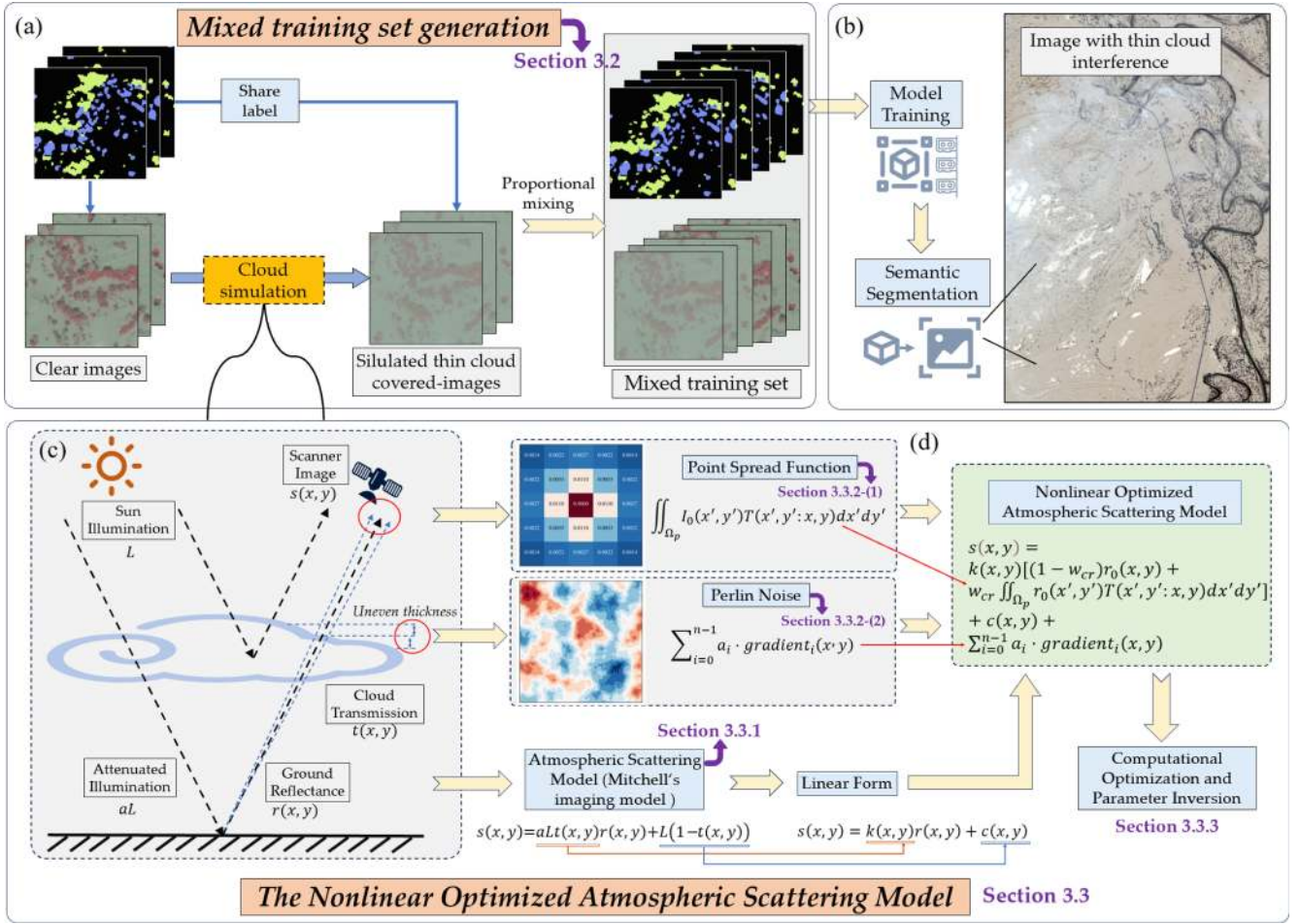


Fig. 2. Flowchart of the methodology. (a) Mixed training set generation; (b) model training and semantic segmentation on a thin cloud-interfered image; (c) illustration of the ASM; (d) the ASM_NL.

independent sample labelling and model training, presenting substantial challenges for deep learning-based remote sensing applications. Therefore, it is crucial to develop a cost-effective and high-fidelity strategy for simulating thin clouds and integrating simulated cloud generation with semantic segmentation.

This study proposes a strategy of mixing high-fidelity STCI in the training set to mitigate thin cloud interference. Firstly, we constructed a semantic segmentation dataset with category labelling for thin cloud and cloudless scenes. Secondly, we provided the mathematical description of the proposed strategy and developed a high-fidelity method for simulating thin clouds using ASM_NL. Thirdly, a series of experiments were conducted to verify the validity and reliability of the proposed strategy. Finally, we discussed some issues of concern.

2. Dataset

This study requires class-labeled datasets with cloudless and thin cloud scenarios. We reviewed relevant open-source datasets and divided them into two categories:

(1) Class-labeled datasets only with cloudless scenarios: This dataset directly selects cloudless images or uses cloud removal as a pre-processing step, resulting in no thin-cloud contamination in the published dataset. Examples of this type of dataset include GID (Tong et al., 2020) and WHLD (Cheng et al., 2023).

(2) Cloud removal (de-hazing) datasets: These types of datasets focus on the cloud removal effect, typically containing clear and thin cloud (fog)-contaminated image pairs but without category labels. Examples of this dataset include T-CLOUD (Ding et al., 2023) and RESIDE (Li et al.,

2019).

Therefore, we created a dataset called *Populus euphratica* (*P. euphratica*) and *Tamarix chinensis* (*T. chinensis*) discrimination (PTD) to support our research, which is the first class-labeled dataset with cloudless and thin cloud-interfered scenarios. All images come from the WorldView-2 satellite, equipped with eight spectral bands (CoastalBlue-427.3 nm, Blue-477.9 nm, Green-546.2 nm, Yellow-607.8 nm, Red-658.8 nm, RedEdge-723.7 nm, NIR1-832.5 nm and NIR2-908 nm) and have a spatial resolution of 0.5 m after pan-sharpening (Zhang, 2004) by the PCI Geomatics software.

All dataset images illustrate desert landscapes in the Tarim River Basin. They only contain three categories: Background (label 0), *P. euphratica* trees (label 1), and *T. chinensis* shrubs (label 2). We chose this dataset because it is hard to distinguish *P. euphratica* from *T. chinensis* due to their similar spectral characteristics. By visually examining the NIR2-Red-Green band combination, it is evident that *P. euphratica* displays a more vivid red color, intricate textures, and prominent shadows in its vicinity. In contrast, the *T. chinensis* shows a slightly lower red color, more uniform textures, and an absence of nearby shadows.

Based on the level of thin cloud interference, we divided the PTD into four subsets (Fig. 1): PTD-CLEAR, PTD-H1, PTD-H2, and PTD-H3. Then, we split them later into training, validation, and testing sets in a 3:1:1 ratio. There are 720, 336, 288, and 288 image pairs and labels in the PTD-CLEAR, PTD-H1, PTD-H2 and PTD-H3, respectively.

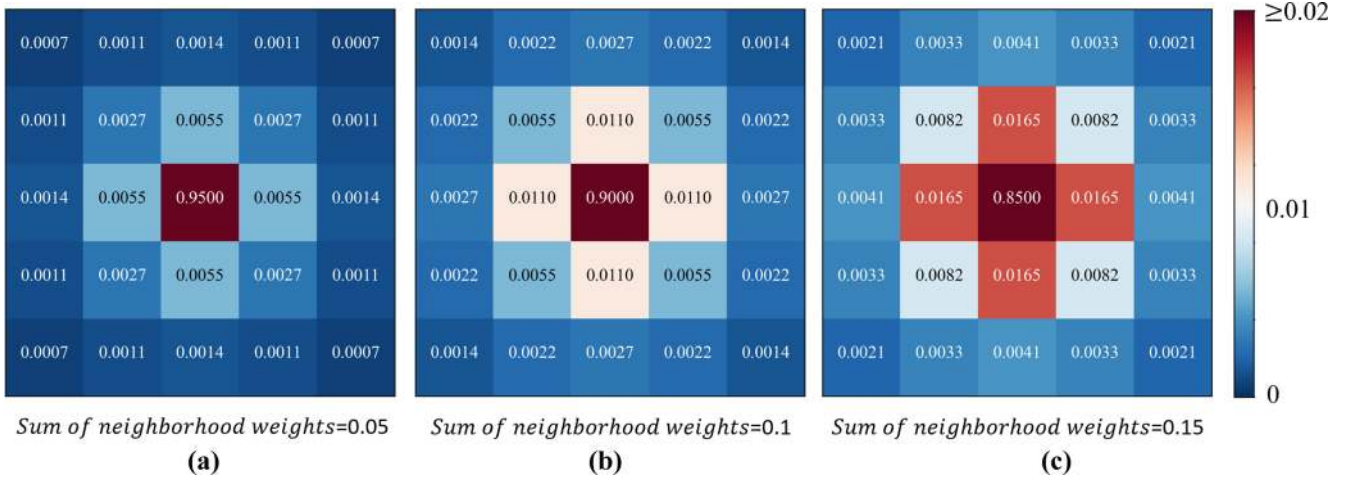


Fig. 3. The distribution of the radiation transmission coefficient: (a)-(c) demonstrates different kernel centers; all values are rounded to five decimal places.

3. Methodology

3.1. Outline of the methodology

This study introduces a novel strategy to improve the accuracy of semantic segmentation in thin-cloud scenarios by mixing simulated cloud-covered samples in the training set, which consists of two main parts. The first part in Section 3.2 is mixed training set generation, and the mixed proportion of the simulated and natural cloud samples is determined through a proportion experiment in Section 4.3 to balance the strategy performance and computational cost. As the fidelity of the simulated clouds is essential for the mapping accuracy of semantic segmentation models, the second part of Section 3.3 focuses on generating high-fidelity simulated thin clouds. Compared with the simplified Atmospheric Scattering Model (ASM), which is commonly used in cloud simulation (Lv et al., 2016), we proposed a nonlinear optimized ASM (ASM.NL) to incorporate two nonlinear components to minimize the distortions between thin clouds and ground objects. The method for parameter inversion is provided in Section 3.3.3. Fig. 2 shows the flowchart of the methodology.

3.2. Mixed training set generation

Acquiring sufficient cloud samples to include various thin cloud scenarios with limited remote sensing images is difficult, while cloud simulation with cloudless images can solve this problem. In order to generate the mixed training set, the simulated cloud samples with various densities and spatial distributions are added to the cloudless samples at a certain proportion. The following is the mathematical description of this process.

The thin cloud image x_{cloud} was simulated with the simulation function F , which takes a cloudless image $x_{\text{cloudless}}$ and a cloud density parameter α as inputs. The cloud density parameter α is randomly sampled from a user-defined distribution $\mathcal{D}$. The simulated cloud function F refers to any simulated cloud generation function such as the ASM:

$$x_{\text{cloud}} = F(x_{\text{cloudless}}, \alpha), \alpha \sim \mathcal{D} \quad (1)$$

The training dataset is defined as $(X_{\text{cloudless}}, Y_{\text{cloudless}})$, consists of the images x and the labels y :

$$(X_{\text{cloudless}}, Y_{\text{cloudless}}) = \{(x, y) | x \in X_{\text{cloudless}}, y \in Y_{\text{cloudless}}\} \quad (2)$$

The simulated thin cloud images are generated based on cloudless images with a probability of p . The probability parameter p is a user-defined parameter to determine the probability of cloud simulation occurrence:

$$(X_{\text{cloud}}, Y_{\text{cloud}}) = \{(F(x, \alpha), y) | x \in X_{\text{cloudless}}, y \in Y_{\text{cloudless}}, \text{ with probability } p\} \quad (3)$$

Finally, the mixed training set $(X_{\text{mixed}}, Y_{\text{mixed}})$ is the union of $(X_{\text{cloudless}}, Y_{\text{cloudless}})$ and $(X_{\text{cloud}}, Y_{\text{cloud}})$:

$$(X_{\text{mixed}}, Y_{\text{mixed}}) = (X_{\text{cloudless}}, Y_{\text{cloudless}}) \cup (X_{\text{cloud}}, Y_{\text{cloud}}) \quad (4)$$

3.3. The nonlinear optimized atmospheric scattering model

3.3.1. The linear atmospheric scattering model

The Atmospheric Scattering Model, known as Mitchell's imaging model (Mitchell et al., 1977), was proposed to simulate how light interacts with the atmosphere (Lv et al., 2016; Sakaridis et al., 2018), and it has been widely used in simulating thin clouds (Lv et al., 2016). This ASM model (Fig. 2 (c)) is expressed as follows:

$$s(x, y) = aLt(x, y)r(x, y) + L(1-t(x, y)) \leq L \quad (5)$$

where (x, y) represents the position of the pixel, $r(x, y)$ is the reflectance of the target object, $t(x, y)$ is the attenuation due to the clouds, L is the solar illumination, and a is the attenuation constant. The values of $r(x, y)$, $t(x, y)$ and a range from 0 to 1.

The ASM in the Eq. (5) requires the parameter inversion of a , L , and $t(x, y)$. As calculating these parameters is complicated, the ASM is usually simplified as a linear model in cloud simulation applications. The Eq. (5) could be restated as:

$$s(x, y) = k(x, y)r(x, y) + c(x, y) \quad (6)$$

where $c(x, y)$ represents the non-target object information component, such as the solar radiation reflected by the clouds and $k(x, y)$, which denotes the proportion of the target object information received by the sensor. Eq. (6) is a linear equation that only requires the inversion of two parameters, $k(x, y)$ and $c(x, y)$, to the simulated clouds.

The simplified linear ASM effectively simulates clouds on homogeneous surfaces (Sakaridis et al., 2018). However, as natural scenarios are usually heterogeneous, distortions are inevitable along the side of the simulated clouds (Hassan et al., 2022).

3.3.2. Nonlinear optimization

The natural thin clouds do not have purely linear effects on the light, and these nonlinear factors include (1) excessive acuteness due to neglecting scattering and (2) homogeneous thickness distribution due to neglecting spatial distribution inhomogeneity (Pei et al., 2021). Therefore, we added two nonlinear factors into the ASM model: the point spread function and the Perlin noise.

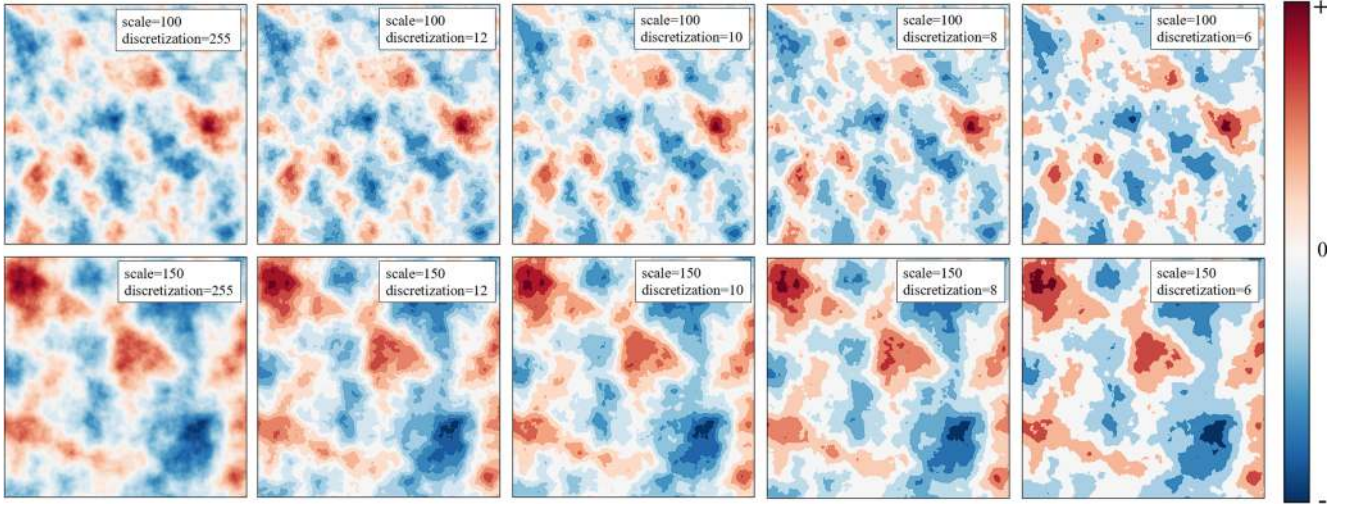


Fig. 4. Perlin noise with a different discretization and scale (marked in the upper right corner of each subplot); deeper red (blue) represents more substantial positive (negative) perturbations. The original ASM model generates a stochastic disturbance through the multiplication of the Perlin noise with the cloud thickness $d(x, y)$. However, the simplified ASM does not account for the variations in target and non-target information components caused by the cloud thickness. Therefore, we include the Perlin noise as an independent noise term in ASM.

(1) Point Spread Function

When a cloud is added to a cloudless image, the radiation of the neighboring pixels around the cloud will be distorted due to cloud scattering, which can be modeled with the point spread function (PSF) (Niu et al., 1997). In the PSF model, the target radiation $r(x, y)$ can be separated into the sum of the target radiation $r_0(x, y)$ and the cross-radiation $r_c(x, y)$, as shown in the Eq. (7).

$$r(x, y) = (1 - w_{cr})r_0(x, y) + w_{cr}r_c(x, y) \quad (7)$$

where w_{cr} is the cross-radiation's weight, the weights of the two components add up to 1. The cross-radiation $r_c(x, y)$ is computed as follows:

$$r_c(x, y) = \iint_{\Omega_p} r_0(x', y') T(x', y' : x, y) dx' dy' \quad (8)$$

where Ω_p denotes the surrounding pixels involved in the calculation; $T(x', y' : x, y)$ denotes the radiation transmission coefficient of (x', y') to (x, y) . The cross-radiation intensity is inversely proportional to the square of the distance to the target pixel, so the peripheral cross-radiation intensity decays rapidly. We tested various convolution sizes, including 3×3 , 5×5 , 7×7 , and 9×9 , to determine the optimal kernel size that balances the precision of the model with its computational cost. Our findings indicate that a 5×5 convolution kernel provides a satisfactory level of detail for capturing the spatial interactions relevant to calculating the radiation transmission coefficient without incurring the excessive computational load associated with larger kernels. Therefore, the radiation transmission coefficient was calculated using a discrete 5×5 convolution to improve computational efficiency. The weight assigned to the center of the kernel is $(1 - w_{cr})$, while the weights of other positions is w_{cr} . The weights are determined by the squared distance from each position to the kernel center. Fig. 3 illustrates the distribution of the radiation transmission coefficient in the convolution kernel with different kernel centers (0.95, 0.9 and 0.85).

(2) Perlin noise

Thin clouds display an uneven spatial distribution caused by airflow, which constitutes a significant nonlinear effect. This effect can be considered a type of spatial noise characterized by its continuous and non-uniform nature. To address this issue, we implemented Perlin noise (Perlin, 2002), a well-established method for representing the uneven distribution of cloud thickness (Zhang et al., 2021). The generation of Perlin noise (pn) can be represented by the following equation:

$$pn(x, y) = \sum_{i=0}^{n-1} a_i \cdot \text{gradient}_i(x, y) \quad (9)$$

where a_i is the random coefficient, $\text{gradient}_i(x, y)$ is the difference between the i^{th} vector and position (x, y) . Moreover, n denotes the vector number. Fig. 4 displays the generation of Perlin noise with different parameters. The degree of intensity and distribution fragmentation is determined by the discretization and scale parameters, which are indicated in the upper right corner of each subplot. Deeper red (blue) in Fig. 4 indicates more significant positive (negative) disturbances.

$$s(x, y) = k(x, y)r(x, y) + c(x, y) + pn(x, y) \quad (10)$$

The Perlin noise is used as a mask on the image during the calculation. However, this method is only suitable for scenes with low disturbances. Typically, the upper and lower limits of the disturbances ($\pm \text{discretization}/2$) should not exceed one-tenth of the scene average radiation to avoid noticeable color distortion. Therefore, by substituting Eq. (5) and Eq. (3) into Eq. (2), the final equation of ASM_NL will be the following:

$$s(x, y) = k(x, y) \left[(1 - w_{cr})r_0(x, y) + w_{cr} \iint_{\Omega_p} r_0(x', y') T(x', y' : x, y) dx' dy' \right] + c(x, y) + \sum_{i=0}^{n-1} a_i \cdot \text{gradient}_i(x, y) \quad (11)$$

3.3.3. Parameter inversion

To obtain the parameters of the simulated cloud model, we propose a statistical analysis method to minimize the gap between the actual and simulated clouds. The process begins by identifying and labeling sampling points in the thin-cloud and cloudless areas. The simulated cloud sampling points were generated based on the points identified in the cloudless areas. Finally, the parameter inversion was performed by minimizing the difference in the discrete probability distribution between the sampling points of real and simulated clouds. The specific strategy can be outlined as follows:

Define E^R , E^C and $E^S = E_\theta^{C \rightarrow S}$ as the sets of the natural cloud points, clear points and simulated cloud points, obtained based on the clear points and parameter space θ . $\{e_{ij}^R\}$, $\{e_{ij}^C\}$ and $\{e_{ij}^S\}$ represent the sets of the radiance values for the land cover category i and spectral band j (this also means inverting each band individually) in the natural, cloudless,

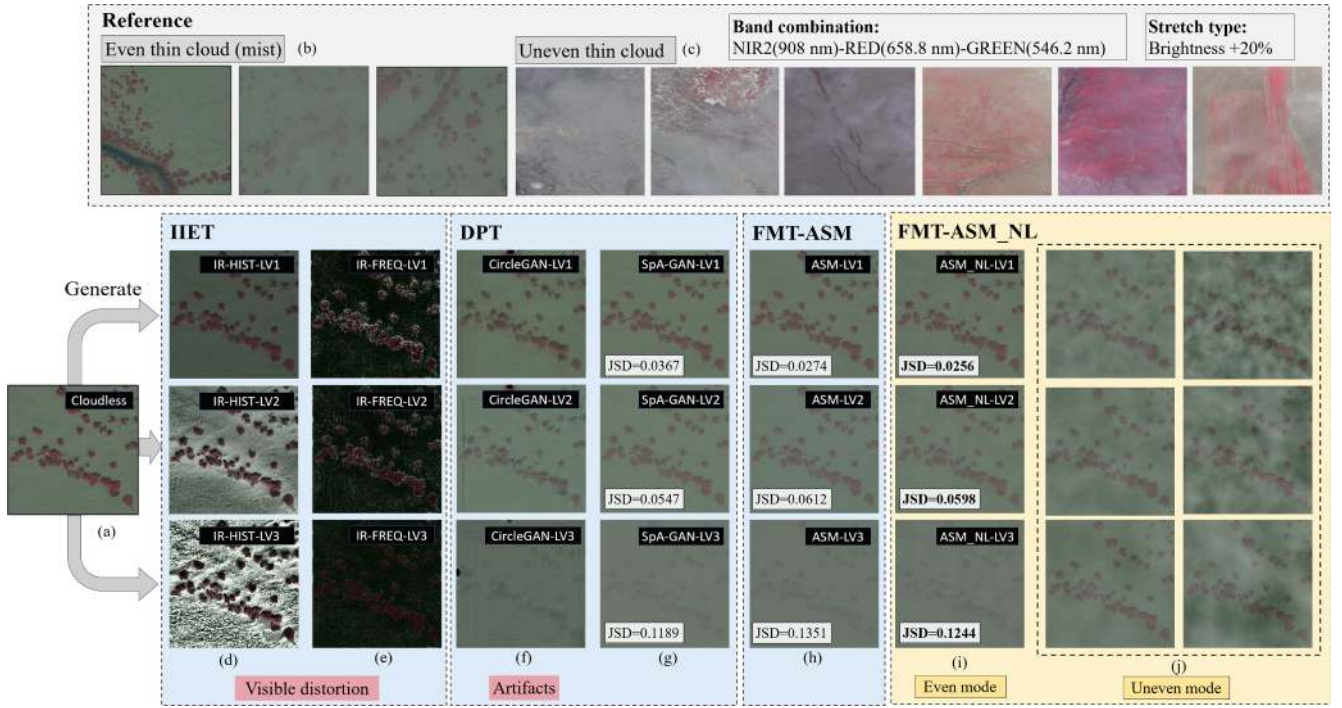


Fig. 5. Visual comparison of different cloud simulation methods: (a) the cloudless image for generating simulated clouds; (b) the natural cloud images including even thin cloud (mist) and uneven thin cloud as a reference; (d-j) simulated clouds generated using different methods. The top right corner of each image indicates the simulated cloud method and density. For the SpA-GAN, ASM and ASM_NL with even mode, the JSD between their discrete probability distribution and the natural cloud is shown in the lower left corner. Lower JSD indicates better simulation.

and simulated cloud points. The discrete probability distributions for $\{e_{ij}^R\}$, $\{e_{ij}^C\}$ and $\{e_{ij}^S\}$ are then calculated using the same approach. For example, the discrete probability distribution $P_{ij}^R(x)$ for $\{e_{ij}^R\}$ is obtained as follows:

$$P_{ij}^R(x) = \frac{1}{N_{ij}^R} \sum_{k=1}^{N_{ij}^R} \delta(x - e_{ij,k}^R) \quad (12)$$

where N_{ij}^R denotes the number of elements in $e_{ij,k}^R$. $\delta(x)$ is the Dirac delta function, which takes the value of 1 when $x = 0$ and 0 otherwise.

In order to minimize the differences between the probability distributions of natural and simulated clouds, it is essential to measure their dissimilarities quantitatively. This study uses the Jensen-Shannon distance (JSD) (Nielsen, 2019) to capture the differences in shape and position between the discrete probability distributions. The JSD between P_{ij}^R and P_{ij}^S can be calculated as follows:

$$JSD(P_{ij}^R || P_{ij}^S) = \sqrt{\frac{1}{2} D_{KL}(P_{ij}^R || M) + \frac{1}{2} D_{KL}(P_{ij}^S || M)} \quad (13)$$

where M is the average probability distribution of P_{ij}^R and P_{ij}^S . D_{KL} is the KL divergence (Kullback and Leibler, 1951), which could be calculated as follows:

$$D_{KL}(P || M) = \sum_{i=0}^{n-1} p_i \log \frac{p_i}{m_i} \quad (14)$$

where p_i and m_i represent the probabilities of P and M , respectively. n represents the number of the image grey levels.

Afterward, the parameters of each band are independently calculated (in this study), and the objective of the parameter solution for band j is:

$$\min_{\theta_j} = \frac{\sum_{i=1}^{n_{class}} JSD(P_{ij}^R || P_{ij}^S)}{n_{class}} \quad (15)$$

where n_{class} is the number of feature classes involved in the calculation.

However, the Perlin noise is not involved in the parameter inversion process as a noise component. Therefore, based on Eq. (11), three parameters need to be inverted: k , c and w_{cr} , with value ranges of $[0, 1]$, $[0, n-1]$ and $[0, 1]$ respectively. We divide the value ranges into 0.01, 0.5, and 0.01, thereby ensuring full parameter space coverage. As the image data type is an 8-bit unsigned integer, the change of the output value should be less than or equal to 1 when any parameter of k , w_{cr} , c is adjusted with its minimum discrete interval. However, searching for all three parameters would have an impractical computational cost, with a search upper limit of 5.12×10^6 . Because the intensity of the cross-radiation in the thin clouds is relatively low, we assume that its shift effect on the discrete probability distribution is smaller than the grey-level quantization. Therefore, we initially set w_{cr} to 0 and focused solely on the inversion of the parameters k and c , and then perform a separate inversion for w_{cr} . The maximum limit for the optimized search is 5.13×10^4 , which poses a manageable computational load.

3.4. Accuracy evaluation

The evaluation indicators include the macro precision (MP), macro recall (MR) and macro F1 score (MF). These indicators represent the arithmetic average of precision, recall and F1 score across all classes, respectively. The calculation formulas are as follows:

$$MP = \frac{\sum_{i=1}^k Precision_i}{k}, Precision = \frac{TP}{TP + FP} \quad (16)$$

$$MR = \frac{\sum_{i=1}^k Recall_i}{k}, Recall = \frac{TP}{TP + FN} \quad (17)$$

$$MF = \frac{\sum_{i=1}^k F1_i}{k}, F1 = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (18)$$

where k represents the number of the classes, which is 3 for the PTD

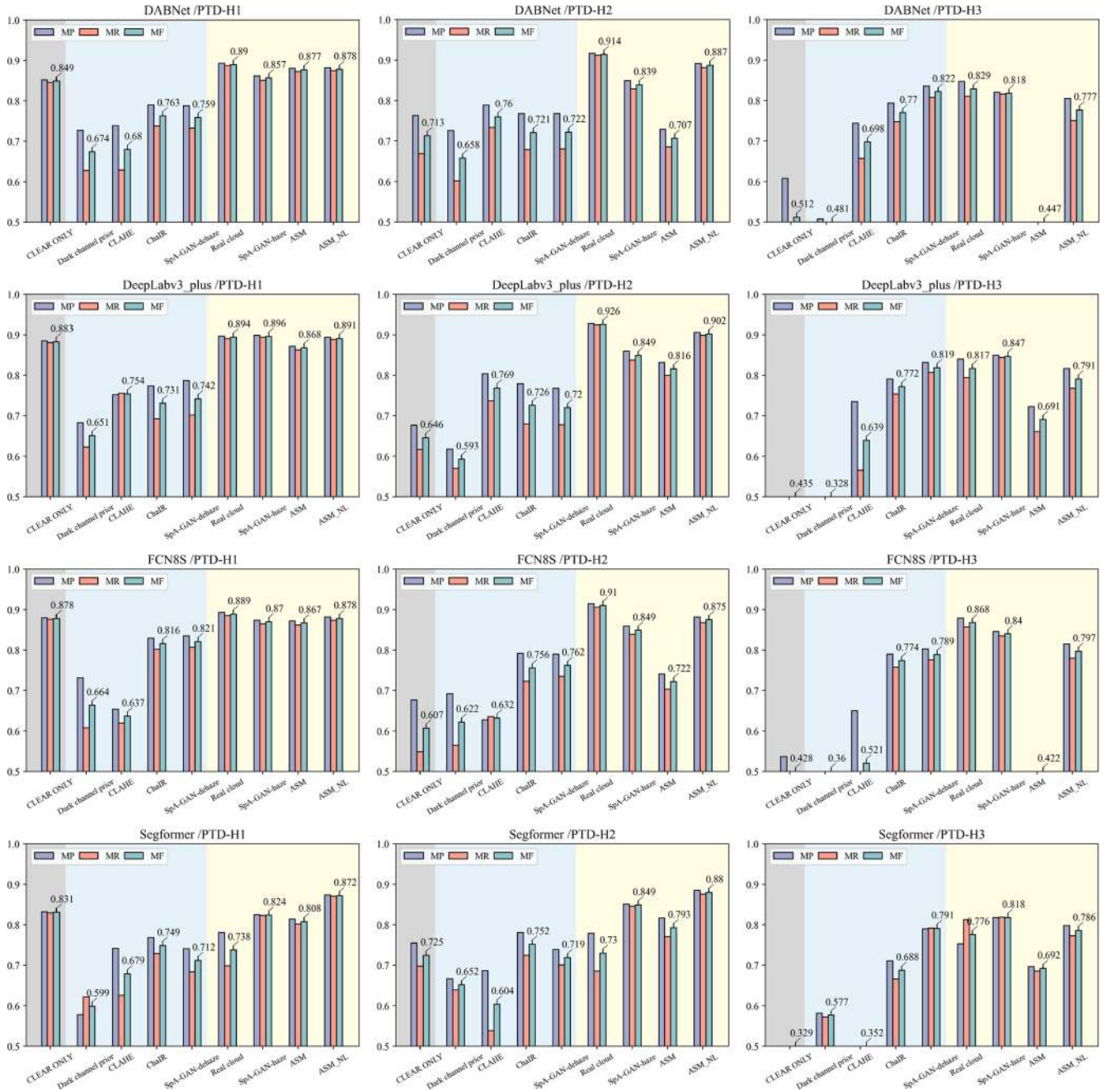


Fig. 6. Accuracy comparison results of different methods on the PTD dataset. Each row corresponds to different models. Each column from left to right corresponds to PTD-H1 ~ H3. The gray background is the control method. The blue background is the cloud removal method. The gold background is the STCI mixing method. The MF value for each accuracy evaluation experiment is labeled on the graph.

dataset, including desert, *P. euphratica* and *T. Chinensis*. *TP* represents the number of true positive samples, *FP* represents the number of false positive samples, and *FN* represents the number of false negative samples.

4. Results

The experimental section is organized into three subsections. Section 4.1 assesses the disparities between the simulated clouds and the real clouds. Section 4.2 comprehensively evaluates the performance of adding STCI to the semantic segmentation training set and compares it with the cloud removal strategy. Section 4.3 investigates the influence of

the proportion of STCI in the training set.

4.1. Cloud simulation realness evaluation

We conducted a comparative experiment to evaluate the commonly used cloud simulation algorithms. These algorithms are divided into three categories: inverse image enhancement techniques (IET), deep learning-based techniques (DPT), and physical modelling techniques (FMT):

IET: Image enhancement techniques (IET) are employed to enhance the image features and eliminate noise without relying on any physical image degradation model (Singh et al., 2022). As clouds are known to

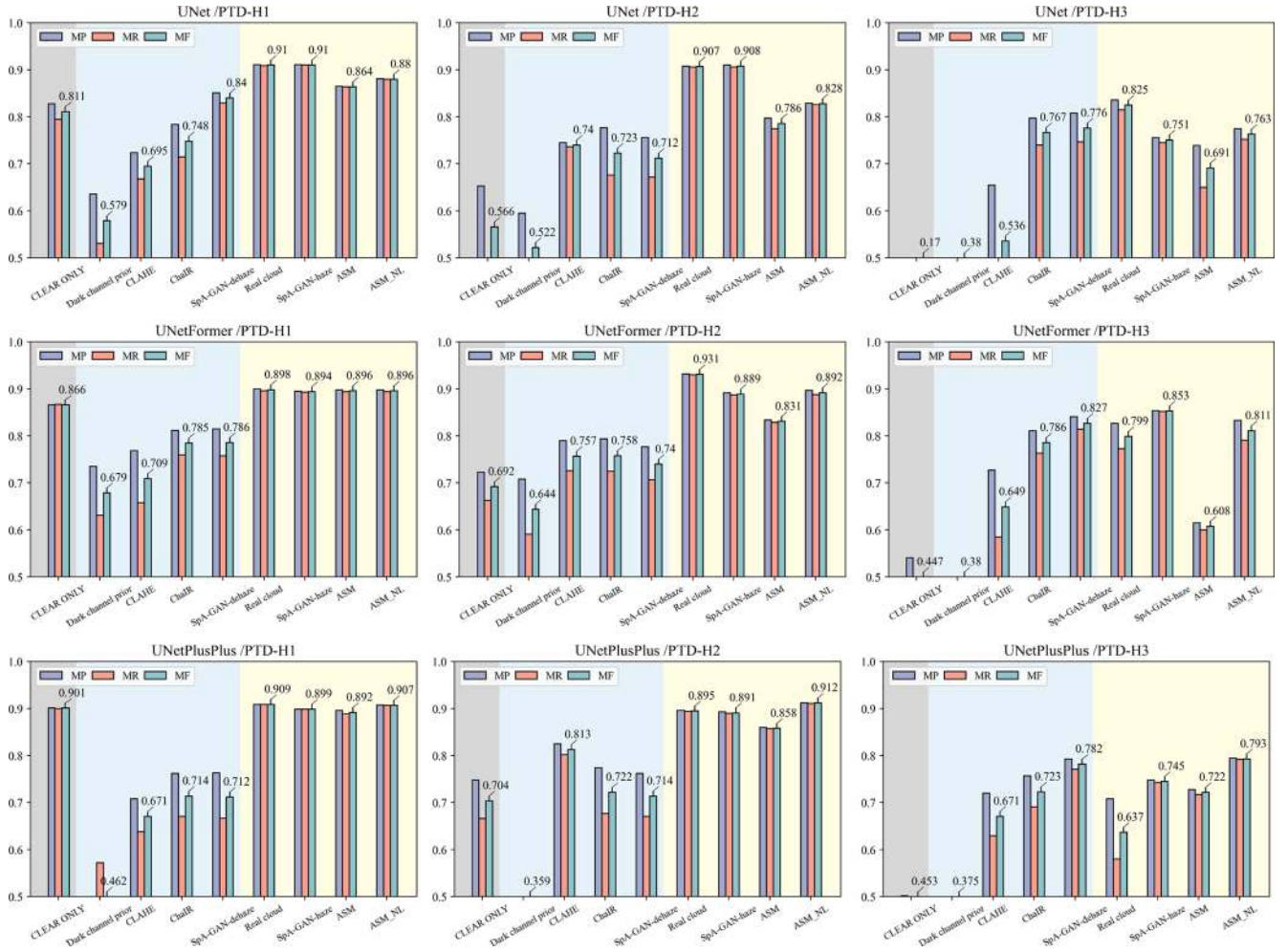


Fig. 6. (continued).

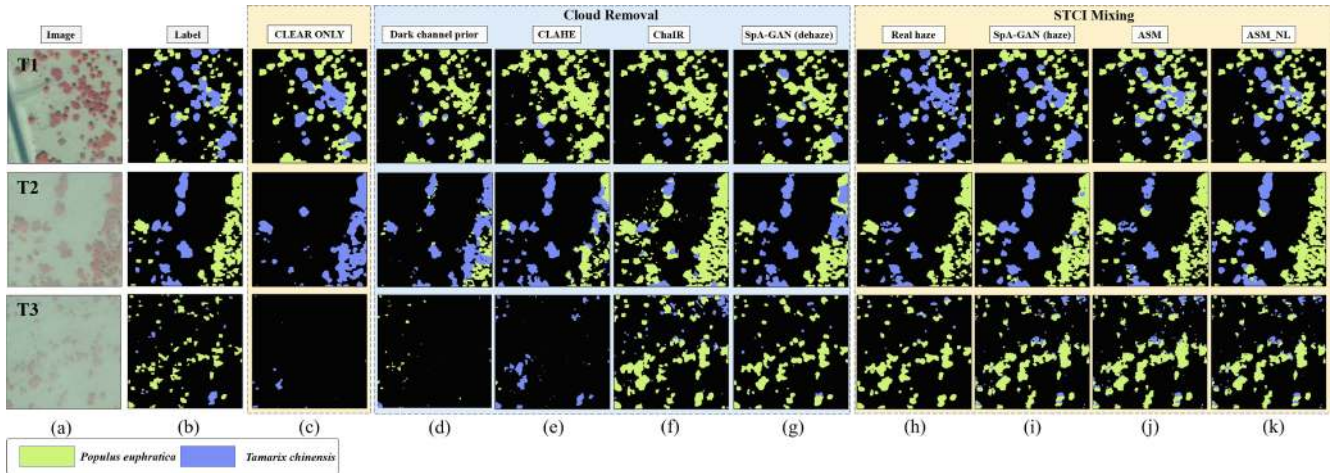


Fig. 7. Visualization of semantic segmentation results of UNet, rows 1–3 corresponds to T1, T2 and T3 from PTD-H1, H2 and H3, respectively. For each column: (a) the original images with all sizes at 256×256 pixels; (b) labels manually created based on visual interpretation; (c–k) visualization of prediction results based on different methods.

degrade image quality, IET is usually applied to mitigate their influence, with some techniques in this category implementing inverse transformations suitable for cloud simulation. For this study, we selected the inverse histogram equalization (IR-HIST) method (Aboshosha et al., 2019) and the inverse frequency transformation (IR-FREQ) method (RL

Easton Jr, 2010) for comparison.

DPT: The DPT typically utilizes a Generative Adversarial Network (GAN) structure to transfer images from the cloudless to the haze domain, producing impressive outcomes. The GAN consists of the Generator and Discriminator blocks. The Generator creates the

Table 1

Results of the simulated cloud proportion experiments: (a) the proportion ranging from 0 to 1, with intervals of 0.1; (b) the proportion ranging from 0 to 0.1, with intervals of 0.167 (1/60); (c) the proportion ranging from 0.9 to 1, with intervals of 0.167 (1/60). The accuracy indicators represent the average of 50 models in each independent sub-experiment. Each test set's highest (lowest) accuracy indicator is bolded with red (blue).

Test Set	Indicator	The Proportion of the STCI.										
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
PTD-CLEAR	MP	0.860	0.872	0.871	0.865	0.872	0.878	0.880	0.863	0.885	0.886	0.568
	MR	0.835	0.846	0.843	0.843	0.845	0.852	0.855	0.848	0.863	0.864	0.606
	MF	0.846	0.859	0.857	0.853	0.858	0.865	0.867	0.854	0.874	0.875	0.586
PTD-H2	MP	0.584	0.800	0.796	0.792	0.782	0.784	0.769	0.763	0.794	0.799	0.757
	MR	0.567	0.744	0.741	0.751	0.732	0.740	0.726	0.736	0.751	0.758	0.703
	MF	0.570	0.771	0.767	0.770	0.755	0.761	0.747	0.748	0.771	0.778	0.729

(a)

Test Set	Indicator	The Proportion of the STCI.						
		0	0.017	0.033	0.050	0.067	0.083	0.1
PTD-CLEAR	MP	0.860	0.870	0.889	0.860	0.896	0.872	0.872
	MR	0.835	0.840	0.870	0.836	0.879	0.843	0.846
	MF	0.846	0.855	0.879	0.847	0.888	0.857	0.859
PTD-H2	MP	0.584	0.769	0.832	0.789	0.839	0.811	0.800
	MR	0.567	0.732	0.794	0.749	0.810	0.759	0.744
	MF	0.570	0.749	0.812	0.767	0.824	0.784	0.771

(b)

Test Set	Indicator	The Proportion of the STCI.						
		0	0.917	0.933	0.950	0.967	0.983	1
PTD-CLEAR	MP	0.886	0.901	0.88	0.886	0.883	0.849	0.568
	MR	0.864	0.888	0.857	0.864	0.88	0.822	0.606
	MF	0.875	0.894	0.868	0.875	0.877	0.835	0.586
PTD-H2	MP	0.799	0.844	0.796	0.807	0.833	0.771	0.757
	MR	0.758	0.826	0.762	0.767	0.812	0.739	0.703
	MF	0.778	0.835	0.779	0.786	0.822	0.755	0.729

(c)

simulated thin clouds that look realistic and have statistical similarities with natural cloud styles. The Discriminator learns by distinguishing between the real and the simulated thin clouds. This adversarial process continues until the simulated thin clouds are visually identical to real thin clouds, making it difficult for the Discriminator to differentiate between the natural and simulated images. State-of-the-art architectures, such as CircleGAN (Zhu et al., 2017) for the unpaired training and SpA-GAN (Emami et al., 2021) for the paired training, are used for comparison.

FMT: The FMT methods establish image degradation models based on the physical effects of the clouds. The ASM and the ASM_NL belong to the FMT category and are included in the comparison.

The images in Fig. 5(b) are used as a reference to visually compare different simulated clouds in Fig. 5(c-h). Each simulated cloud method generated clouds at different levels (LV1-LV3 correspond to PTD H1-H3, respectively), and the most realistic results in the parameter space were selected using a traversal strategy. The simulated clouds based on IET (Fig. 5 (c-e)) show noticeable visual differences compared to the reference natural clouds. On the other hand, the simulated clouds based on DPT and FMT (Fig. 5 (f-i)) display higher fidelity. The Circle-GAN-based simulated cloud shows noticeable artefacts due to using an unpaired training set. However, the simulated clouds generated by SpA-GAN, ASM, and ASM_NL are visually indistinguishable from the real clouds.

Furthermore, we examined their discrete probability distributions through statistical analysis. The quantitative results are presented in the lower-left corner of the sub-figure of Fig. 5, which shows the JSD between the discrete probability distributions of the simulated clouds (SpA-GAN, ASM and ASM_NL) and the natural clouds. The parameter inversion covers the entire parameter space, and the JSD values are the minimum of each inversion. The result shows that the JSD values of the ASM_NL are consistently lower than those of ASM, with differences amounting to 0.0256, 0.0598 and 0.1244 for the cloud densities LV1, LV2 and LV3, respectively. Additionally, we noticed that an increasing cloud density leads to higher JSD values between the simulated and

natural clouds, indicating that the inherent deviation between the simulated and the natural clouds is amplified as the cloud density increases.

4.2. STCI mixing effectiveness evaluation

This section compared the performance of various methods for mitigating interference from thin clouds. Ten methods were compared and divided into three groups. The first group contained only one method, CLEAR_ONLY, as a control, where the model was trained with PTD-CLEAR and no cloud removal measures were applied. The second group consists of four cloud removal methods: Dark channel prior (Kaiming et al., 2009), Contrast-limited adaptive histogram equalization (CLAHE) (Pizer et al., 1990), ChaIR (Cui and Knoll, 2023), and SpA-GAN-dehaze (Dash et al., 2024). The third group contained all simulated thin cloud-covered images (STCI) mixing methods. The natural thin-cloud images (NTCI) were derived from the training sets of PTD-H1 ~ H3, and the STCIs were generated based on SpA-GAN-haze, ASM, and ASM_NL. All these NTCI and STCI were randomly mixed with the PTD-CLEAR according to a 20 % additional ratio. The accuracy evaluation was performed on the test set of the PTD-H1 ~ H3.

Moreover, the commonly used CNN-based and Transformer-based semantic segmentation models, including UNet (Ronneberger et al., 2015), UNet++ (Zhou et al., 2018), UNetFormer (Wang et al., 2022), DeepLab-v3+ (Chen et al., 2018), DABNet (Li et al., 2020), FCN8S (Shelhamer et al., 2017), and Segformer (Xie et al., 2021) were involved in the comparison to demonstrate the generalizability of the proposed method. None of these models utilized pre-trained models to prevent interference from external data. All models involved in the comparison achieved the highest accuracy in 10 independent trainings with different random number seeds. To ensure consistency, the FocalLoss loss function (Lin et al., 2020) and the ADAM optimizer (Kingma and Ba, 2014) used the same set of hyperparameters.

The experimental results are displayed in Fig. 6, with different color

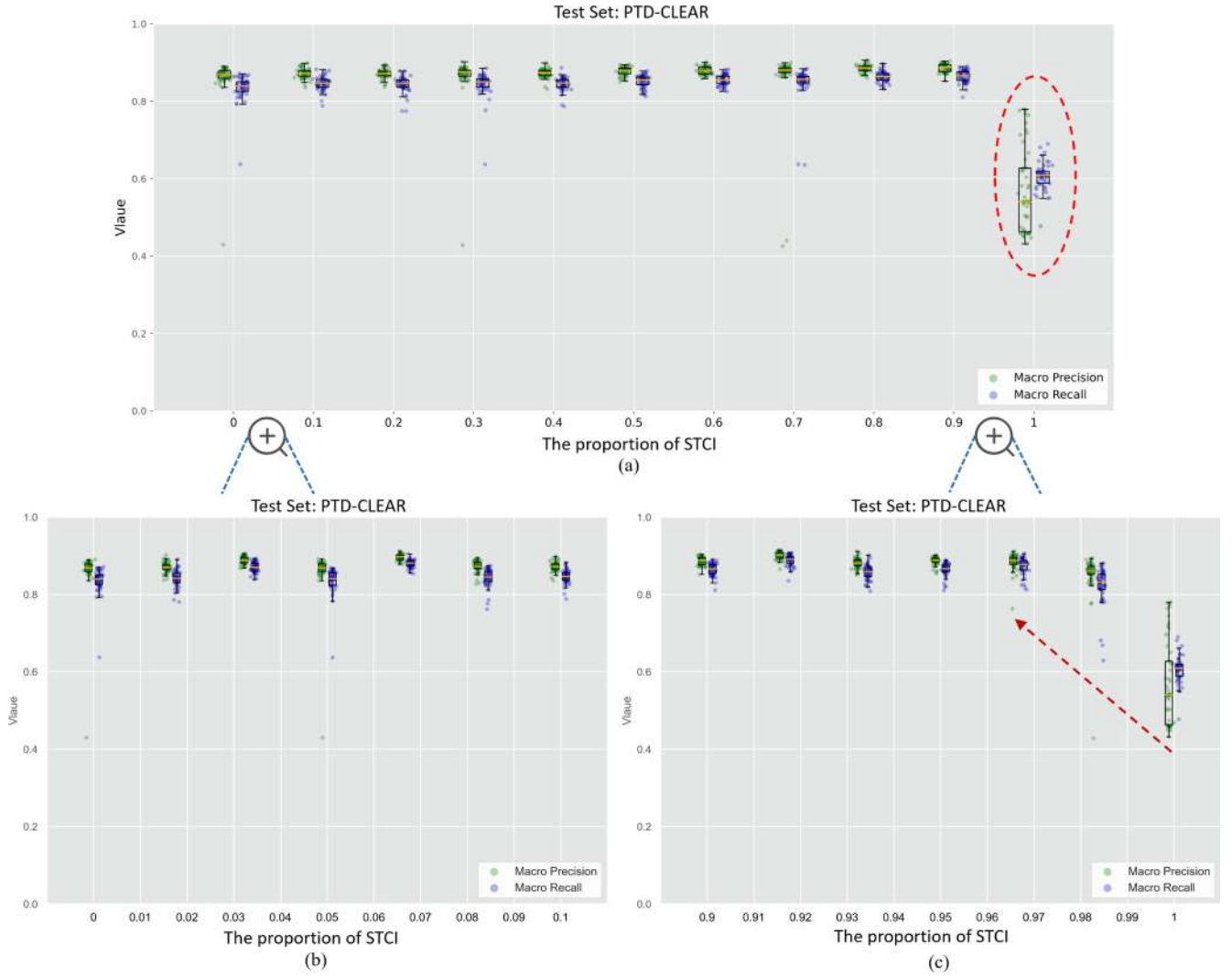


Fig. 8. Boxplots of proportion experiments on the test set of PTD-CLEAR: (a) the proportion ranging from 0 to 1, with intervals of 0.1; (b) the proportion ranging from 0 to 0.1, with intervals of 0.167 (1/60); (c) the proportion ranging from 0.9 to 1, with intervals of 0.167 (1/60).

backgrounds assigned to different types of methods for comparison: grey background for the control method, blue background for the cloud removal methods, and gold background for the STCI mixing methods.

The mean MF values of the STCI mixing methods on PTD-H1, H2, and H3 were 0.846, 0.832, and 0.711, respectively, which were better than those of the cloud removal methods (0.675, 0.668, and 0.617) and the control method (0.812, 0.643, and 0.373). The accuracy of all methods decreases as cloud density increases, demonstrating that the information loss caused by thin cloud interference cannot be completely mitigated. For PTD-H1, which has the lowest cloud density, both the control and STCI mixing methods have high accuracy.

In contrast, the cloud removal methods have significantly lower accuracy than other methods due to color distortion. Specifically, for PTD-H2 with medium cloud density, the average MF values of the STCI mixing methods are significantly higher than that of the control and cloud removal methods by 0.189 and 0.164, respectively. The DPT-based cloud removal methods (ChaIR, SpA-GAN-dehaze) perform better than the control method, while the overall accuracy is still lower than that of the STCI mixing methods.

As for T3, characterized by the highest cloud density, even experienced experts could barely identify *P. euphratica* and *T. chinensis* under a specific image stretching. At this stage, the highest MF of the control method is only 0.512. The average MF of the STCI hybrid methods declined to 0.711 but is still higher than that of the cloud removal

methods, which registered an average MF of 0.617.

The STCI mixing methods did not show a significant difference in accuracy on PTD-H1. The ASM_NL achieved an MF of 0.849, which closely matches the 0.8493 MF of SpA-GAN-haze, surpassing the 0.746 MF of ASM on PTD-H2. As for PTD-H3, the MF value of ASM on several models, such as DABNet, FCN8S, etc., dropped dramatically, and the corresponding models failed to converge correctly even after multiple independent trainings. The MF values of ASM_NL were significantly higher than those of ASM on all the models and were similar to those of SpA-GAN-haze. The NTCI mixing method, which serves as the control, had a higher MF than the STCI mixing methods on PTD-H2 and H3.

Three images, T1, T2, and T3 from PTD-H1, H2, and H3, respectively, were utilized to assess model performance. We performed semantic segmentation using UNet, followed by applying methods designed to mitigate the interference of thin clouds corresponding to Fig. 6. The results are shown in Fig. 7. For T1, which had the lowest cloud density, all methods could accurately map the vegetation and the desert. However, inaccuracies were observed in differentiating between *T. chinensis* and *P. euphratica*. For T2, which had the medium cloud density, *P. euphratica* and *T. chinensis* cannot be distinguishable with the CLEAR ONLY method. Therefore, the STCI mixing methods demonstrate more stable performance, with more precise vegetation contours and higher distinguishing accuracies between *P. euphratica* and *T. chinensis*. As for T3, which had the highest cloud density, the *P. euphratica* and

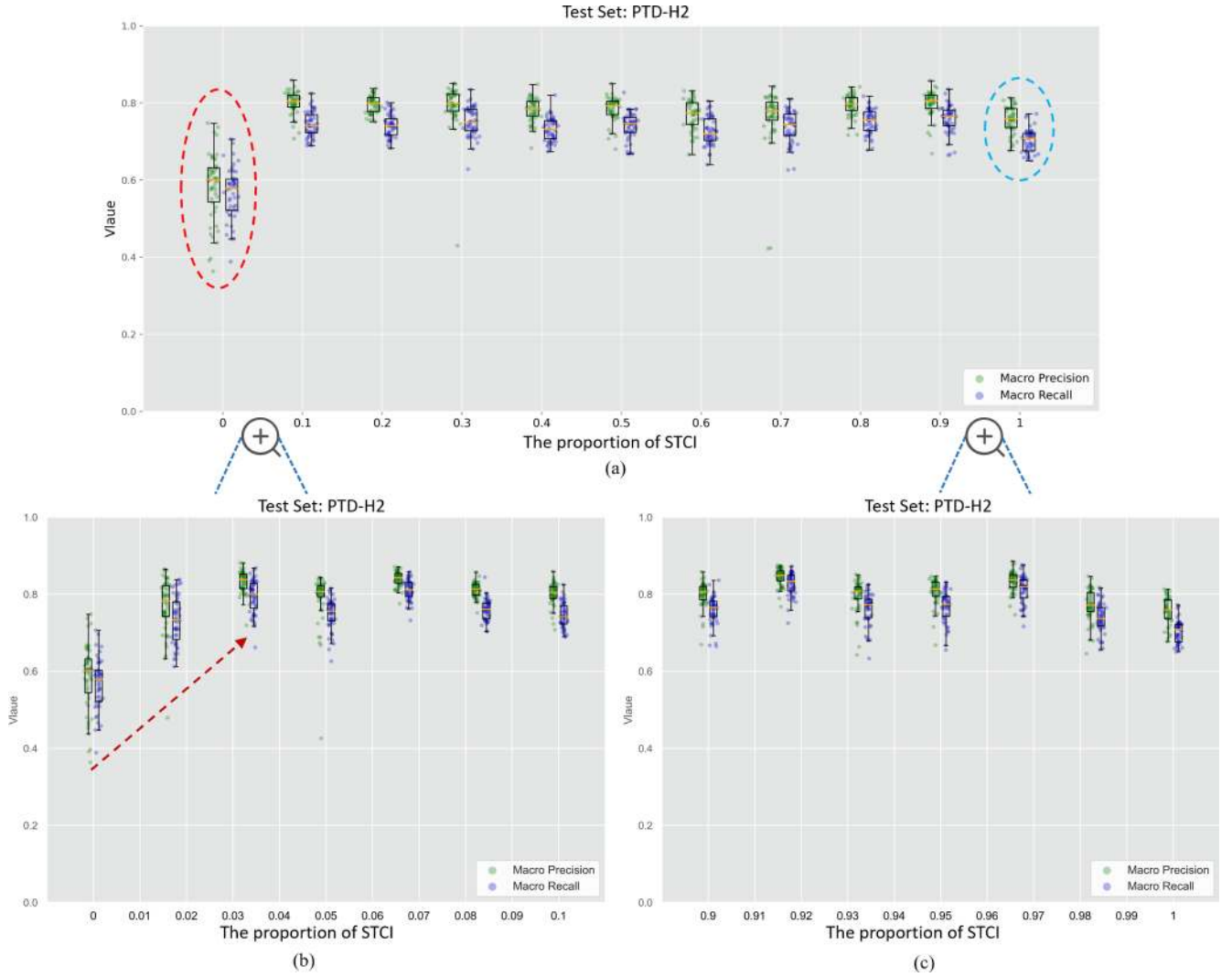


Fig. 9. Boxplots of proportion experiments on PTD-H2: (a) the proportion ranging from 0 to 1, with intervals of 0.1; (b) the proportion ranging from 0 to 0.1, with intervals of 0.167 (1/60); (c) the proportion ranging from 0.9 to 1, with intervals of 0.167 (1/60).

T. chinensis can hardly be identified even under specific image stretching. At this stage, both CLEAR ONLY and IET-based cloud removal methods have nearly lost their ability to recognize vegetation, while the other methods also recognized significantly unstable vegetation contours. Overall, the visual results in Fig. 7 align well with the quantitative experimental result.

4.3. STCI proportion experiment

In this section, we added different proportions of STCI in the training set to assess the model's performance and determine the optimal proportion. The experiments were conducted by mixing the ASM_NL-generated STCI (cloud density equal to PTD-H2) proportionally into the PTD-CLEAR and dividing those into three groups: a proportion range from 0 to 1 with an interval of 0.1, a proportion range from 0 to 0.1 with an interval of 0.167 (1/60), and a proportion range from 0.9 to 1 with an interval of 0.167 (1/60). This results in a total of 21 independent sub-experiments. The UNet model is the only semantic segmentation model used in the sub-experiment.

Table 1 displays the quantitative results of the accuracy evaluation for the simulated cloud proportion experiment, with the corresponding boxplots of each independent sub-experiment shown in Fig. 8 and Fig. 9. Firstly, we carried out experiments in the proportion range from 0 to 1. It was observed that, except for the scenarios with a proportion of 0 and 1,

the performance of other sub-experiments was relatively stable, with no significant differences. The performance at proportions 0 and 1 was considerably lower than the one of other proportions on the test set of PTD-H2 and PTD-CLEAR, respectively (marked with red ellipses in Fig. 8 and Fig. 9). These poor performance results aligned with the test image types missing from the training set, suggesting that even with a small inclusion of other style images in the training set, the model could develop a recognition capability for these styles.

The refinement experiment was conducted to determine the upper and lower limits of the simulated cloud proportion, ranging from 0 to 0.1 and 0.9 to 1. The experimental results showed a considerable improvement in the model's performance on the test set of PTD-H2 (PTD-CLEAR), notably when the training set incorporated even the minimal proportion of simulated thin cloud (cloudless) images at a ratio of 1/60. Our analysis indicates that once the proportion of simulated thin cloud (clear) images in the training set ascended to 2/60, the model's performance aligned closely with that observed at other specific proportions, as delineated by the red dashed lines in Fig. 8 and Fig. 9. Additionally, we observed that at a proportion of 1, there was a noticeable decline in performance in the test set of PTD-H2, as indicated by a blue ellipse in Fig. 9 (a). This observation implies that images from cloudless scenes also played a supporting role in the recognition of the thin cloud-covered scenes.

5. Discussion

The research recognized that real-world scenarios involve nonlinear factors that can cause specific differences between simulated clouds generated by ASM and natural clouds. We have highlighted the significance of these nonlinear components and introduced the main nonlinear effects, PSF and Perlin noise, into the linear ASM. Our experiments have confirmed the effectiveness of this enhancement. The experimental results have also shown that subtle differences between the training and test sets can strongly impact the model's performance. Whether the simulated clouds are based on ASM or ASM_NL, they visually resemble natural clouds and exhibit only minor differences in their distribution. However, their performance differs significantly from natural clouds, possibly due to the exclusion of additional nonlinear factors such as quantization errors, hardware noise, and the inherent randomness of artificial labeling. Since modeling these factors is challenging, it is crucial to balance model performance and computational cost in future work.

The results of the STCI proportion experiment in Section 4.3 showed that as cloud coverage increases, the mapping accuracy does not significantly improve, regardless of whether natural or simulated clouds are used. This means that information loss caused by data-dependent or model-dependent methods alone cannot fully compensate for information loss caused by clouds, and the model's capacity is inherently limited by its training data. Additionally, the experiment on the proportion of simulated cloud images revealed a "catalyst effect." Adding a small proportion of simulated cloud images to the training set can significantly improve the model's ability in thin cloud scenarios. However, such a small proportion of simulated cloud images is insufficient for the model to recognize a thin cloud scenario. This indicates that the "catalyst effect" is driven by activating the model's transferability.

6. Conclusion

Thin clouds would weaken the transmission of ground surface information, which affects semantic segmentation models for surface analysis. This paper comprehensively investigates the strategy that mitigates thin cloud interference in semantic segmentation by mixing the STCI. The main contributions include the generation of annotated datasets with both thin cloud and cloudless scenes, the formulaic description of the STCI mixing strategy, and a cost-balanced ASM_NL for generating high-fidelity simulated thin clouds. Large-scale experiments evaluated the strategy's feasibility and advantages compared to cloud removal methods. Adding 1/60 high-fidelity STCI to the training set could mitigate thin cloud interference significantly. We also demonstrate that the STCI mixing strategy might activate the model transferability. These findings significantly contribute to designing and optimizing deep learning-based models when mitigating the interference caused by thin clouds.

Several challenges and issues should be addressed. These include defining and measuring the domain gaps between simulated and natural clouds and exploring high-fidelity simulated cloud models with appropriate computational costs. We would leave these challenges for future research.

CRedit authorship contribution statement

Haoyu Wang: Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Junli Li:** Writing – review & editing, Funding acquisition, Conceptualization. **Zhanfeng Shen:** Supervision, Methodology, Funding acquisition. **Zihan Zhang:** Methodology. **Linze Bai:** Writing – review & editing, Visualization. **Ruifeng Li:** Methodology. **Chenghu Zhou:** Supervision. **Philippe De Maeyer:** Supervision. **Tim Van de Voorde:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported in part by the National Natural Sciences Foundation of China (U2003201), the Tianshan Talent-Science and Technology Innovation Team (2022TSYCTD0006), the Third Integrated Scientific Expedition Project in Xinjiang (2021xjkk1403), and the Chinese Academy of Sciences President's International Fellowship Initiative (2024PVB0064).

We sincerely appreciate the assistance Jiudan Zhang and Shuangshuang Li provided in the sample labelling process. We would like to thank Ms Sabine Cnudde from the Department of Geography at Ghent University for her contribution to language editing.

References

- Aboshosha, S., Zahran, O., Dessouky, M.I., Abd El-Samie, F.E., 2019. Resolution and quality enhancement of images using interpolation and contrast limited adaptive histogram equalization. *Multimedia Tools Appl.* 78, 18751–18786. <https://doi.org/10.1007/s11042-018-7022-1>.
- Atapour-Abarghouei, A., Breckon, T.P., 2018. Real-Time Monocular Depth Estimation Using Synthetic Data with Domain Adaptation via Image Style Transfer, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2800–2810.
- Chen, Y., Li, W., Sakaridis, C., Dai, D., Gool, L.V., 2018b. Domain Adaptive Faster R-CNN for Object Detection in the Wild, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3339–3348.
- Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), *Computer Vision – ECCV 2018*. Springer Int. Publishing, Cham, pp. 833–851.
- Cheng, Y., Wang, W., Ren, Z., Zhao, Y., Liao, Y., Ge, Y., Wang, J., He, J., Gu, Y., Wang, Y., Zhang, W., Zhang, C., 2023. Multi-scale feature fusion and transformer network for urban green space segmentation from high-resolution remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* 124, 103514 <https://doi.org/10.1016/j.jag.2023.103514>.
- Cui, Y., Knoll, A., 2023. Exploring the potential of channel interactions for image restoration. *Knowledge-Based Syst.* 282, 111156 <https://doi.org/10.1016/j.knsys.2023.111156>.
- Dash, A., Ye, J., Wang, G., 2024. A review of generative adversarial networks (GANs) and its applications in a wide variety of disciplines: from medical to remote sensing. *IEEE Access* 12, 18330–18357. <https://doi.org/10.1109/ACCESS.2023.3346273>.
- Ding, H., Zi, Y., Xie, F., 2023. Uncertainty-Based thin cloud removal network via conditional variational autoencoders. In: Part, I.I.I. (Ed.), *Computer Vision – ACCV 2022: 16th Asian Conference on Computer Vision*, Macao, China, December 4–8, 2022. Proceedings. Springer-Verlag, Macao, China, pp. 52–68.
- Emami, H., Aliabadi, M.M., Dong, M., Chinnam, R.B., 2021. SPA-GAN: spatial attention GAN for image-to-image translation. *IEEE Trans. Multimedia* 23, 391–401. <https://doi.org/10.1109/TMM.2020.2975961>.
- Hassan, H., Mishra, P., Ahmad, M., Bashir, A.K., Huang, B., Luo, B., 2022. Effects of haze and dehazing on deep learning-based vision models. *Appl. Intell.* 52, 16334–16352. <https://doi.org/10.1007/s10489-022-03245-5>.
- He, K., Sun, J., Tang, X., 2011. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353. <https://doi.org/10.1109/TPAMI.2010.168>.
- Jia, J., Pan, M., Li, Y., Yin, Y., Chen, S., Qu, H., Chen, X., Jiang, B., 2023. GLTF-Net: deep-learning network for thick cloud removal of remote sensing images via global-local temporality and features. *Remote Sens.* 15, 5145.
- Kaiming, H., Jian, S., Xiaoou, T., 2009. Single image haze removal using dark channel prior, 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1956–1963.
- Kingma, D., Ba, J., 2014. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*.
- Kullback, S., Leibler, R.A., 1951. On information and sufficiency. *Ann. Math. Stat.* 22, 79–86. <https://doi.org/10.1214/aoms/117729694>.
- Li, G., Jiang, S., Yun, I., Kim, J., Kim, J., 2020. Depth-wise asymmetric bottleneck with point-wise aggregation decoder for real-time semantic segmentation in urban scenes. *IEEE Access* 8, 27495–27506. <https://doi.org/10.1109/ACCESS.2020.2971760>.
- Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z., 2019. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* 28, 492–505. <https://doi.org/10.1109/TIP.2018.2867951>.

- Li, X., Yu, H., Zhao, C., Fan, C., Zou, L., 2023. DADRnet: Cross-domain image dehazing via domain adaptation and disentangled representation. *Neurocomputing* 544, 126242. <https://doi.org/10.1016/j.neucom.2023.126242>.
- Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2020. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 318–327. <https://doi.org/10.1109/TPAMI.2018.2858826>.
- Long, M., Ding, G., Wang, J., Sun, J., Guo, Y., Yu, P.S., 2013. Transfer Sparse Coding for Robust Image Representation, 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 407–414.
- Lv, H., Wang, Y., Shen, Y., 2016. An empirical and radiative transfer model based algorithm to remove thin clouds in visible bands. *Remote Sens. Environ.* 179, 183–195. <https://doi.org/10.1016/j.rse.2016.03.034>.
- Ma, X., Wang, Q., Tong, X., Atkinson, P.M., 2022. A deep learning model for incorporating temporal information in haze removal. *Remote Sens. Environ.* 274, 113012 <https://doi.org/10.1016/j.rse.2022.113012>.
- Mitchell, O.R., Delp, E.J., Chen, P.L., 1977. Filtering to remove cloud cover in satellite imagery. *IEEE Trans Geosci Electron* 15, 137–141. <https://doi.org/10.1109/TGE.1977.6498971>.
- Nielsen, F., 2019. On the jensen-shannon symmetrization of distances relying on abstract means. *Entropy* 21, 485.
- Niu, Z., Zhu, C., Wang, C., 1997. Analysis of atmospheric cross radiation in a slant viewing angle. *National Remote Sensing Bulletin* 88–93. <https://doi.org/10.11834/jrs.19970202>.
- Pei, Y., Huang, Y., Zou, Q., Zhang, X., Wang, S., 2021. Effects of image degradation and degradation removal to CNN-based image classification. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 1239–1253. <https://doi.org/10.1109/TPAMI.2019.2950923>.
- Perlin, K., 2002. Improving noise. *ACM Trans. Graph.* 21, 681–682. <https://doi.org/10.1145/566654.566636>.
- Pizer, S.M., Johnston, R.E., Ericksen, J.P., Yankaskas, B.C., Muller, K.E., 1990. Contrast-limited adaptive histogram equalization: speed and effectiveness, *Proceedings of the First Conference on Visualization in Biomedical Computing*, pp. 337–345.
- RL.Easton.Jr, 2010. Multidimensional Fourier Transforms, *Fourier Methods in Imaging*, pp. 325–346.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. *Springer International Publishing, Cham*, pp. 234–241.
- Sakaridis, C., Dai, D., Van Gool, L., 2018. Semantic foggy scene understanding with synthetic data. *Int. J. Comput. Vision.* 126, 973–992. <https://doi.org/10.1007/s11263-018-1072-8>.
- Shao, Y., Li, L., Ren, W., Gao, C., Sang, N., 2020. Domain adaptation for image dehazing. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2805–2814.
- Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>.
- Shen, H., Li, H., Qian, Y., Zhang, L., Yuan, Q., 2014. An effective thin cloud removal procedure for visible remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 96, 224–235. <https://doi.org/10.1016/j.isprsjprs.2014.06.011>.
- Singh, P., Komodakis, N., 2018. Cloud-gan: cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks. In: *IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1772–1775.
- Singh, M., Laxmi, V., Faruki, P., 2022. Visibility enhancement and dehazing: Research contribution challenges and direction. *Comput. Sci. Rev.* 44 <https://doi.org/10.1016/j.cosrev.2022.100473>.
- Song, R., Li, T., Li, T., 2023. Ship detection in haze and low-light remote sensing images via colour balance and DCNN. *Appl. Ocean Res.* 139, 103702 <https://doi.org/10.1016/j.apor.2023.103702>.
- Sun, X., Wang, B., Wang, Z., Li, H., Li, H., Fu, K., 2021. Research progress on few-shot learning for remote sensing image interpretation. *IEEE J Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 2387–2402. <https://doi.org/10.1109/jstars.2021.3052869>.
- Tong, X.-Y., Xia, G.-S., Lu, Q., Shen, H., Li, S., You, S., Zhang, L., 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* 237, 111322 <https://doi.org/10.1016/j.rse.2019.111322>.
- Tseng, D.C., Tseng, H.T., Chien, C.L., 2008. Automatic cloud removal from multi-temporal SPOT images. *Appl. Math. Comput.* 205, 584–600. <https://doi.org/10.1016/j.amc.2008.05.050>.
- Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., Atkinson, P.M., 2022. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS J. Photogramm. Remote Sens.* 190, 196–214. <https://doi.org/10.1016/j.isprsjprs.2022.06.008>.
- Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P., 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv Neural Inf Proces Syst* 34, 12077–12090.
- Xu, M., Deng, F., Jia, S., Jia, X., Plaza, A.J., 2022. Attention mechanism-based generative adversarial networks for cloud removal in Landsat images. *Remote Sens. Environ.* 271, 112902 <https://doi.org/10.1016/j.rse.2022.112902>.
- Yuan, Z., Li, Y., Yang, J., 2022. Improving hazy image recognition by unsupervised domain adaptation, 2022 17th international conference on control. Automation, Robotics and Vision (ICARCV), pp. 311–316.
- Zhang, Y., 2004. Understanding image fusion. *Photogramm Eng Remote Sens* 70, 657–661.
- Zhang, L., Zhu, A., Zhao, S., Zhou, Y., 2021. Simulation of atmospheric visibility impairment. *IEEE Trans Image Process.* 30, 8713–8726. <https://doi.org/10.1109/TIP.2021.3120044>.
- Zheng, W.J., Zhao, X.L., Zheng, Y.B., Lin, J., Zhuang, L., Huang, T.Z., 2023. Spatial-spectral-temporal connective tensor network decomposition for thick cloud removal. *ISPRS J. Photogramm. Remote Sens.* 199, 182–194. <https://doi.org/10.1016/j.isprsjprs.2023.04.006>.
- Zhou, H., Chang, Y., Yan, W., Yan, L., 2023. Unsupervised cumulative domain adaptation for foggy scene optical flow. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9569–9578.
- Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. UNet++: A Nested U-Net architecture for medical image segmentation. *Springer International Publishing, Cham*, pp. 3–11.
- Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251.