remote sensing

MDPI

Article

An Innovative Approach for Effective Removal of Thin Clouds in Optical Images Using Convolutional Matting Model

Renzhe Wu [1], Guoxiang Liu [1,2,*], Jichao Lv [1], Yin Fu [1], Xin Bao [1], Age Shama [1], Jialun Cai [3], Baikai Sui [1], Xiaowen Wang [1,2] and Rui Zhang [1,2]

[1] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China; mrwurenzhe@my.swjtu.edu.cn (R.W.); lvjichao@my.swjtu.edu.cn (J.L.); rsyinfu@my.swjtu.edu.cn (Y.F.); baoxin@my.swjtu.edu.cn (X.B.); shamaage@my.swjtu.edu.cn (A.S.); baikaisui@my.swjtu.edu.cn (B.S.); insarwxw@swjtu.edu.cn (X.W.); zhangrui@swjtu.edu.cn (R.Z.)

[2] State-Province Joint Engineering Laboratory of Spatial Information Technology of High-Speed Rail Safety, Southwest Jiaotong University, Chengdu 611756, China

[3] School of Environment and Resource, Southwest University of Science and Technology, Chengdu 621010, China; caijialun@my.swjtu.edu.cn

* Correspondence: rsgxliu@swjtu.edu.cn

Abstract: Clouds are the major source of clutter in optical remote sensing (RS) images. Approximately 60% of the Earth's surface is covered by clouds, with the equatorial and Tibetan Plateau regions being the most affected. Although the implementation of techniques for cloud removal can significantly improve the efficiency of remote sensing imagery, its use is severely restricted due to the poor timeliness of time-series cloud removal techniques and the distortion-prone nature of single-frame cloud removal techniques. To thoroughly remove thin clouds from remote sensing imagery, we propose the Saliency Cloud Matting Convolutional Neural Network (SCM-CNN) from an image fusion perspective. This network can automatically balance multiple loss functions, extract the cloud opacity and cloud top reflectance intensity from cloudy remote sensing images, and recover ground surface information under thin cloud cover through inverse operations. The SCM-CNN was trained on simulated samples and validated on both simulated samples and Sentinel-2 images, achieving average peak signal-to-noise ratios (PSNRs) of 30.04 and 25.32, respectively. Comparative studies demonstrate that the SCM-CNN model is more effective in performing cloud removal on individual remote sensing images, is robust, and can recover ground surface information under thin cloud cover without compromising the original image. The method proposed in this article can be widely promoted in regions with year-round cloud cover, providing data support for geological hazard, vegetation, and frozen area studies, among others.

Keywords: image superposition; image matting; cloud removal; deep learning; remote sensing image

Citation: Wu, R.; Liu, G.; Lv, J.; Fu, Y.; Bao, X.; Shama, A.; Cai, J.; Sui, B.; Wang, X.; Zhang, R. An Innovative Approach for Effective Removal of Thin Clouds in Optical Images Using Convolutional Matting Model. *Remote Sens.* **2023**, *15*, 2119. https://doi.org/10.3390/rs15082119

Academic Editors: Xiangrong Zhang, Yansheng Li, Lichao Mou, Licheng Jiao and Xu Tang

Received: 16 March 2023
Revised: 12 April 2023
Accepted: 14 April 2023
Published: 17 April 2023

1. Introduction

Since the first remote sensing (RS) satellite termed Sputnik-1 was launched in 1957, the collection of RS images from space has been uninterrupted [1]. The imaging mode of visible RS images is most consistent with human visual perception. This characteristic has led to the development of several ground detection techniques and continuous monitoring methods. In addition, it has found considerable applications in disaster relief, geology, environmental monitoring, and engineering construction, accompanied by extensive promotion and adoption [2]. Nonetheless, the global surface cloud coverage can fluctuate between 58% [3] and 61% [4], which, subsequently, leads to a reduction in the availability of most visible remote sensing products.

Therefore, enhancing the efficacy of visible RS images is of utmost importance. Furthermore, the process of cloud removal can be regarded as a type of image reconstruction, which is heavily dependent on precise cloud detection [5]. The principal techniques for

cloud removal can be divided into two primary categories based on the number of RS images utilized: multi-temporal methods and single-image methods [6–9].

Cloud removal in time-series remote sensing images is a widely used method, particularly in big data platforms such as the Google Earth Engine, which involves combining same-track images for efficient image element replacement. Researchers have established various cloud removal restoration models, such as those using dictionary learning [10,11], spatially and temporally weighted regression [12], low-rank representation [13], deep learning [9,14], and nonnegative matrix factorization [15]. Despite this, removing time-series remote sensing images may result in inconsistent image tone and surface information due to the temporal consistency and seasonal variability of remote sensing images, especially in tropical and mountainous regions, where the number of cloud-free images is limited, further exacerbating information variability.

Single remote sensing image cloud removal can be achieved using four primary methods: image enhancement, spatial interpolation, atmospheric transport model, and generate adversarial network-based surface information reconstruction. Image enhancement methods alter contrast and weaken image components to suppress clouds, resulting in unrealistic visual completion that cannot be used for remote sensing target detection and quantitative analysis [16–18]. Spatial interpolation techniques such as Kriging spatial interpolation and the neighborhood similar pixel interpolator can eliminate speckled clouds, but they face challenges in effectively recovering the surface information under a wide range of clouds [19–21]. Atmospheric transport models such as haze-optimized transformation and dark channel defogging exploit prior knowledge and can effectively remove clouds, but changes in the prior knowledge may lead to biased image estimates [7,8,22,23]. Generative adversarial networks have shown promise in cloud removal and surface information reconstruction, but they do not recognize when thin and thick clouds coexist, resulting in significant disparities between the output and original image [24–28].

Recently, researchers have proposed novel techniques for improving the stability of remote sensing images through cloud removal. Li et al. developed a deep-learning-based approach, CR-MSS, which leverages short-wave infrared to eliminate thin clouds from high-resolution images [29]. Image matting technology has also made significant advances in recent years [30–34], finding applications in fields such as image fusion [35], shadow removal [36], semantic segmentation [37], image defogging [38], cloud estimation [39], and beyond. By overlaying foreground and background images and measuring foreground transparency, precise foreground information can be extracted.

CNNs are highly competent in extracting image features as a result of their local connectivity, parameter sharing, and translational invariance, resulting in a more comprehensive feature extraction compared to generative adversarial networks [40]. Moreover, CNN optimization is simpler, and the model quickly and stably converges. In this study, our primary goal is to estimate cloud opacity, which may be considered a type of noise (akin to Perlin noise). While the discriminator can create a more natural, smooth, and visually appealing generated image in generative adversarial networks, the generator can easily generate noise signals to deceive the discriminator, which may result in limited significance of the loss function presented by the discriminator and instability in the model training process. Thus, we suggest employing a CNN to assess the background information and extract the foreground information in the image. Based on this, we propose the "SCM-CNN" (Saliency Cloud Matting Convolutional Neural Network) model that integrates deep learning techniques with image matting for remote sensing image cloud recognition, cloud opacity estimation, and cloud removal.

The conventional method for image matting necessitates the incorporation of a "Trimap", which involves the triple classification of images. To lessen the model's dependence on Trimaps, we employed a saliency monitoring network to enhance the model's applicability.

It is crucial to acknowledge that the surface characteristics will evolve over time, and the levels of solar radiation and aerosol concentration will vary with each RS satellite

observation. Obtaining an actual control group of overcast and clear images is unfeasible. Therefore, creating an RS image that is as close to reality as possible is a prerequisite for the SCM-CNN model to converge successfully. In this regard, we refer to Matting dataset construction scenarios such as the alpha-matting dataset [41], the portrait image matting dataset [42], and the traditional RS image process for generating samples for cloud detection, such as L7 Irish [43] and L8 SPARCS [44], in this study. Cloud opacity is determined based on the color range extracted from Sentinel-2 satellite images covering the sea's surface. Subsequently, the samples are aggregated into a single image to form the precise label, and then the training and validation dataset required for the study are constructed using the RGB band images of cloud-free Sentinel-2 as the base image.

The contributions of this article are summarized as follows.

1. We constructed a cloud-matting dataset generation method and created a set of high-quality cloud-matting datasets based on Sentinel-2 imagery. The cloud-matting dataset outperforms the commonly used semantic labels for cloud detection, by effectively distinguishing the clouded and cloud-free areas in RS images with 100% accuracy. Moreover, it accurately describes the mixing degree of image elements in cloud-covered areas. As the cloud-matting dataset closely resembles cloud scenes in real RS images, it enables various applications such as accurate cloud detection, image reconstruction, and cloud opacity estimation.
2. This work presents an integrated model for cloud detection, cloud opacity estimation, and cloud removal based on the principle of deep learning Image-matting. The model utilizes the saliency detection function to eliminate the need for a "Trimap", enabling cloud removal from a different perspective. Cloud removal in this model relies on cloud identification and similarity analysis with the original image. It effectively recovers the surface area beneath thin clouds, even in scenarios of coexisting thick and thin clouds.
3. The proposed method includes a "Channel Global Max/Average Pooling" structure that estimates the reflection of the foreground pixel efficiently with minimal parameters. The structure efficiently extracts the foreground pixel reflection with minimal parameters from the feature map.
4. Our proposed method is a multi-objective loss function gradient optimization approach that calculates the gradient deviation of the loss function rather than relying on the conventional weight linear combination of multiple loss functions. The model gradient is updated accordingly.

The following is the organization of this article. Section 2 presents the superposition model of the remote sensing (RS) images, the formulation of cloud removal under the image matting framework, and the dataset and evaluation metrics that were utilized. Section 3 provides a detailed exposition of the experimental procedures and results. In Section 4, we analyze the relative advantages and disadvantages of various methods and propose possible enhancements. Finally, Section 5 draws conclusive remarks.

2. Materials and Methods

2.1. Dataset

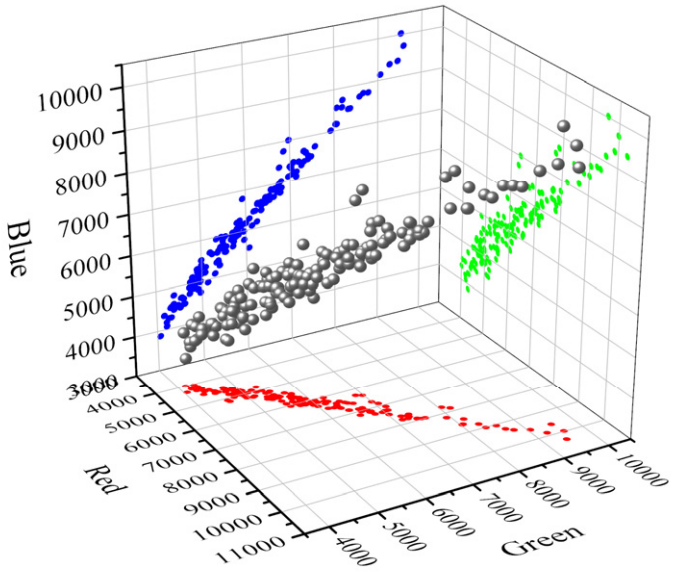
The current cloud dataset is primarily designed for cloud detection, using a mask to distinguish cloud areas from others. However, this dataset is unsuitable for cloud–fog retrieval tasks, and its inherent bias may significantly reduce the accuracy of machine learning models. To address these issues, we employ a colorimetric approach and Sentinel-2 satellite ocean imagery to obtain simulated cloud images and establish a normalized cloud opacity layer. By overlaying and merging opacity layers of cloud layers from multiple scenes, we were able to generate highly realistic cloud simulations.

Cloud-Free Image Construction: Our objective was to remove thin clouds under less cloudy mountainous conditions. To achieve this, we used a dataset of 24 Sentinel-2 remote sensing images covering the southeastern Tibetan region. The images were processed with S2Cloudless to detect clouds and calculate cloud shadows based on satellite incidence

angles and surface dark pixel information. A sliding window of a 512 × 512 pixel size with a step of 256 pixels was then used to traverse the images, generating cut-out remote sensing images without clouds and cloud shadows and discarding those with clouds or cloud shadows.

Cloud Opacity Image Construction: Some researchers utilized manual masking [9] or Perlin noise [45] to generate cloud images, while simulating cloud images that approximated real cloud distributions. Therefore, we extracted cloud distributions from actual ocean surface remote sensing images to simplify the process and reduce errors. Multiple Sentinel-2 remote sensing images were selected, and the opacity of clouds was extracted using a color range. The opacity of clouds was normalized, and image closing operations were performed to eliminate holes caused by convective cloud shadows. Multiple cloud opacity images with different cloud thicknesses and shapes were obtained by multiple extractions and merging.

Simulating Cloud Image Generation: In Sentinel-2 images, there is no evident distinction between the red, green, and blue bands of cloud areas. They exhibit a strong linear relationship with some perturbations around the linear regression function. In Figure 1a, we randomly selected some points for fitting, and the results indicated that there was little difference in pixel brightness between the red, green, and blue bands in the range of 4000–8000 pixels. As red, green, and blue all belong to the visible light band, their wavelengths vary slightly, resulting in similar penetration capabilities for clouds. Therefore, pixel value perturbations are more likely caused by factors such as surface objects, particles in the cloud layer, and aerosols. Simulating these perturbations is a complex task. To avoid human errors and simplify the model, we assume that the RGB bands of cloud layers in RS images have the same opacity value. According to the assumption in Section 2.2 that ε_{cloud} can be a constant in a local area, we simulated cloud images based on reference formula $\varepsilon = (1 - \alpha)\varepsilon_{ground} + \alpha\varepsilon_{cloud}$, where $\varepsilon_{cloud} \in [4000, 8000]$ and α are 512 × 512 random blocks in the opacity map of cloud layers, as shown in Figure 1b.



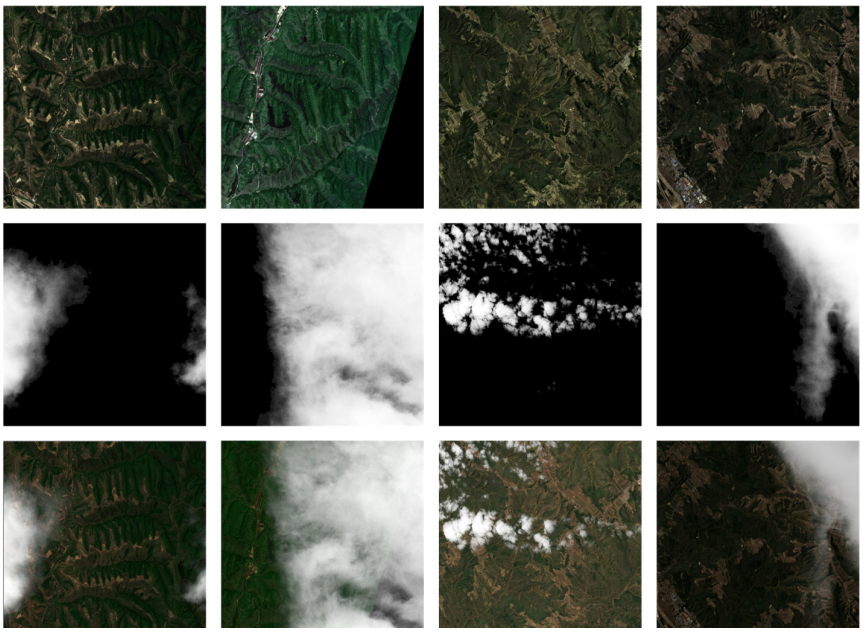
$$Green = 642 + 0.876 \times Red$$

$$Blue = 827 + 0.875 \times Red$$

Red	Green	Blue
4000	4327	4146
5000	5022	5202
6000	5898	6077
7000	6774	6952
8000	7650	7827

(**a**)

Figure 1. *Cont.*



(**b**)

Figure 1. (**a**) Pixel value statistics of red, green, and blue wave segments in Sentinel-2's cloud region. Two sets of linear regression calculations were carried out based on the red band, and the values of 4000–8000 pixels were instantiated. (**b**) dataset generation for cloud matting. The first row contains the bottom image of a Sentinel-2 RS image without clouds (Background); the second row is foreground pixel reflection intensity multiplied by the cloud opacity (Foreground); the third row contains the generated RS image with clouds.

2.2. Superposition Model of RS Images

In Figure 2, the cloud represents the foreground image, while the surface object acts as the background image, with the cloud's opacity being corresponding to the foreground image's opacity (α). As a result, the satellite RS image is formed after superimposing the background pixel reflection signal (ε_{ground}) and the foreground pixel reflection signal (ε_{cloud}). The thin cloud effect is created by combining the multiplicative factor that attenuates the light from below the clouds and the additive factor that adds reflected light at the cloud tops.

$$\varepsilon = (1-\alpha)\varepsilon_{ground} + \alpha\varepsilon_{cloud}, \tag{1}$$

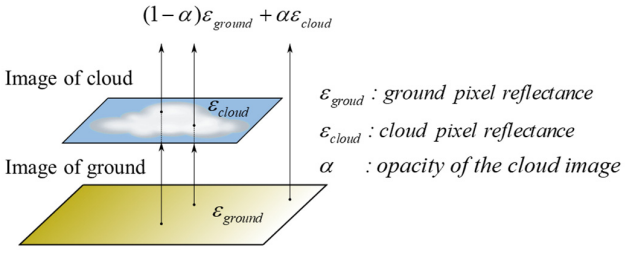


Figure 2. RS image imaging process schematic diagram. α is the opacity of the foreground, and the foreground and background pixel reflection signals are ε_{cloud} and ε_{ground}, respectively.

This paper adopts image matting for RS image cloud removal, and the cloud shadow is not considered for the time being. If assuming that cloud shadow does not exist, Equation (1) is a simplified model which has three unknowns. Substantially, we only need to obtain the α and ε_{cloud} to calculate ε_{ground}.

$$\varepsilon_{ground} = \frac{\varepsilon - \varepsilon_{cloud}}{1-\alpha} + \varepsilon_{cloud}, \tag{2}$$

Nevertheless, it should be noted that ε_{cloud} is independent of α if we use α to solve the derivative at both ends of Equation (2).

$$\frac{\partial \varepsilon_{ground}}{\partial \alpha} = \frac{\varepsilon - \varepsilon_{cloud} + (1-\alpha)\partial \varepsilon / \partial \alpha}{(1-\alpha)^2}, \tag{3}$$

Equation (3) shows the relationship between ε_{ground} and α after each cloud removal restoration process in the image. When $\alpha \to 1$, the denominator is infinitely close to 0, and even a small perturbation will lead to a significant error, reducing the model's reliability. For that reason, we classify clouds into thin clouds ($\alpha \leq 0.5$) and thick clouds ($1 \geq \alpha > 0.5$), according to their opacity.

2.3. Overview of SCM-CNN

The SCM-CNN model suggested in this research performs three significant roles, which are as follows.

I. Cloud matting: The central focus of the research presented in this paper is the concept of cloud matting, which ultimately leads to cloud opacity. In order to reduce model complexity, we have discontinued the use of the "Trimap" input and "Trimap" generating strategy, although this decision has resulted in significant challenges for model inference. To address this issue, we propose the use of a salient target detection function, which has proven to be highly effective in detecting the desired target type. This approach involves learning multi-scale features from datasets with diverse backgrounds and cloud combinations of varying brightness and opacity followed by the merging the multi-features of pixel similarity and spatial similarity to produce high-precision cloud matting results.

II. Cloud Maximum Digital Number (CMaxDN) value: On a local scale, foreground pixel reflection intensity is a constant approximation to solar radiation. Thus, we assume that the maximum brightness of clouds (ε_{cloud}) in a scene of RS images is consistent and may, likewise, be viewed as a constant.

This assumption can greatly simplify model operation. Because of hardware constraints, the convolutional neural network cannot simultaneously read and write the entire RS image. We usually train the model by image slicing, thus $\widetilde{\varepsilon}_{cloud} = \max_{\Omega} \varepsilon_{cloud}(\gamma)$ is used in the actual operation, γ representing the solar radiation, Ω representing the slicing range of the image, and $\widetilde{\varepsilon}_{cloud}$ indicating the maximum reflected brightness of the cloud within the slicing range.

III. Cloud removal: Cloud removal and cloud matting are inextricably linked. Equation (3) explores the relationship between the ε_{ground} and α, demonstrating that a single RS image cannot restore surface information that thick clouds have covered. Combining Equation (3) and observing the generated cloud images, we found that even slight disturbances in the estimation of cloud opacity of a single image with $\alpha > 0.8$ could greatly affect the result of cloud removal. That is to say that the effect of cloud removal for some single images with $\alpha > 0.8$ is poor. In order to improve the reliability of the model for cloud removal and image restoration, we adopted the method of setting a threshold to constrain the range of α values considered, and the process for cloud removal and restoration is shown in Equation (4).

$$\varepsilon_{ground} = \frac{\varepsilon - \varepsilon_{cloud}}{1 - \min(\alpha, 0.8)} + \varepsilon_{cloud}, \tag{4}$$

2.4. Model and Algorithm

SCM-CNN employs a superior saliency detection network capable of analyzing RS images from many sizes and scenes and fusing multi-stage features. As depicted in Figure 3a, the SCM-CNN consists of two primary components: the automatic production of the cloud-matting mask and the slice maximum digital number estimation. SCM-CNN's backbone network is U^{2}Net [46], and the RSU (Residual U-block) is employed for feature extraction and feature fusion. The traditional Residual block (RES) can be expressed

as $h(x) = f_2(f_1(x)) + x$, where $h(x)$ denotes the mapping result of the input feature's expectation; $f_1 f_2$ denote the two-weight layers, respectively. Each residual operation of RSU adopts a U-shaped structure, and the overall operation of RSU is expressed as $h_{RSU}(x) = u(f_1(x)) + f_1(x)$. Each residual operation yields global multi-scale features, and the RES module is superior in the model perception field and feature redundancy. The FLOPs of the RSU module are similar to the RES module, which can be faster for training and model inference (15FPS, with the input size of $512 \times 512 \times 3$ on P4000 GPU). The model reference Feature Pyramid Networks (FPN) structure has 12 outputs, and a total of 14 outputs are obtained after further stacking to obtain two outputs of fusion features. Seven outputs represent the cloud-matting mask, and the others represent the CMaxDN value.

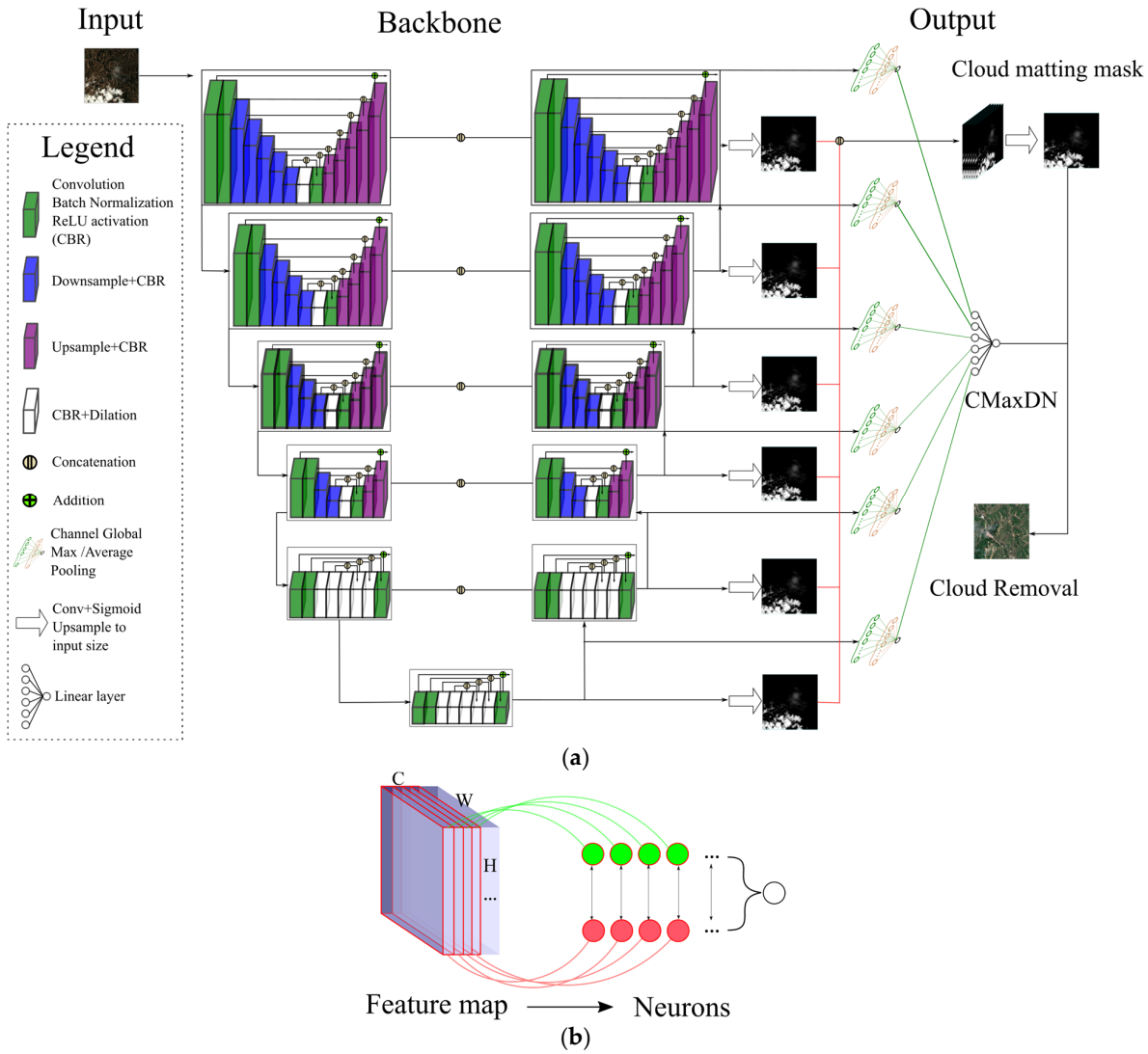


Figure 3. (**a**) SCM-CNN(U^{2}Net) model. (**b**) Channel global Average and Max Pooling. Where C, W, and H represent the number of channels, width and height of the image, respectively. Firstly, each channel of the feature map is fused into one neuron, and secondly, multiple neurons are further fused into one neuron. Furthermore, this neuron is the CMaxDN.

The CMaxDN is a piece of intrinsic information contained within the feature map. However, using the convolutional branch to perform operations based on the feature maps leads to a considerable increase in computational complexity since there are a large number of feature maps. To simplify the calculation process and obtain the CMaxDN value comprehensively based on multi-channel feature maps, we have designed a structure known as "Channel Global Average/Max Pooling", as shown in Figure 3b.

Typically, the CMaxDN value approached the global brightest value. Therefore, we used global max pooling to obtain the brightest value of the feature map. However, there are times when CMaxDN was greater than the brightest value of the feature map, such as in the case of low cloud opacity. In these scenarios, we hope that the model could extract more realistic CMaxDN values from the feature map. Consequently, we utilized global average pooling to summarize the feature map information.

In addition, we incorporated two Linear layers to fuse the maximum/average pooling values obtained from each feature map. Nonlinear transformation was then performed using the Relu activation function. Finally, we fused the multiple outputs of the Feature Pyramid Networks (FPN) to obtain the predicted CMaxDN. The "Channel Global Average/Max Pooling" structure only used 7144 parameters from the essence of CMaxDN to yield excellent prediction results. It integrated features from multiple channels and could effectively assist in cloud removal tasks.

The "Channel Global Average & Max Pooling" structure consisted of two linear layers. The primary objective of the first linear layer was to calculate the global average pooling and global maximum pooling values for each channel of every feature map, with the aim of generating a condensed representation of the feature map. The second linear layer employed multiple feature map compression expressions to establish the CMaxDN value.

2.5. Loss Function

The output of SCM-CNN consisted of two parts. Thus, we use the combined loss function to evaluate the model outcomes based on pixel and structural similarity, respectively. The 1-norm's prediction results are more precise and consistent with human visual perception than other norms. Then, Equations (5) and (6) used 1-norm to establish $L_{||\alpha||}$ and $L_{||\varepsilon_{cloud}||}$ to measure the gap between pixels for the cloud-matting mask ($\widetilde{\alpha}$) and CMaxDN value ($\widetilde{\varepsilon}_{cloud}$), respectively.

$$L_{||\alpha||} = \sum_{i=1}^{n} |\widetilde{\alpha}_i - \alpha_i|, \tag{5}$$

$$L_{||\varepsilon_{cloud}||} = |\widetilde{\varepsilon}_{cloud} - \varepsilon_{cloud}|, \tag{6}$$

We also considered the image's general structural similarity to improve the efficacy of cloud removal. However, the structural similarity $\widetilde{\alpha}$ was not apparent. To solve this problem, we used $\widetilde{\alpha}$ and $\widetilde{\varepsilon}_{cloud}$ pollute the cloud-free tag image (Equation (7)) and judge the structural similarity between the contaminated image and the input image. This strategy could prevent the occurrence of a zero denominator and ensure the smooth execution of model training. "MS-SSIM" is an image quality evaluation method that combines image features with different resolutions, which may be evaluated comprehensively based on the two images' brightness, contrast, and structural similarity [47]. The "MS-SSIM" loss function is calculated as shown in Equation (8), M represents the scale factor, $[\mu_{\widetilde{\varepsilon}}, \mu_{\varepsilon}]$ represents the mean value of the predicted feature map and the actual image, $[\sigma_{\widetilde{\varepsilon}}, \sigma_{\varepsilon}]$ represents the standard deviation between the predicted image and the actual image, $\sigma_{\widetilde{\varepsilon}\varepsilon}$ represents the covariance between the predicted image and the actual image, $[\beta_m, \gamma_m]$ represents the importance between the two multipliers, and $[c_1, c_2]$ is a constant term to prevent the divisor from being zero.

$$\widetilde{\varepsilon} = (1 - \widetilde{\alpha})\varepsilon_{ground} + \widetilde{\alpha}\widetilde{\varepsilon}_{cloud}, \tag{7}$$

$$L_{ms-ssim}(\widetilde{\varepsilon}, \varepsilon) = 1 - \prod_{m=1}^{M} \left(\frac{2\mu_{\widetilde{\varepsilon}}\mu_{\varepsilon} + c_1}{\mu_{\widetilde{\varepsilon}}^2 + \mu_{\widetilde{\varepsilon}}^2 + c_1} \right)^{\beta_m} \left(\frac{2\sigma_{\widetilde{\varepsilon}\varepsilon} + c_2}{\sigma_{\widetilde{\varepsilon}}^2 + \sigma_{\varepsilon}^2 + c_2} \right)^{\gamma_m}, \tag{8}$$

Equation (9) superimposes all the error functions to obtain the loss function used in the model training.

$$L_{SCM-CNN} = \beta_1 L_{||\alpha||} + \beta_2 L_{||\varepsilon_{cloud}||} + \beta_3 L_{ms-ssim}, \tag{9}$$

It is worth mentioning that this paper has seven sets of output results, thus we finally get the loss value as shown in Equation (10).

$$L_{SCM-CNN} = \sum_{i=1}^{7} \beta_1 L_{||\alpha||_i} + \beta_2 L_{||\varepsilon_{cloud}||_i} + \beta_3 L_{ms-ssim_i}, \tag{10}$$

The optimization objectives of the three loss functions were distinct from each other. The common practice is to fuse multiple loss functions by linearly combining them with weights. Through careful observation and experimentation, it was discovered that $\beta_1 = 0.2$ $\beta_2 = 0.4$ $\beta_3 = 0.4$ could effectively balance multiple loss functions. However, linear weighting may not always produce optimal results due to the differences in gradient direction among loss functions. Instead of using a simple linear weighting approach, this paper adopted an algorithm inspired by multi-task learning, which was suitable for optimizing multiple objectives in a single task [48–50]. As depicted in Algorithm 1, the algorithm calculated the deviation of the gradient norm for each loss function and assigned smaller weights to those with larger deviations and larger weights to those with smaller deviations. This ensured that loss functions with larger gradients do not dominate during model updates, thus ensuring equal attention is given to all loss functions.

Algorithm 1 Automatic weighting of loss functions

Input: [loss1.loss2], Model, Learning Rate:
Output: Model

function $(Grad_{opt})[loss1, loss2], Model, LearningRate$
%CALCULATE THE GRADIENT OF ALL LOSS FUNCTIONS
$grad_i \leftarrow \nabla loss_i$
%FAN OF THE GRADIENT OF THE LOSS FUNCTION
$norm_i \leftarrow \frac{grad_i - grad_i}{\sigma_i}$, $\sigma_i = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(grad_i - grad_i)^2}$
%AVERAGE OF THE NORM OF THE GRADIENT OF THE LOSS FUNCTION
$std_i \leftarrow STD(norm_i)$
%DEVIATION FROM THE NORM OF THE GRADIENT OF THE LOSS FUNCTION
$dev_i \leftarrow \frac{norm_i - norm_i}{std_i}$
%CALCULATE THE WEIGHTS ACCORDING TO THE DEGREE OF DEVIATION
$weight_i \leftarrow \exp(-dev_i)$
%NORMALISATION OF THE OBTAINED WEIGHTS
for j **in** $(weight_i)$ **do**
$j \leftarrow j/\sum weight_i$
end for
%CALCULATE THE WEIGHTED GRADIENT
$grad_i \leftarrow weight_i \times grad_i$
%UPDATE THE MODEL PARAMETERS ACCORDING TO THE GRADIENT
for $param, g$ **in** $(Model.parameters(), grad_i)$ **do**
$param \leftarrow LearningRate \times g$
end for
end function

2.6. Evaluation Metrics

We evaluated the model output and actual samples using metrics based on pixel similarity and structural similarity. Three major types of evaluation metrics were used. (1) The mean square error MSE (Equation (11)) was used to verify the computational

accuracy of cloud opacity directly; the actuarial accuracy of the CMaxDN and pixel gap between the synthetic cloud image and the input image was denoted as $RMSE(\widetilde{\alpha}, \alpha)$, $RMSE(\widetilde{\varepsilon}_{cloud}, \varepsilon_{cloud})$ and $RMSE(\widetilde{\varepsilon}, \varepsilon)$, respectively. (2) The peak signal-to-noise ratio PSNR (Equation (12)) and structural similarity SSIM (Equation (13)) were used to measure the structure of cloud opacity, synthetic cloud image, and actual sample. The disparity was noted as $PSNR(\widetilde{\alpha}, \alpha)$, $SSIM(\widetilde{\alpha}, \alpha)$, $PSNR(\widetilde{\varepsilon}, \varepsilon)$, and $SSIM(\widetilde{\varepsilon}, \varepsilon)$, respectively. (3) Accuracy (Equation (14)) was used to measure the classification approximation of cloud opacity and was noted as $ACC(\widetilde{\alpha}, \alpha)$.

$$RMSE(X, Y) = \sqrt{\sum_{i=1}^{n} (Y_i - X_i)^2}, \tag{11}$$

$$PSNR(X, Y) = 10 \log_{10} \left(\frac{MAX_Y^2}{MSE(X, Y)} \right), \tag{12}$$

$$SSIM(X, Y) = \frac{(2\mu_X \mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)}, \tag{13}$$

$$ACC(X, Y) = \frac{TP + TN}{TP + TN + FP + FN}, \tag{14}$$

In Equations (11)–(14), X represents the prediction image, Y represents the reference image, and $[\mu_X, \mu_Y]$ represents the mean of the predicted image and the reference image, respectively. $[\sigma_X, \sigma_Y]$ represents the standard deviation between the predicted image and the reference image, respectively. σ_{XY} represents the covariance between the predicted image and the reference image, $[C_1, C_2]$ is a constant term to prevent the divisor from being 0, and $[TP, TN, FP, FN]$ represents the TruePositive, FalseNegative, TrueNegative, and FalsePositive, respectively.

3. Results

This paper presented a novel deep learning-based matting model named SCM-CNN, which significantly improved the accuracy of cloud removal in visible imagery. However, existing cloud removal techniques, such as image interpolation and enhancement, have shown fewer promising results compared to the chosen two categories, which were compared in this study. (1) Atmospheric transport models, "Dark Channel" [22], and "HOT" [23]. (2) Image element reconstruction, "SpA-GAN" [24,27], "Pix2pix" [28], and "CR-GAN" [26]. Additionally, a standard image-matting method: "Closed-form matting" [51].

However, the usage of "Trimap" as prior input in "Closed-form matting" poses a significant challenge in practical applications. Nevertheless, because of the "Closed-form matting" strong deductibility, it only proves the feasibility of cloud removal based on matting in simulated data sets. "Hot", "Pix2pix", and "CRGAN" have all failed and will cause damage to the original image. Limited to space, we put this part of the results and the initial results of "Dark Channel" and "SpA-GAN" in Figure A1.

In Figure 4, "SCM-CNN", "Pix2pix", "CR-GAN", and "SpA-GAN" are trained using the simulated dataset. The cloud removal effect of "SCM-CNN" is generally better than other methods, and "SpA-GAN" has an excellent cloud removal effect, and there is almost no color difference between slices. The cloud removal effect of "Dark Channel" is second only to "SpA-GAN". Below that, the image changes with "Dark Channel" and "SpA-GAN"; although the clouds are removed, the overall amount of information changes significantly, and brighter objects such as roads and snow become uniform. Neither "Dark Channel" nor "SpA-GAN" can accurately obtain the cloud mask and cannot achieve post-filtering in thick cloud regions, seriously impacting its practical application. In contrast, "SCM-CNN" can generate better cloud opacity information and removal results, and image colors remain pristine.

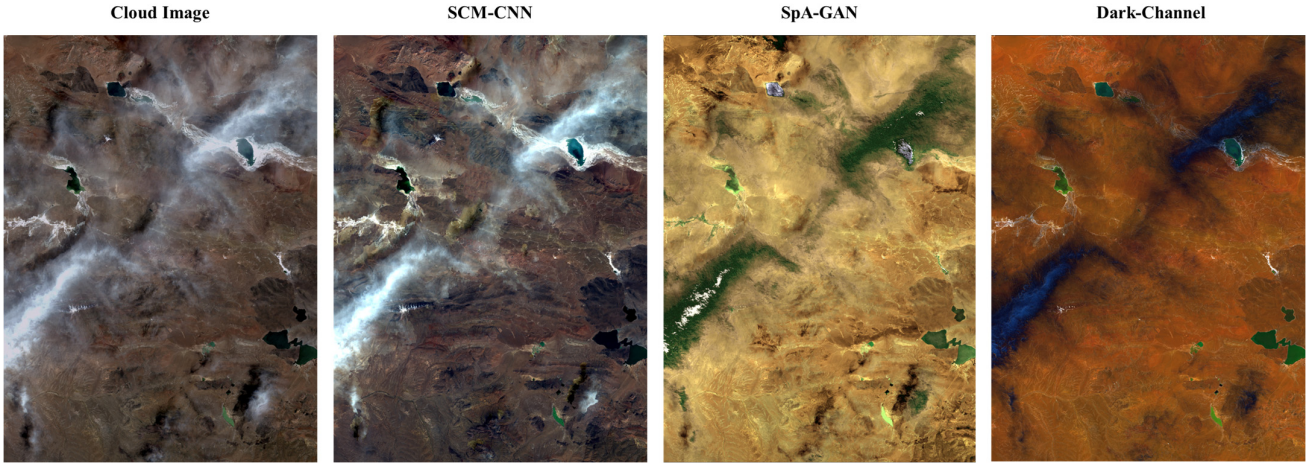


Figure 4. Sentinel-2B observation image and its cloud removal results from orbit 119 on 29 June 2020. This image contains thick and thin clouds and complex information such as snow, mountains, and cloud shadows, which could be used to test each model's accuracy and reliability effectively. Among them, the predicted images from SpA-GAN and Dark Channel are too dark, thus "normalized pixel stretching" is used for comparison. The original image refers to the attached illustration A2.

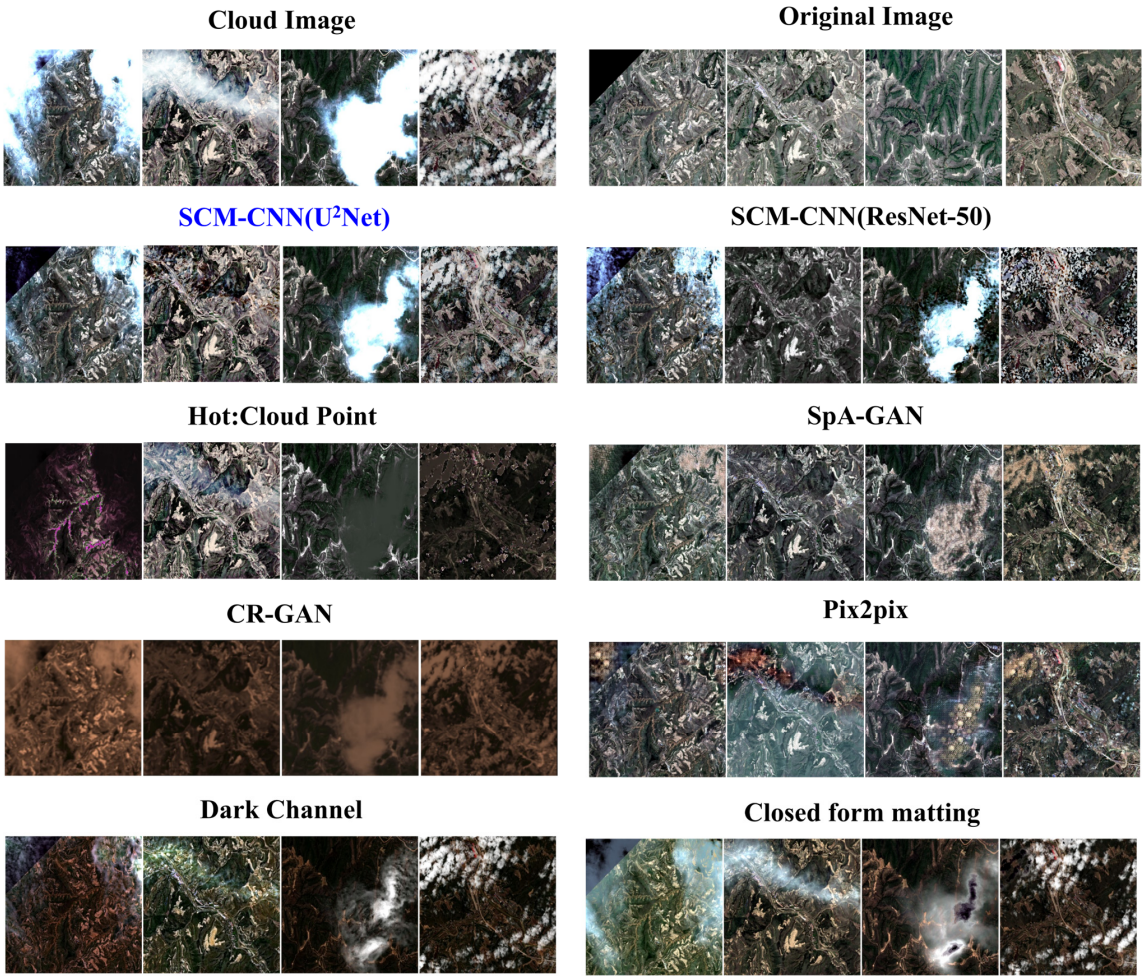
Due to differences in time, atmosphere, lighting, and other factors in each RS image, the hue and saturation of the image may vary significantly. Methods such as histogram matching can make image colors look similar, but they also reduce the realism of image elements. Therefore, considering the above factors, Table 1 only uses the "Cloud Image" as the noise image and the prediction result as the denoising result to measure the cloud removal effect of the image roughly with the PNSR function. However, the evaluation score is very different from human visual perception.

Table 1. The score of psnr. We utilized the WHUS2-CR dataset to validate the accuracy of our model on actual remote sensing images [29]. Specifically, we selected a subset of 10 high-quality image pairs with and without clouds for comparative analysis.

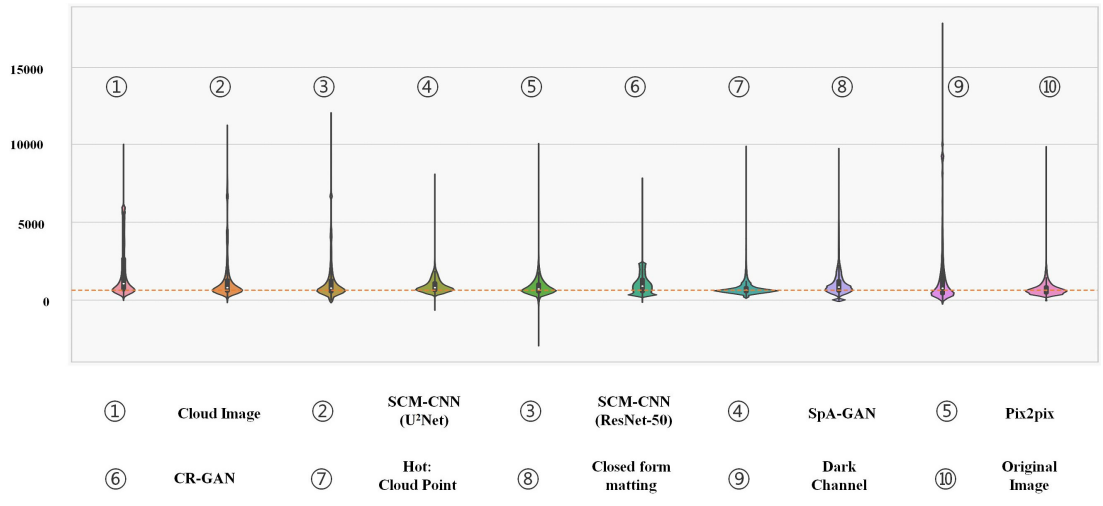
	SCM-CNN	SpA-GAN	Pix2pix	CRGAN	Hot	Dark Channel
PSNR	25.32	18.94	11.16	15.52	13.756	12.128

Figure 4 shows the usefulness of SCM-CNN for removing real RS image clouds, but evaluating the quality of RS image cloud removal results is challenging. Therefore, this work focuses on the validation of simulation data sets.

In Figure 5a, the "Trimap" was employed as an input to direct the "Closed-form matting" method. The "Dark Channel" approach utilized a 15 × 15 filtering window for dark visual elements. "SpA-GAN", "CR-GAN", and "Pix2pix" are variants of CGAN (Condition Generative Adversarial Networks), with "SpA-GAN" and "CR-GAN" incorporating an attention mechanism. In this paper, the "SCM-CNN" is trained to utilize the U^2Net and ResNet-50 backbone networks, respectively. All cloud removal techniques, including deep learning and non-deep-learning methods, can produce superior results when only thin clouds are in the image. However, when thick and thin clouds coexist in the image, the accuracy of "CGANs" (SpA-GAN, Pix2pix, CR-GAN) decreases significantly. "HOT", "Dark-channel", "Closed-form matting", and "SCM-CNN" can still assess the thin cloud zone more accurately to recover surface details.



(**a**)



(**b**)

Figure 5. (**a**) Comparison of cloud removal results of various methods applied to RS image slices. (**b**) The violin diagram depicts the pixel distribution of cloud removal outcomes generated by various approaches and illustrates the features of pixel dispersion. The orange dashed line indicates the primary image element distribution of the original image and the primary image distribution of cloudy RS images. The primary distribution of cloud removal findings should be located on this dashed line.

4. Discussion

Due to cloud contamination, the local pixel features of the synthesized cloudy image are altered in Figure 5b, and the overall image has fewer dark features than bright features. However, the image elements in the uncontaminated region will preserve their original distribution characteristics. Consequently, the primary elements continue to be dispersed around the orange dashed line. Based on the various model outputs, we can make the following preliminary assessment:

1. The "Hot" method adjusts the pixel brightness distribution by setting the unmistakable skyline. Therefore, the model is more sensitive to the ground feature, leading to variable variations in the unmistakable skyline, and it is prone to over-correction and color distortion.
2. Because of differences between traditional and satellite RS images (Refer to Figure A2 for details). "Dark Channel" will cause the image's overall color to darken.
3. "Closed-form matting" demonstrates the feasibility of matting in remote sensing image cloud removal, which uses the color line model and ridge regression optimization algorithm. However, this method requires an accurate "Trimap" as an a priori input, significantly limiting the applicable scenarios.
4. "SpA-GAN", "Pix2pix", and "CR-GAN" are advanced conditional generative adversarial models. However, the most crucial point is that partial pixel loss may lead to multiple entities may be invisible in RS images. Therefore, the generative adversarial network can show significant distortion in areas covered by thick clouds.
5. "SCM-CNN (U^2Net)" and "SCM-CNN (ResNet-50)" receive good cloud removal results, with almost no color deviation from the original image. The reason is that "SCM-CNN" is based on image-matting, which embedded cloud detection. The cloud removal results of the "SCM-CNN (U^2Net)" model are more reliable. Additionally, with more accurate cloud opacity estimation and less patchiness. This further demonstrates the superiority of our chosen saliency detection network.

To summarize, the different approaches to cloud removal substantially alter the values of image elements, even in the areas of the image that are free from clouds. Consequently, meeting the requirements for secondary production is a challenging task. On the contrary, the "SCM-CNN" technique identifies clouds preferentially, produces cloud opacity and CMaxDN data, and then employs the Image matting formula to restore the image after removing the clouds. Therefore, the outcomes are highly dependable and do not alter the image elements in the cloud-free regions.

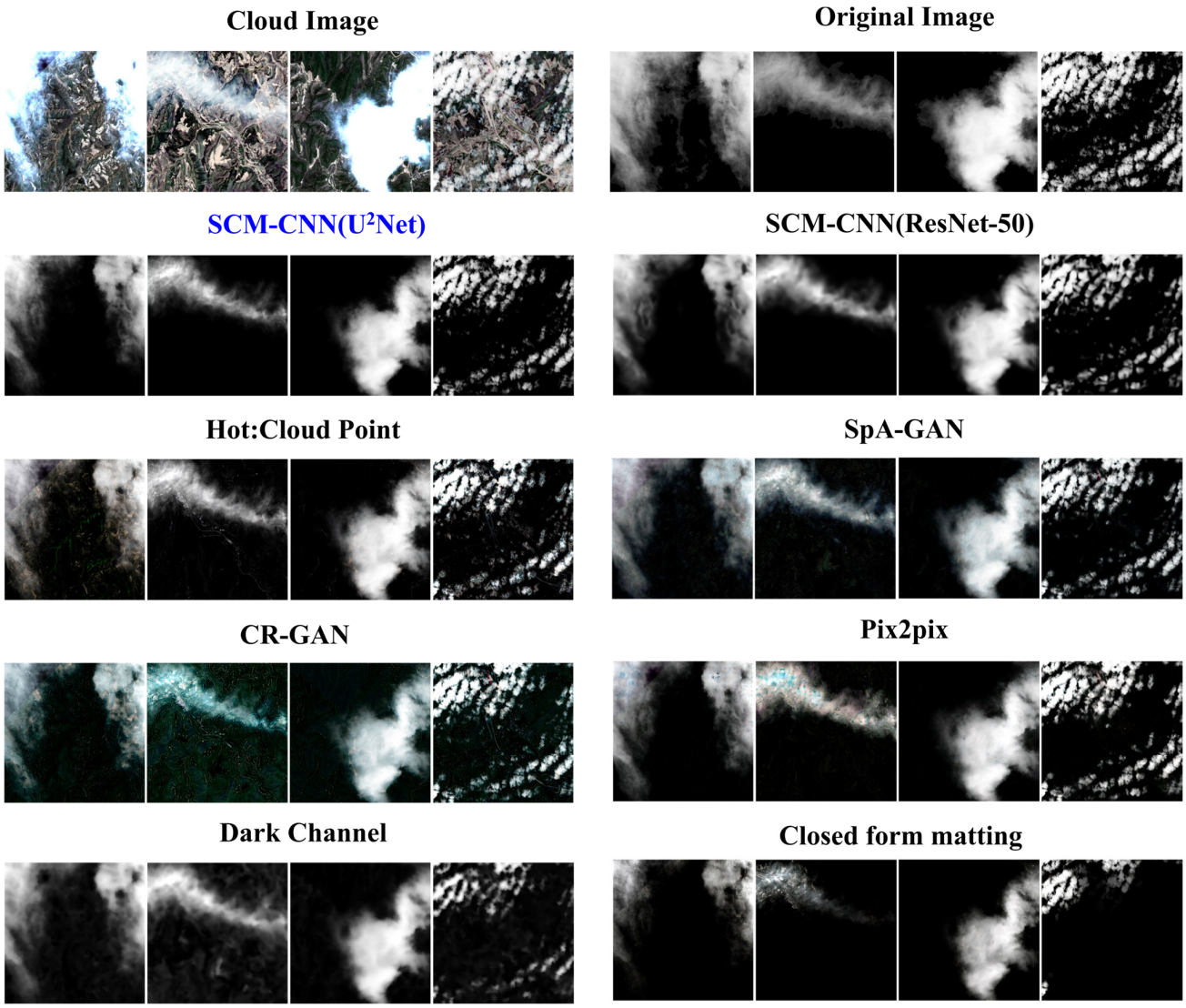
Figure 6a,b represent the results from the estimated image of cloud opacity to intuitively evaluate the quality of image de-clouding. Due to Equation (3), theoretically, all the calculated α values are greater than or equal to 0. However, the "Hot", "Dark Channel", and "CGANs" will have unreasonable values, and we set the $\alpha > 1$ values to 1 and the $\alpha < 0$ values to 0. Since the noisy background of the image no longer limits it, it can reflect the effect of model cloud removal more intuitively.

Visual observation can evaluate from two essential points: 1. The purity of the image. 2. The light and dark changes of the image. The purity of the image reflects the effect of foreground rejection during the cloud opacity operation, and the higher the purity the lower the noise of the cloud removal result. On the contrary, the cloud removal process will significantly modify the background image elements. The light and dark changes in the image reflect the accuracy of the cloud opacity calculation, and the closer to the label the higher the accuracy. The combined performance is that Figure 6a has higher clarity and contrast, and Figure 6b's scatter points are clustered around the red line. With careful consideration of the purity and the light and dark changes of the images, "U^2Net" is better than "ResNet-50"as the backbone.

Nevertheless, biases based on visual interpretation, such as "SpA-GAN", can deceive human vision more effectively by replacing cloud-contaminated image elements with those of a similar hue. However, there is a significant discrepancy between the simulated and actual surface information. To precisely measure the cloud removal effect, we begin

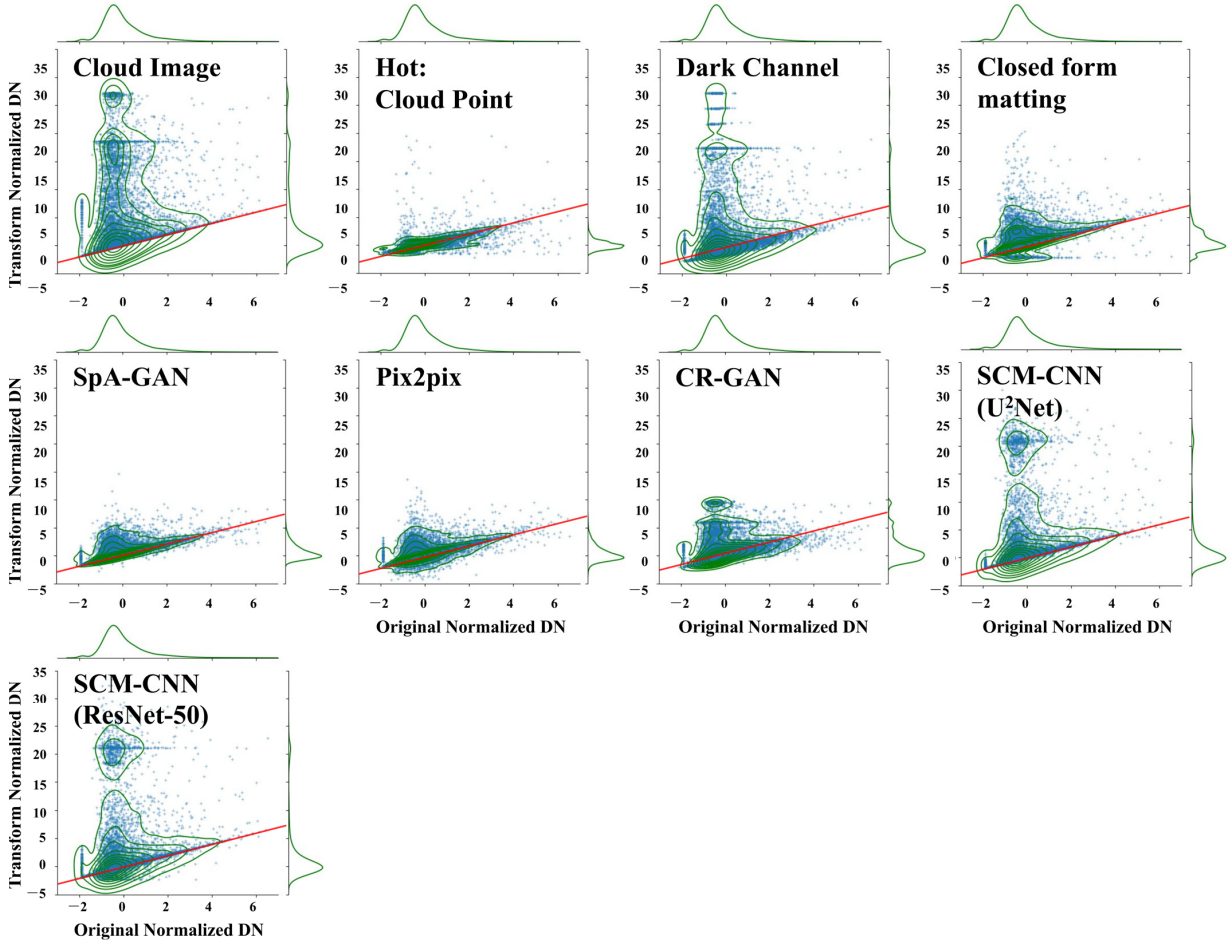
the mathematical statistic comparison by validating the cloud removal and cloud opacity estimation results.

We developed only 20 sets of typical image slices for statistical purposes because the Hot technique requires manual screening operations and is less automated. Other methods employed 640 sets of image slices to assess the model's performance based on image element similarity RMSE, structural similarity SSIM, peak signal-to-noise ratio PSNR, and classification accuracy ACC, respectively. The evaluation outcomes are presented in Table 2 and Figure 7.



(**a**)

Figure 6. *Cont.*



(**b**)

Figure 6. (**a**) Comparison of several cloud opacity estimating techniques. SCM-CNN and Closed-form matting generate the cloud opacity layer directly, whereas Dark Channel is based on the transmissivity layer back-calculated to obtain the cloud opacity layer. The remaining methods are based on cloud removal results using the actual CMaxDN values directly back-calculated to obtain. (**b**) Results of cloud opacity estimation against real values for various cloud removal techniques. One hundred thousand points are selected at random from the results of the opacity estimation, and their distributions are calculated. The Figure's red line represents the ideal estimate; the closer the data distribution is to the line the more trustworthy the cloud removal results. The green curve depicts the probability density of the dispersed dots.

Table 2 displays the various classification metrics produced for cloud removal and opacity estimation images using various methods. "CGANs" map the cloudy image element into the cloud-free image distribution characteristics, minimizing the difference between the cloudy image element and the cloud-free image element. However, the reflection characteristics of the image element are not identical to the semantic information. Therefore, "GANs" contribute to the deceptive image element of the metrics and the low reliability of the direct measurement of cloud removal findings. Consequently, we perform the metrics operation based on each model cloud opacity image. The estimations of cloud opacity, namely, RMSE (Alpha), SSIM (Alpha), and PSNR (Alpha), demonstrate that the efficacy of "SCM-CNN (U^2Net)" surpasses that of other methods by a significant margin, while "SCM-CNN (ResNet-50)" also achieves commendable results. On the other hand, although Hot and CGANs exhibit better RMSE (Alpha) scores in terms of image element similarity, their performance in the structural similarity index SSIM (Alpha) is generally subpar.

Table 2. Comparison of classification metrics for multiple cloud removal methods. Two metrics were obtained for each cloud removal method, representing the method's optimal and average values. By comparing multiple methods, we use green and blue to denote the best metrics values obtained by each method.

	RMSE ↓ (Image)	SSIM ↑ (Image)	PSNR ↑ (Image)	RMSE ↓ (Alpha)	SSIM ↑ (Alpha)	PSNR ↑ (Alpha)	ACC101 ↑ (Alpha)	ACC11 ↑ (Alpha)
Hot:	0.0091	0.9858	40.7270	0.0155	0.9878	36.1503	0.5616	0.9306
Cloud Point	0.0458	0.8499	23.6225	0.0478	0.9113	28.7166	0.4159	0.7758
Dark	0.0233	0.8198	32.6296	0.0059	0.9928	44.5602	0.6593	0.9672
Channel	0.1234	0.4115	19.3394	0.0803	0.8171	23.8009	0.0865	0.4274
Closed	0.0065	0.9942	43.6871	0.0159	0.9953	35.9338	0.8572	0.9738
form matting	0.1429	0.7418	20.0632	0.1141	0.8588	21.1993	0.5497	0.7473
SpA-GAN	0.0121	0.9959	44.1723	0.0071	0.9941	43.1151	0.8302	0.9761
	0.1098	0.8321	26.7704	0.0314	0.8616	30.5172	0.5476	0.8522
Pix2Pix	0.0107	0.9888	39.3703	0.0175	0.9902	35.0956	0.8159	0.9509
	0.0969	0.7942	24.4502	0.0437	0.9340	28.4119	0.4907	0.8353
CR-GAN	0.0141	0.9685	36.9748	0.0167	0.9820	35.5416	0.7971	0.9608
	0.1309	0.7687	21.3918	0.0596	0.8795	24.9252	0.4331	0.7058
SCM-CNN	0.0019	0.9991	54.0098	0.0061	0.9989	46.1551	0.8735	0.9874
(U^2Net)	0.1552	0.8465	30.0411	0.0262	0.9909	33.8204	0.6718	0.8990
SCM-CNN	0.0023	0.9984	52.4571	0.0077	0.9925	42.2312	0.8687	0.9792
(ResNet-50)	0.1731	0.7673	27.9535	0.0297	0.9626	32.1196	0.6389	0.8671

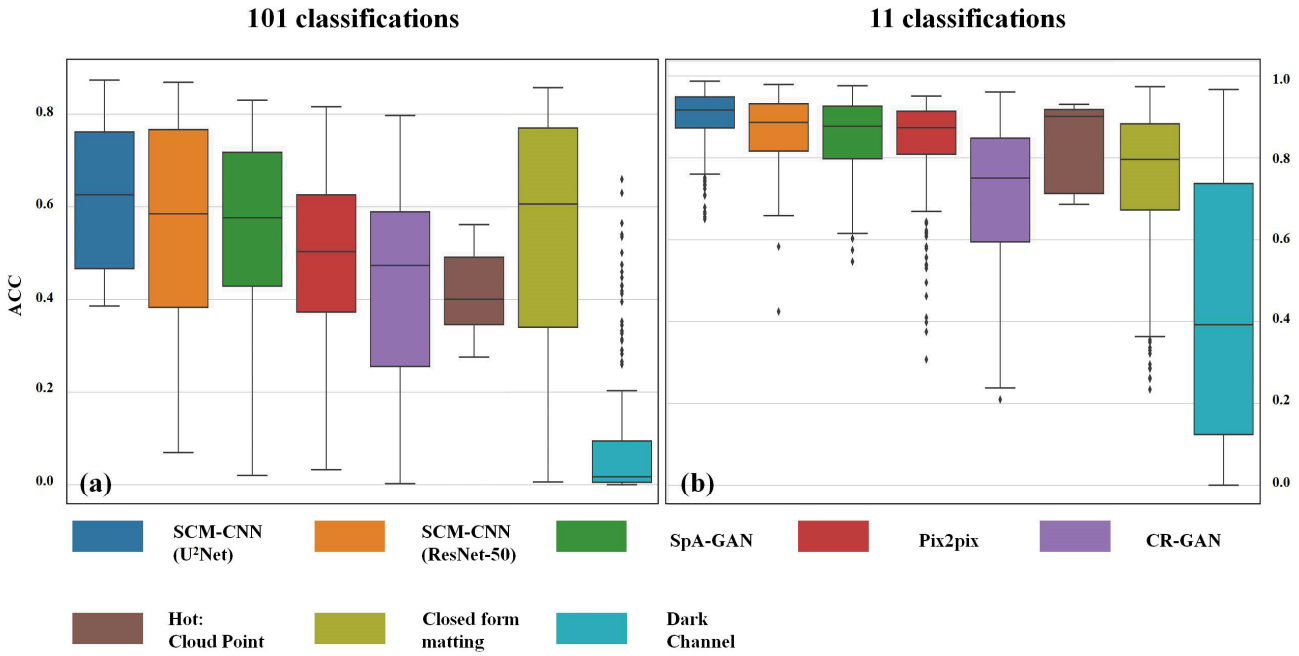


Figure 7. The results of Figure 6a are transformed into 101 (**a**) and 11 classes (**b**). We divided the predicted cloud opacity results into 101 classes and 11 classes based on intervals such as 0–1 and obtained the prediction pixel accuracy based on matching the prediction results to them. The Hot methods utilized 20 sets of image slices, whereas the others utilized 640 sets of image slices.

In addition, "Hot" and "CGANs" also modify many image elements in the original image's cloud-free region, resulting in many noise points in the cloud opacity estimation map. "SpA-GAN" output results are very similar to the global features, and the obtained ones are filled with a large number of ground surface texture features, but the pixel gap is small. Figure 7 divided the prediction results into 101 and 11 classes based on intervals from 0–1, and then obtained the prediction pixel accuracy based on matching the prediction results to labels. In Figure 7, the ACC101 (Alpha) shows that "SCM-CNN" has the best performance, while CGANs and Closed-form matting metrics are unstable. In addition, ACC11 (Alpha) shows that each method of cloud removal possesses better accuracy and the probability density of CGANs is more concentrated on a particular part.

The disadvantage of "Dark Channel" transmittance estimation is that it does not conform to the imaging mechanism of RS images and will enhance or weaken the image according to the image element brightness. The "Hot" is more sensitive to the feature, leading to the apparent skyline variation. The magnitude is variable, and all the image elements that deviate from the clear skyline must return to the clear skyline, but this step introduces much noise. The methods employed in this study for image generation are known as conditional generative adversarial networks (CGANs), which are currently considered state-of-the-art. The comparison of the three CGANs utilized in this study, SpA-GAN, CR-GAN, and Pix2pix, indicates that the overall effect of SpA-GAN is greater than that of CR-GAN and Pix2pix. This is primarily due to the attention mechanism introduced by SpA-GAN and CR-GAN. For more details, please refer to Figure A5, which presents the attention images under different conditions. However, the models cannot distinguish between thin and thick clouds. When complex meteorological conditions occur, the model results appear to be "fabricated," reducing the realism of the cloud removal results. In contrast, our proposed "SCM-CNN" is based on cloud detection, which achieves cloud removal by estimating clouds' opacity and maximum reflectance in RS images without modifying the image elements in cloud-free areas.

In addition, it is worth mentioning that as the cloud opacity increases, the confidence level of the surface image elements recovered results will gradually decrease. Depending on the scenario, "SCM-CNN" can be artificially set as a threshold in practical applications to assist human interpretation and data processing (Figure 8).

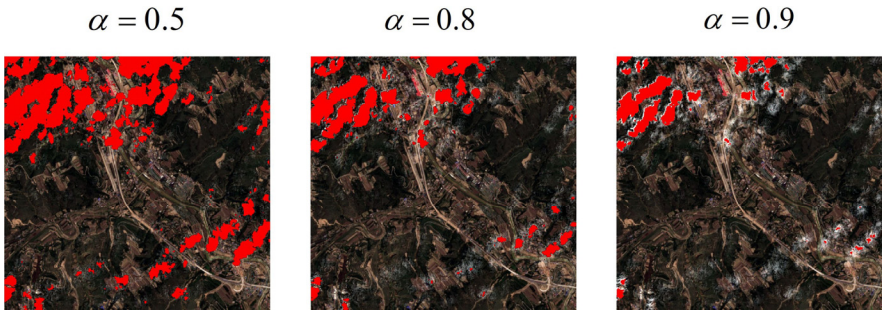


Figure 8. Setting different cloud opacity thresholds and overlaying the cloud removal results. Effectively improves the application of cloud removal results by masking the areas with lower confidence regions and taking $\alpha = 0.5$ in this image yields optimal results.

"SCM-CNN" demonstrates considerable generalization and stability in removing clouds from a single image, but it still has some shortcomings. Firstly, it is worth noting that the Sentinel-2 image data are Uint16, and the upper limit of image reflection brightness is variable. To normalize the training sample, we divide the atmospherically corrected Sentinel-2 image by 10,000. However, since the reflection value of the data obtained in this way is generally small, we will use "z-score" normalization to organize the data in the subsequent study to improve its universality. Secondly, the simulated data resemble natural remote sensing imagery, and the model can be effectively transferred to natural Sentinel-2 satellite imagery cloud scenes. However, the model presents some blurry artifacts in

the cloud shadow areas of remote sensing images due to shadow issues being neglected. Thirdly, to achieve accurate estimation and removal of cloud opacity, this paper employs a backbone network based on saliency detection and a multi-scale feature extraction and fusion strategy. Compared with existing algorithms for thin cloud removal, SCM-CNN can remove thin clouds better under the condition of coexisting thin and thick clouds. Therefore, it has good applicability for the frozen zone with severe fog and cloud cover. However, the model still exhibits some misjudgment when clouds are located above highly reflective ice and snow, typically recognizing highly reflective snow as thick clouds.

In the following research, we propose adopting strategies to optimize the algorithm: 1. Replace the CNN network with the Transformer network, which has performed better in recent years to achieve the attention mechanism and multi-scale feature fusion [52]. 2. Collect heterogeneous region image base map to improve the robustness of the model. 3. Fuse more wavebands and other auxiliary information (DEM, DSM, and others.) from RS images into the model to achieve higher accuracy of cloud-snow differentiation. 4. Replacing "Channel Global Average & Max Pooling" with an MLP-Mixer may lead to better generalization and accuracy improvements [53].

5. Conclusions

In this research, we approached the subject of single-frame RS image cloud removal from the image-matting angle. Moreover, we discussed the principles, advantages, and disadvantages of various single-frame image cloud removal methods. We established an open-source "SCM-CNN" model and supporting data.

We can draw the following conclusions from the research results: 1. Single-frame RS image cloud removal can only recover the surface information covered by thin clouds. Then, improving model stability in thick and thin cloud scenes is very important. 2. "SCM-CNN" adopts a saliency detection network, multi-scale extraction, and the fusion of image features to achieve high-precision cloud opacity generation. Furthermore, the cloud removal process is more consistent with the imaging mechanism of RS images. 3. "SCM-CNN" is based on cloud detection, and the results of cloud removal do not interfere with the original images of cloud-free areas. 4. A synthetic data set is created. All cloud removal methods can operate efficiently on our datasets. 5. The effect of the "SCM-CNN" method is significantly better than other comparison methods. It is worth mentioning that CGANs' image element reconstruction ability is powerful, thus it is easy to obtain similar but not identical image elements. 6. Changing the conditions of "CGANs" can also obtain better cloud opacity estimation results (Figures A3 and A4). Thus, cloud matting can be effectively migrated to other types of deep learning models.

Above all, "SCM-CNN" can effectively build the cloud opacity information from single-frame RS images. Moreover, "SCM-CNN" demonstrates good anti-interference performance in the coexistence of thick and thin clouds. Cloud matting is very helpful for remote sensing image processing in cloudy areas, and we will continue to intensify our efforts in this area.

Author Contributions: Conceptualization, G.L. and R.Z.; methodology, R.W.; validation, J.C. and B.S.; formal analysis, X.B. and Y.F.; data curation, R.W., J.L. and A.S.; writing—original draft preparation, R.W.; writing—review and editing, R.W., X.W. and R.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was jointly funded by the National Natural Science Foundation of China (Grant 42171355 and Grant 42071410); the Sichuan Science and Technology Program (No. 2021YFH0038, 2018JY0564, 2019ZDZX0042, 2020JDTD0003); and the Southwest University of Science and Technology Doctoral Fund (Grant 22ZX7171).

Data Availability Statement: "SCM-CNN" code URL: https://github.com/wurenzhe163/SCM-CNN (accessed on 2 December 2022); Cloud matting dataset URL: https://doi.org/10.5281/zenodo.7188292 (accessed on 2 December 2022).

Acknowledgments: Renzhe Wu Author thanks Xuebin Qin, author of U^{2}Net. U^{2}Net is a very effective convolutional neural network model and has greatly aided our research.

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

Appendix A

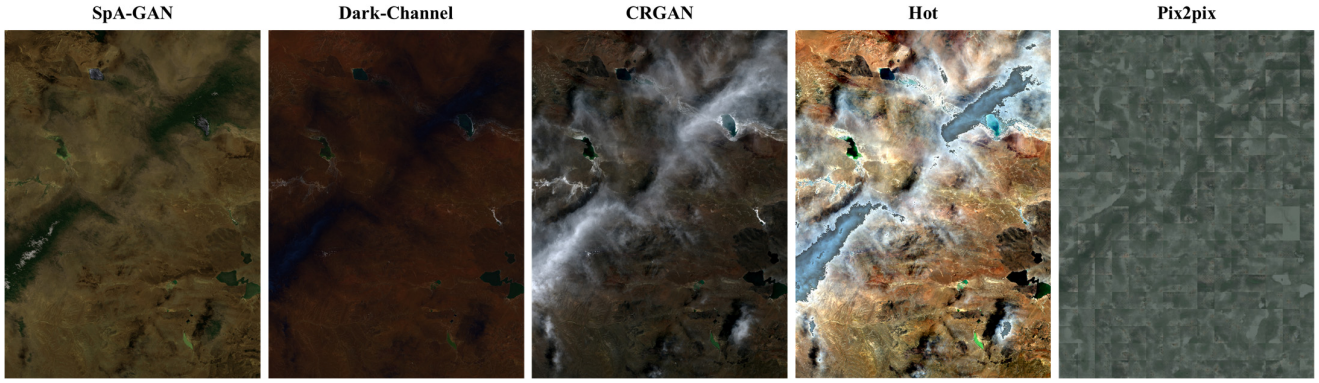


Figure A1. Displays cloud removal results from the comparison method. SpA-GAN and Dark Channel show the absolute luminance value after image rendering.

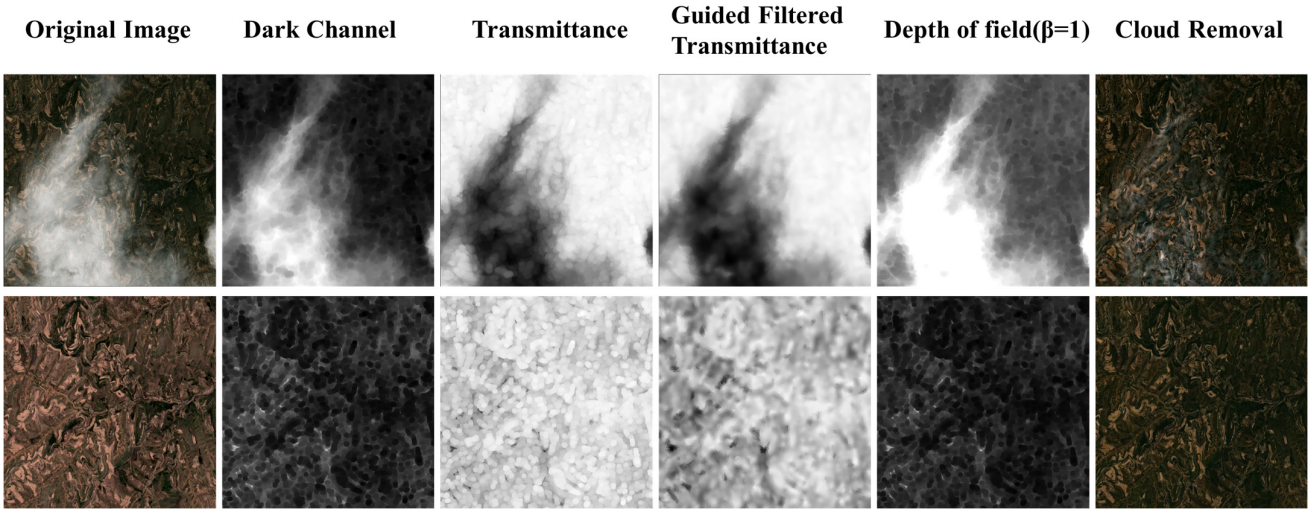


Figure A2. Intermediate results of the Dark Channel method for de-clouding.

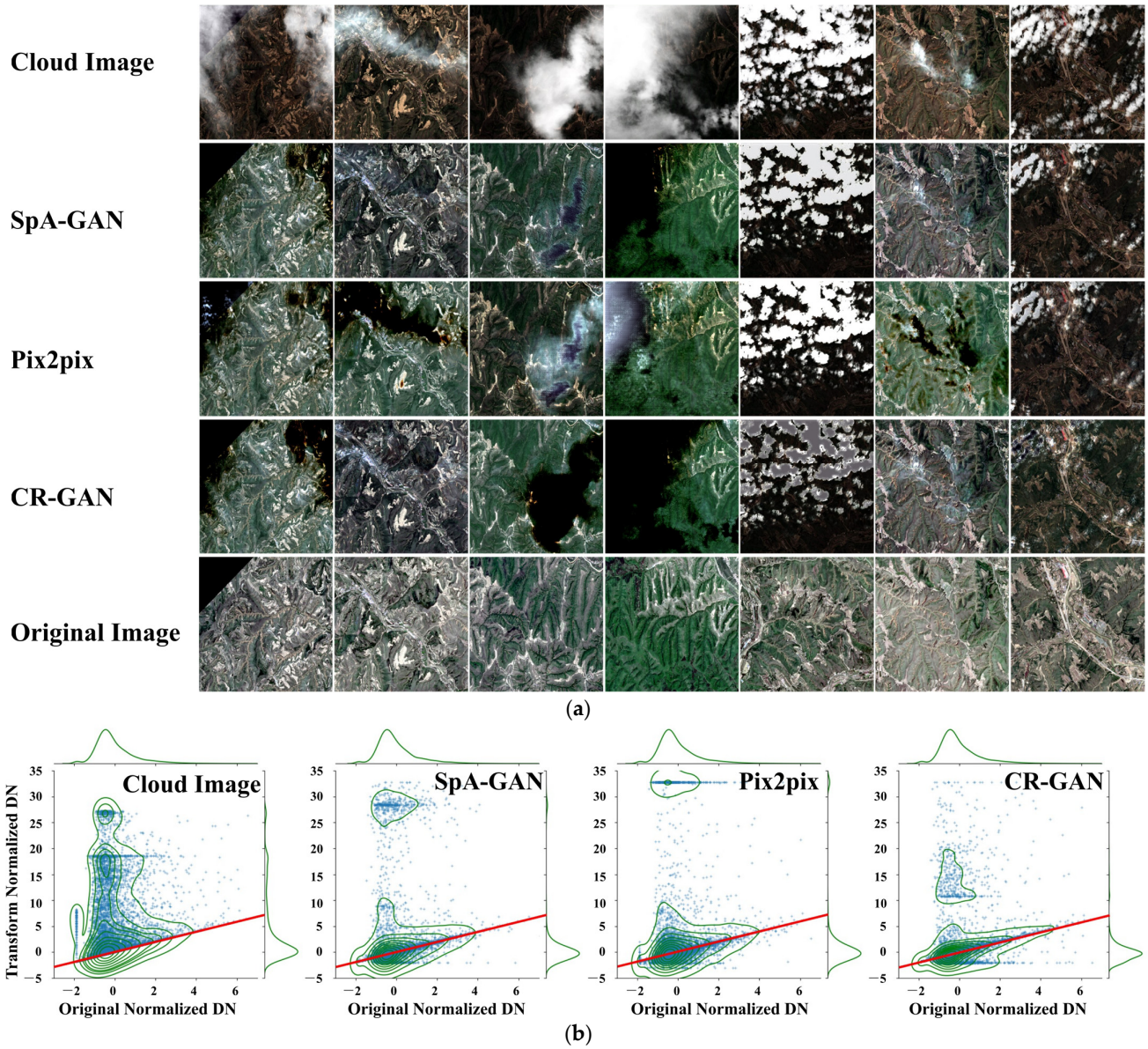


Figure A3. (**a**) CGANs after changing the conditions. (**b**) Cloud removal results from various methods, and cloud-free images are compared for pixel distribution characteristics. We first use normalization to convert the image to a normal distribution, and then we randomly sample the distribution of 100,000 points statistics in the normalized image due to the wide range of data distribution. The red line represents the best estimate; points below are considered overcorrected, and those above are considered undercorrected. The green curve represents the probability density of the scattered points.

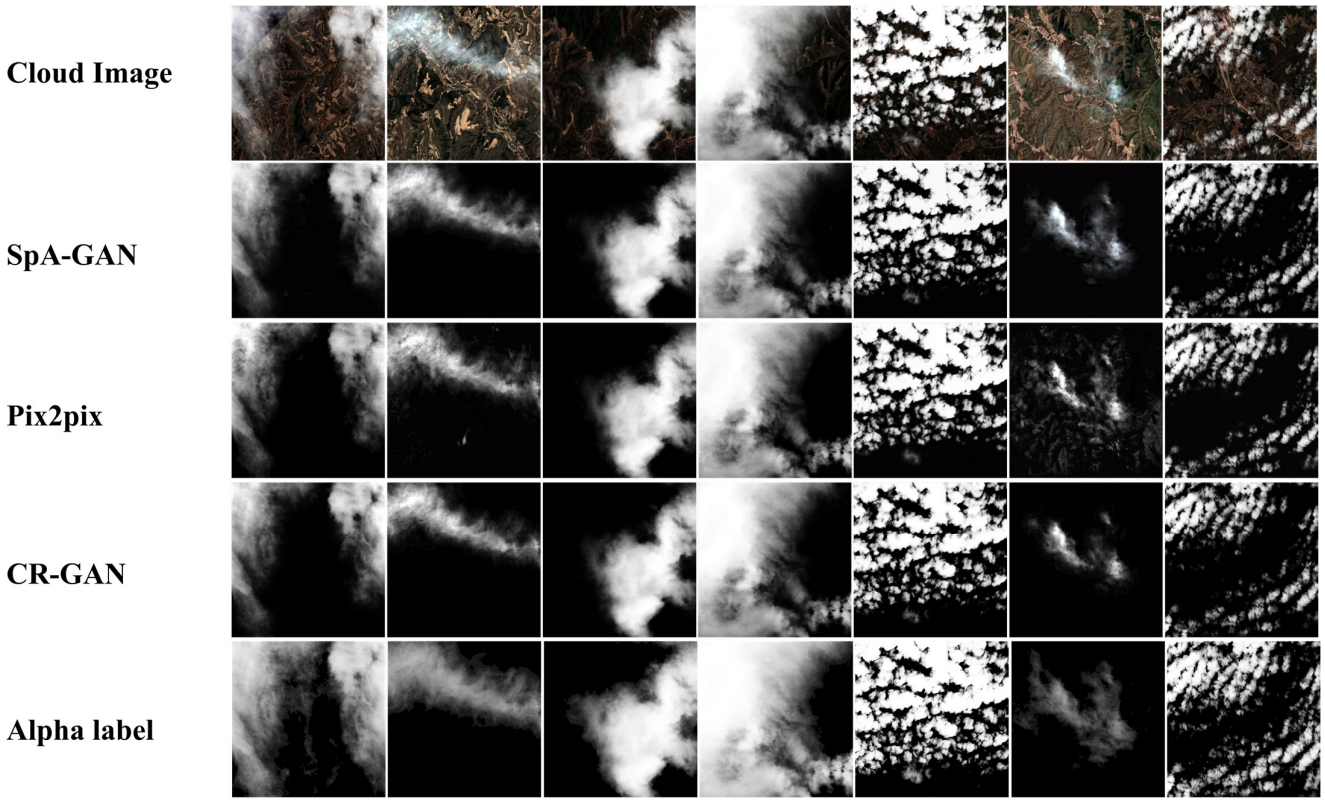


Figure A4. (**a**) Results of cloud opacity estimation of CGANs after changing the conditions, and the elements in Figure are consistent with Figure A3a. (**b**) Cloud opacity estimation results compared to actual values for various cloud removal methods. One hundred thousand points are randomly sampled from the opacity estimation results, and their distributions are counted. The red line in the figure represents the optimal estimate, and the closer the data distribution is to the line the more reliable the cloud removal results are. The green curve represents the probability density of the scattered points.

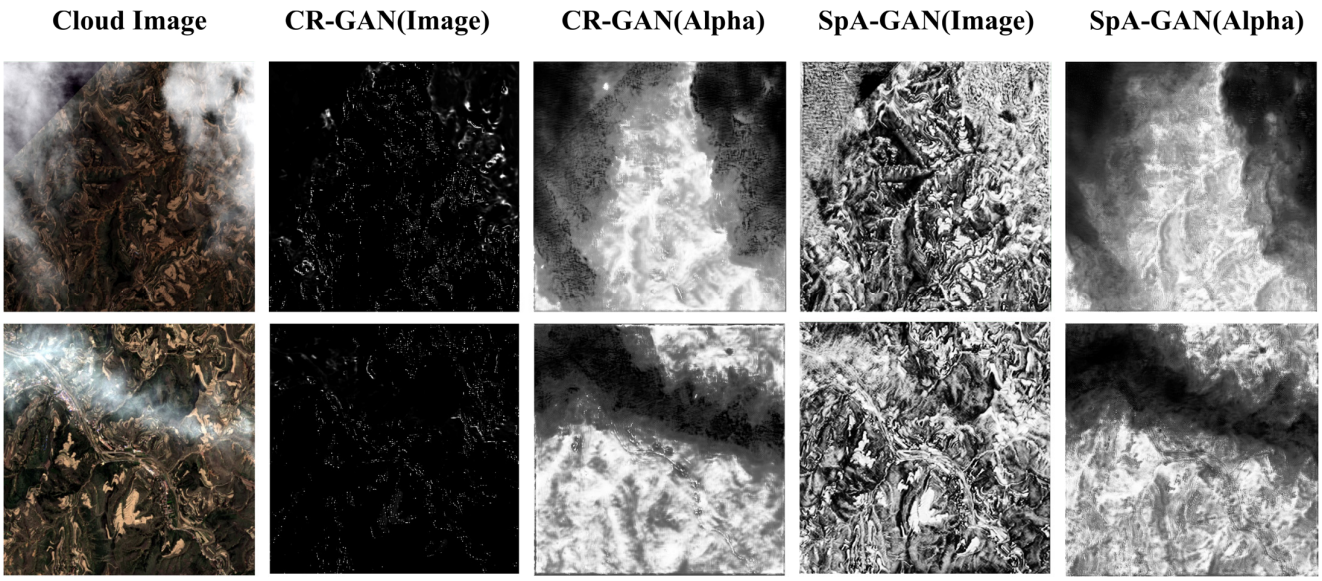


Figure A5. Visualization of the attention mechanism from CR-GAN and SpA-GAN networks under different conditions.

References

1. Efremenko, D.; Kokhanovsky, A. *Introduction to Remote Sensing BT—Foundations of Atmospheric Remote Sensing*; Efremenko, D., Kokhanovsky, A., Eds.; Springer International Publishing: Cham, Switzerland, 2011; ISBN 978-3-030-66745-0.
2. Inglada, J.; Vincent, A.; Arias, M.; Tardy, B.; Morin, D.; Rodes, I. Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sens.* **2017**, *9*, 95. [CrossRef]
3. Rossow, W.B.; Schiffer, R.A. Advances in Understanding Clouds from ISCCP. *Bull. Am. Meteorol. Soc.* **1999**, *80*, 2261–2287. [CrossRef]
4. Zhang, Y.; Rossow, W.B.; Lacis, A.A.; Oinas, V.; Mishchenko, M.I. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *J. Geophys. Res. Atmos.* **2004**, *109*, 1–27. [CrossRef]
5. Wang, Y.; Yuan, Q.; Li, T.; Shen, H.; Zheng, L.; Zhang, L. Large-scale MODIS AOD products recovery: Spatial-temporal hybrid fusion considering aerosol variation mitigation. *ISPRS J. Photogramm. Remote Sens.* **2019**, *157*, 1–12. [CrossRef]
6. Shen, H.; Li, X.; Cheng, Q.; Zeng, C.; Yang, G.; Li, H.; Zhang, L. Missing Information Reconstruction of Remote Sensing Data: A Technical Review. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 61–85. [CrossRef]
7. Pan, X.; Xie, F.; Jiang, Z.; Yin, J. Haze Removal for a Single Remote Sensing Image Based on Deformed Haze Imaging Model. *IEEE Signal Process. Lett.* **2015**, *22*, 1806–1810. [CrossRef]
8. Xie, F.; Chen, J.; Pan, X.; Jiang, Z. Adaptive haze removal for single remote sensing image. *IEEE Access* **2018**, *6*, 67982–67991. [CrossRef]
9. Zhang, Q.; Yuan, Q.; Li, J.; Li, Z.; Shen, H.; Zhang, L. Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 148–160. [CrossRef]
10. Li, X.; Shen, H.; Zhang, L.; Zhang, H.; Yuan, Q.; Yang, G. Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7086–7098. [CrossRef]
11. Xu, M.; Jia, X.; Pickering, M.; Plaza, A.J. Cloud removal based on sparse representation via multitemporal dictionary learning. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2998–3006. [CrossRef]
12. Chen, B.; Huang, B.; Chen, L.; Xu, B. Spatially and Temporally Weighted Regression: A Novel Method to Produce Continuous Cloud-Free Landsat Imagery. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 27–37. [CrossRef]
13. Zhang, Y.; Wen, F.; Gao, Z.; Ling, X. A Coarse-to-Fine Framework for Cloud Removal in Remote Sensing Image Sequence. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5963–5974. [CrossRef]
14. Pelletier, C.; Webb, G.I.; Petitjean, F. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sens.* **2019**, *11*, 523. [CrossRef]
15. Li, X.; Wang, L.; Cheng, Q.; Wu, P.; Gan, W.; Fang, L. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. *ISPRS J. Photogramm. Remote Sens.* **2019**, *148*, 103–113. [CrossRef]
16. Liang, S.; Fang, H.; Morisette, J.T.; Chen, M.; Shuey, C.J.; Walthall, C.L.; Daughtry, C.S.T. Atmospheric correction of Landsat ETM+ land surface imagery. II. Validation and applications. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2736–2746. [CrossRef]
17. Cao, S.; Li, H.; Ma, W. Removing thin cloud arithmetic based on mathematic morphology for remote sensing image. *Geopraphy Geo-Inf. Sci.* **2009**, *4*, 30–33.

18. Cai, W.; Liu, Y.; Li, M.; Cheng, L.; Zhang, C. A Self-Adaptive Homomorphic Filter Method for Removing Thin Cloud. In Proceedings of the 2011 19th International Conference on Geoinformatics, Shanghai, China, 24–26 June 2011; pp. 1–4.
19. Rossi, R.E.; Dungan, J.L.; Beck, L.R. Kriging in the shadows: Geostatistical interpolation for remote sensing. *Remote Sens. Environ.* **1994**, *49*, 32–40. [CrossRef]
20. Zhu, X.; Gao, F.; Liu, D.; Chen, J. A modified neighborhood similar pixel interpolator approach for removing thick clouds in landsat images. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 521–525. [CrossRef]
21. Bertalmio, M.; Vese, L.; Sapiro, G.; Osher, S. Simultaneous Structure and Texture Image Inpainting. *IEEE Trans. Image Process.* **2003**, *12*, 882–889. [CrossRef]
22. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341. [CrossRef]
23. Zhang, Y.; Guindon, B.; Cihlar, J. An image transform to characterize and compensate for spatial variations in thin cloud contamination of Landsat images. *Remote Sens. Environ.* **2002**, *82*, 173–187. [CrossRef]
24. Emami, H.; Aliabadi, M.M.; Dong, M.; Chinnam, R.B. Spa-gan: Spatial attention gan for image-to-image translation. *IEEE Trans. Multimed.* **2020**, *23*, 391–401. [CrossRef]
25. Xu, M.; Deng, F.; Jia, S.; Jia, X.; Plaza, A.J. Attention mechanism-based generative adversarial networks for cloud removal in Landsat images. *Remote Sens. Environ.* **2022**, *271*, 112902. [CrossRef]
26. Ramjyothi, A.; Goswami, S. Cloud and Fog Removal from Satellite Images Using Generative Adversarial Networks (GANs). 2021. Available online: https://hal.science/hal-03462652 (accessed on 12 January 2023).
27. Pan, H. Cloud Removal for Remote Sensing Imagery via Spatial Attention Generative Adversarial Network. *arXiv* **2020**, arXiv:2009.13015. [CrossRef]
28. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; Volume 2017-Janua, pp. 5967–5976.
29. Li, J.; Wu, Z.; Hu, Z.; Li, Z.; Wang, Y.; Molinier, M. Deep learning based thin cloud removal fusing vegetation red edge and short wave infrared spectral information for sentinel-2A imagery. *Remote Sens.* **2021**, *13*, 157. [CrossRef]
30. Lin, S.; Ryabtsev, A.; Sengupta, S.; Curless, B.L.; Seitz, S.M.; Kemelmacher-Shlizerman, I. Real-Time High-Resolution Background Matting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 21–24 June 2021; pp. 8762–8771.
31. Sun, Y.; Tang, C.-K.; Tai, Y.-W. Semantic Image Matting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 21–24 June 2021; pp. 11120–11129. [CrossRef]
32. Qiao, Y.; Liu, Y.; Zhu, Q.; Yang, X.; Wang, Y.; Zhang, Q.; Wei, X. Multi-scale Information Assembly for Image Matting. *Comput. Graph. Forum* **2020**, *39*, 565–574. [CrossRef]
33. Xu, N.; Price, B.; Cohen, S.; Huang, T. Deep Image Matting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; Volume 2017-Janua, pp. 2970–2979. [CrossRef]
34. Chen, Q.; Ge, T.; Xu, Y.; Zhang, Z.; Yang, X.; Gai, K. Semantic Human Matting. In Proceedings of the MM 2018—Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 22–26 October 2018; pp. 618–626. [CrossRef]
35. Xu, Y.; Sun, B.; Yan, X.; Hu, J.; Chen, M. Multi-focus image fusion using learning based matting with sum of the Gaussian-based modified Laplacian. *Digit. Signal Process. A Rev. J.* **2020**, *106*, 102821. [CrossRef]
36. Khan, S.; Pirani, Z.; Fansupkar, T.; Maghrabi, U. Shadow Removal from Digital Images using Multi-channel Binarization and Shadow Matting. In Proceedings of the 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) , Palladam, India, 12–14 December 2019; pp. 723–728. [CrossRef]
37. Amin, B.; Riaz, M.M.; Ghafoor, A. Automatic image matting of synthetic aperture radar target chips. *Radioengineering* **2020**, *29*, 228–234. [CrossRef]
38. Golts, A.; Freedman, D.; Elad, M. Unsupervised Single Image Dehazing Using Dark Channel Prior Loss. *IEEE Trans. Image Process.* **2020**, *29*, 2692–2701. [CrossRef]
39. Li, W.; Zou, Z.; Shi, Z. Deep Matting for Cloud Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8490–8502. [CrossRef]
40. Jeppesen, J.H.; Jacobsen, R.H.; Inceoglu, F.; Toftegaard, T.S. A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sens. Environ.* **2019**, *229*, 247–259. [CrossRef]
41. Rhemann, C.; Rother, C.; Wang, J.; Gelautz, M.; Kohli, P.; Rott, P. A Perceptually Motivated Online Benchmark for Image Matting. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, Miami, FL, USA, 20–25 June 2009; pp. 1826–1833.
42. Shen, X.; Tao, X.; Gao, H.; Zhou, C.; Jia, J. Deep Automatic Portrait Matting. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9905 LNCS, pp. 92–107. [CrossRef]
43. Irish, R.R.; Barker, J.L.; Goward, S.N.; Arvidson, T. Characterization of the landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 1179–1188. [CrossRef]
44. Hughes, M.J.; Hayes, D.J. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. *Remote Sens.* **2014**, *6*, 4907–4926. [CrossRef]

45. Ebel, P.; Meraner, A.; Schmitt, M.; Zhu, X.X. Multisensor Data Fusion for Cloud Removal in Global and All-Season Sentinel-2 Imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5866–5878. [CrossRef]
46. Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O.R.; Jagersand, M. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit.* **2020**, *106*, 107404. [CrossRef]
47. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multi-Scale Structural Similarity for Image Quality Assessment. In Proceedings of the Conference Record of the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2.
48. Sener, O.; Koltun, V. Multi-task learning as multi-objective optimization. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 527–538.
49. Naik, A.; Rangwala, H. *Multi-Task Learning*; Springer: Cham, Switzerland, 2018; ISBN 1461375274.
50. Zhang, Y.; Yang, Q. An overview of multi-task learning. *Natl. Sci. Rev.* **2018**, *5*, 30–43. [CrossRef]
51. Xiao, C.; Liu, M.; Xiao, D.; Dong, Z.; Ma, K.-L. Fast closed-form matting using a hierarchical data structure. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *24*, 49–62. [CrossRef]
52. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 9992–10002.
53. Tolstikhin, I.O.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J. Mlp-mixer: An all-mlp architecture for vision. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 24261–24272.