Weakly Supervised Adversarial Training for Remote Sensing Image Cloud and Snow Detection

Jiajun Yang, Wenyuan Li, Keyan Chen, Zili Liu, Zhenwei Shi, *Senior Member, IEEE*, and Zhengxia Zou⋆, *Senior Member, IEEE*

***Abstract*—Cloud and snow detection in remote sensing images has advanced significantly with the aid of deep learning methods. However, deep learning methods necessitate a large quantity of labeled data, which consumes a substantial amount of human and material resources. Numerous studies have focused on weakly supervised methods to reduce the workload of annotation, but the majority of these methods concentrate on cloud detection and involve snow detection only infrequently. In this paper, we propose a novel weakly supervised cloud and snow detection (WCSD) method. Under the guidance of the remote sensing imaging mechanism, we design generative adversarial networks (GAN) to generate cloud and snow images and pseudo labels for training detection networks. The proposed method can generate clouds of different states and reproduce snow's texture. For both the cloud GAN model and snow GAN model, with only image-level annotation training supervision, the models produce both pixel-level cloud/snow reflectance and cloud opacity to obtain the generated remote sensing images and corresponding pseudo labels. Compared to other weakly supervised methods, our method achieves superior cloud and snow detection performance.**

***Index Terms*—Cloud and snow detection, deep learning, generative adversarial networks, weakly supervised learning, remote sensing images**

I. Introduction

REMOTE sensing technology has greatly expanded humanity's capacity to understand the earth. Remote sensing satellite image processing and analysis is playing an increasingly crucial role in modern agriculture, disaster prevention, and resource exploration [1–13]. Clouds and snow are common components in remote sensing images and are also limiting factors affecting ground feature analysis. For instance, ground objects are frequently obscured by clouds in multispectral remote sensing images. Studies have indicated that, at various times of the day, more than 60% of the world's surface is covered by clouds [14, 15]. Additionally, in some regions of high latitude or altitude, snow may always be present.

Cloud and snow detection is an essential step in the preprocessing phase of remote sensing image production. Numerous studies investigate how to rapidly extract cloud and snow coverage areas from remote sensing images using rule-based methods [16–23], machine learning-based methods [24–27], and deep learning methods [28–43].

The rule-based cloud and snow detection method mainly constructs relevant rules to detect clouds and snow based on the spectral information of different bands. The two most representative methods are the Automatic Cloud Cover Assessment (ACCA) [18, 19] and the Function of masks (Fmask) [20–23]. ACCA method uses the 2nd-6th bands of Landsat 7 ETM+ satellite images and sets different thresholds based on the reflection conditions of each band to extract cloud layers precisely. The Fmask method takes the 1st-7th bands as input and detects clouds and snow by quantitatively modeling the bands' relationship. The rule-based cloud and snow detection methods are simple to calculate and require few computing resources. They are generally based on the relationship between various spectra and have distinct physical meanings. However, in high latitudes and mountainous terrain, these methods have low detection accuracy. In addition, they require the input image to contain sufficient spectral information, which makes it difficult to apply to remote sensing images other than Landsat.

The machine learning-based methods [24, 26, 27] require the manual design of feature extraction algorithms in order to extract the distinguishing characteristics of clouds, snow, and ground objects. The machine learning-based detection methods first extract image features of clouds and snow such as visual texture from remote sensing images, and then use machine learning algorithms such as support vector machine [44] or random forest [45] to classify image pixels based on the extracted image features. The classic detection methods include brightness feature-based method [45, 46], texture feature-based method [24] and local statistical feature-based method [47, 48]. These methods are less stringent than the rule-based cloud and snow detection methods for the input bands. They also show higher detection accuracy than rule-based methods in complex regions. Despite the fact that machine learning-based methods have improved cloud and snow detection performance, they require a large number of manually designed image features. Subsequent development of deep learning methods makes it possible to automatically extract features and detect snow and clouds from remote sensing images.

The deep learning-based cloud and snow detection methods

The work was supported by the National Key Research and Development Program of China (Grant No. 2022ZD0160401), the National Natural Science Foundation of China under Grants 62125102 and 623B2013, the Beijing Natural Science Foundation under Grant JL23005, and the Fundamental Research Funds for the Central Universities. *(Corresponding Author: Zhengxia Zou (zhengxiazou@buaa.edu.cn)).*

Jiajun Yang, Keyan Chen, Zili Liu, and Zhenwei Shi are with the Image Processing Center, School of Astronautics, the Beijing Key Laboratory of Digital Media, the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China, and with the Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China.

Wenyuan Li is with the Department of Geography, the University of Hong Kong, Hong Kong 999077, China.

Zhengxia Zou is with the Department of Guidance, Navigation and Control, School of Astronautics, Beihang University, Beijing 100191, China.

[15, 28–39, 41, 42, 49, 50] have significantly improved the cloud and snow detection in remote sensing images. The early deep learning-based detection methods were realized through image block classification [28], and then image segmentation networks based on fully convolutional networks such as U-Net dominated the field of cloud and snow detection [32, 51]. The segmentation models based on Transformer further promoted the development of cloud and snow detection technology [52]. In addition, many weakly supervised cloud and snow detection methods [53, 54] have been proposed to solve the problem of insufficient high-quality labeled data. With a large number of images that are pixel-by-pixel labeled, the strong representation ability of deep learning methods significantly improves the cloud and snow detection accuracy.

However, deep learning methods usually require a large amount of labeled data, and the annotation process of remote sensing images consumes a lot of time and resources and relies on professionals. Moreover, clouds and snow usually exhibit very similar visual characteristics in remote sensing images, making it difficult to distinguish with previous methods.

To reduce the need for labeled data, weakly supervised learning provides an effective way of producing fine-grained results with coarse-grained labels, such as implementing pixel-by-pixel segmentation tasks using only the classification labels of images. In semantic segmentation [55–58], object localization [59–62], and etc., weakly supervised learning methods have produced favorable results. Class activation map (CAM) [63] is the most common method, which weights the features prior to global pooling with parameters of the full connection layer to obtain the response areas of various classes on the input image, thereby achieving segmentation or localization. The CAM method is simple but effective, which serves as the foundation for numerous weakly supervised learning methods [53, 54, 64–67].

Weakly supervised learning has also been developed for the problem of cloud detection in remote sensing images. Li et al. [54] proposed a CAM-based method for weakly supervised cloud detection in remote sensing images, and obtained favorable results. In addition to the CAM-based method, generative adversarial networks can be used to generate cloud remote sensing images and their labels for weakly supervised cloud detection. Zou et al. [53] proposed a weakly supervised cloud detection method with generative adversarial networks [68–70], which can generate cloud remote sensing images and their pseudo-cloud labels with the input of cloud images and background images. Li et al. [67] proposed a weakly supervised cloud detection method GAN-CDM based on synergistic combination of generative adversarial networks and a physics-based cloud distortion model. GAN-CDM model trained on Landsat images can provide accurate Landsat cloud detection results and has good Sentinel-2 image transferability. Liu et al. [71] proposed a weakly supervised cloud detection framework that uses the bidirectional threshold segmentation and adaptive gating mechanism to generate pseudo masks, which are then used as weak supervision to optimize the heuristic cloud detection network.

In satellite remote sensing images, clouds and snow often appear simultaneously, making accurate snow identification a significant challenge in cloud detection tasks. Because the visual characteristics of clouds and snow are highly similar, they are easily confused. We aim to establish a unified cloud and snow hybrid imaging mechanism that clearly defines both clouds and snow, helping neural networks learn more efficiently to generate visually and physically authentic images of clouds and snow. Based on this, the detection networks are trained jointly in a weakly supervised manner to distinguish between clouds and snow. The existing weakly supervised learning related works mainly focus on cloud detection task, rarely considering snow detection task. There has been a lack of efforts to address cloud and snow detection issues by deeply exploring the distribution relationships and imaging characteristics of clouds and snow in remote sensing images. Previous cloud detection methods often struggle to resolve the interference problem between snow and clouds when dealing with detection tasks involving cloud-snow mixed images.

In this paper, we propose a novel **W**eakly supervised **C**loud and **S**now **D**etection (WCSD) method for remote sensing images. Under the guidance of the imaging mechanism of snow/cloud images, we design a generative framework consisting of a pair of Generative Adversarial Networks (GAN) - a cloud GAN model and a snow GAN model, that can produce cloud and snow remote sensing images and their corresponding pixel-wise labels. Then, we design a cloud/snow detection network, using synthetic data and pseudo labels to achieve weakly supervised cloud and snow detection.

In our method, the input of the cloud GAN model consists of randomly selected cloud and cloud-free remote sensing images. The model generates both cloud reflectance and cloud opacity. By overlaying the cloud reflectance with cloud-free remote sensing images according to cloud opacity, we can generate not only cloud images but also their corresponding pixel-level pseudo labels for cloud detection. For the snow GAN model, the input consists of randomly selected snow and background remote sensing images. It produces snow reflectance, synthesizes new images with snow, and pixel-level labels for snow detection. In addition, the generated snow images can also be further input into cloud GAN to generate cloud-snow mixed images as well as pixel-level labels.

We construct cloud detection networks and snow detection networks based on ResNet50 [72]. Using the generated data with their pseudo cloud and snow labels, weakly supervised cloud detection and snow detection can be achieved. We experiment on Levir_CS [42], a public remote sensing cloud/snow detection dataset. The experiments suggest that only trained with coarse-grained image-level labels, our method can produce pixel-level cloud and snow detection results. Our method achieves superior cloud and snow detection performance by comparing it with several well-known weakly supervised semantic segmentation methods and even matches the accuracy of fully supervised cloud and snow detection methods trained on pixel-level labels.

II. Methods

Fig. 2 illustrates the proposed method's overall structure. It involves a cloud detection task and a snow detection task. In

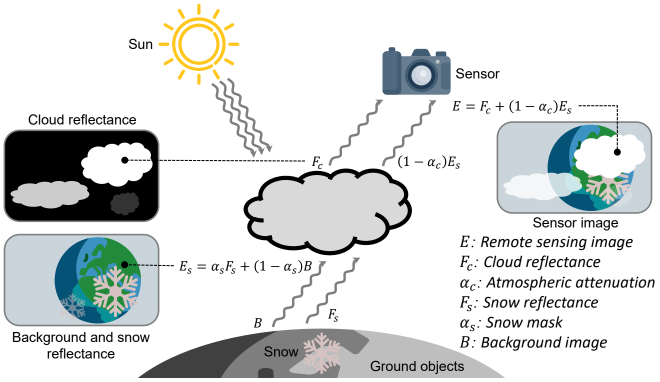

Fig. 1. An illustration of the cloud and snow imaging model. The energy E captured by each unit of the imaging sensor can be approximately decomposed into a linear combination of cloud reflectance energy F_c and background energy E_s. Background E_s can be approximately expressed as the superposition of snow reflectance energy F_s and ground objects reflectance energy B.

each task, we introduce a GAN model to generate cloud/snow images as well as pixel-level labels based on the imaging mechanisms. Then, the cloud/snow detection networks are trained based on the generated data. The subsequent sections will introduce the remote sensing imaging models of cloud and snow respectively, the image generation networks for cloud and snow, the loss functions of different GAN models, the cloud and snow detection methods, and the implementation details of the proposed framework.

A. Cloud and Snow Imaging Model

Inspired by the cloud imaging model [53, 73], we extend the image model with a hybrid form to handle the interference from both clouds and snow in remote sensing images. In this subsection, we introduce the hybrid image model used in our method.

As shown in Fig. 1, the sensor onboard receives energy from both ground objects and clouds at the same time. According to [53, 73], remote sensing image E can be regarded as the linear combination of cloud reflectance image F_c and a ground object image G, which can be expressed as follows:

$$E = F_c + (1-\alpha_c)G, \quad \alpha_c \in [0,1], \tag{1}$$

where the energy of G comes from the reflected energy and the radiation of ground objects. α_c is the atmospheric attenuation factor or the opacity of clouds: the larger the α_c, the thicker the cloud, $\alpha_c = 0$ indicates there is no cloud, while $\alpha_c = 1$ indicates the ground objects are completely obscured.

Considering the ground object may be also covered with snow, although snow and cloud are similar in vision, their imaging principle is quite different. Snow is usually attached to ground objects and usually appears below clouds, which presents a non-transparent state different from clouds.

Suppose F_s denotes the snow reflectance, B denotes the ground objects, and when there are no clouds, the snow image E_s can be expressed as follows:

$$E_s = \alpha_s F_s + (1-\alpha_s)B, \quad \alpha_s \in \{0,1\}, \tag{2}$$

where α_s denotes a binary mask indicating the location of snow pixels. $\alpha_s = 1$ means that the ground objects are completely covered by snow, and $\alpha_s = 0$ means that there is no snow and the ground objects are clearly visible.

Since clouds and snow may both exist in the image at the same time and snow is usually attached to the surface of ground objects and under the cloud, by combining Eq. 2 and Eq. 1, and let $E_s = G$, a complete form of cloud and snow imaging model can be expressed as follows:

$$\begin{aligned} E &= F_c + (1-\alpha_c)G \\ &= F_c + (1-\alpha_c)(\alpha_s F_s + (1-\alpha_s)B), \end{aligned} \tag{3}$$

where $0 \le \alpha_c \le 1$, and $\alpha_s \in \{0,1\}$.

B. Cloud and Snow Image Generation

We propose a weakly supervised approach for cloud detection. Given a set of images with and without clouds, our cloud generation network G_c can synthesize new images with clouds and pixel-level cloud reflectivity labels. The generated data and labels are further provided to the cloud detection network as training samples. Finally, pixel-level cloud detection is achieved with only image-level annotations.

The input of our cloud generator is a pair of images - a cloud-free image x_b and a cloud image x_c. The output of the cloud generator is also a pair of images - the generated cloud reflectance $\tilde{r}_c$ and the generated opacity $\tilde{\alpha}_c$, with the same sizes as the input pair. In the above process, we do not directly employ the networks' outputs, but instead learn to extract the cloud foreground image from the input cloud image, thereby ensuring the generated cloud image's texture and color authenticity. Suppose the direct output of the cloud generator is represented as r_c, which is utilized to extract cloud reflectance $\tilde{r}_c$ from the input cloud image:

$$\tilde{r}_c = r_c \cdot \max(\gamma(x_c), 0), \tag{4}$$

where $\gamma(x_c)$ performs a nonlinear stretch on x_c:

$$\gamma(x_c) = \beta_1(e^{\beta_2 x_c} - 1)x_c, \tag{5}$$

where β_1 and β_2 are the stretch factors. For the generated cloud opacity, since it is highly related to the cloud reflectance (higher reflectivity clouds may have greater thickness), we consider cloud opacity $\tilde{\alpha}_c$ as a linear transformation of the cloud reflectance:

$$\tilde{\alpha}_c = \delta \tilde{r}_c, \tag{6}$$

where δ represents a scaling factor.

Given the generated reflectance $\tilde{r}_c$ and opacity $\tilde{\alpha}_c$, a new image y_c with cloud coverage can be synthesized based on the imaging model Eq. 1 as follows:

$$\begin{aligned} y_c &= G_c(x_c, x_b) \\ &= \tilde{r}_c + (1-\tilde{\alpha}_c)x_b. \end{aligned} \tag{7}$$

We apply a linear combination of the generator's output instead of generating the cloud image directly from noise. We apply a nonlinear stretch mapping to the input cloud image, enhancing the details of the clouds by adjusting the contrast and brightness of the cloud regions, making the subtle and complex

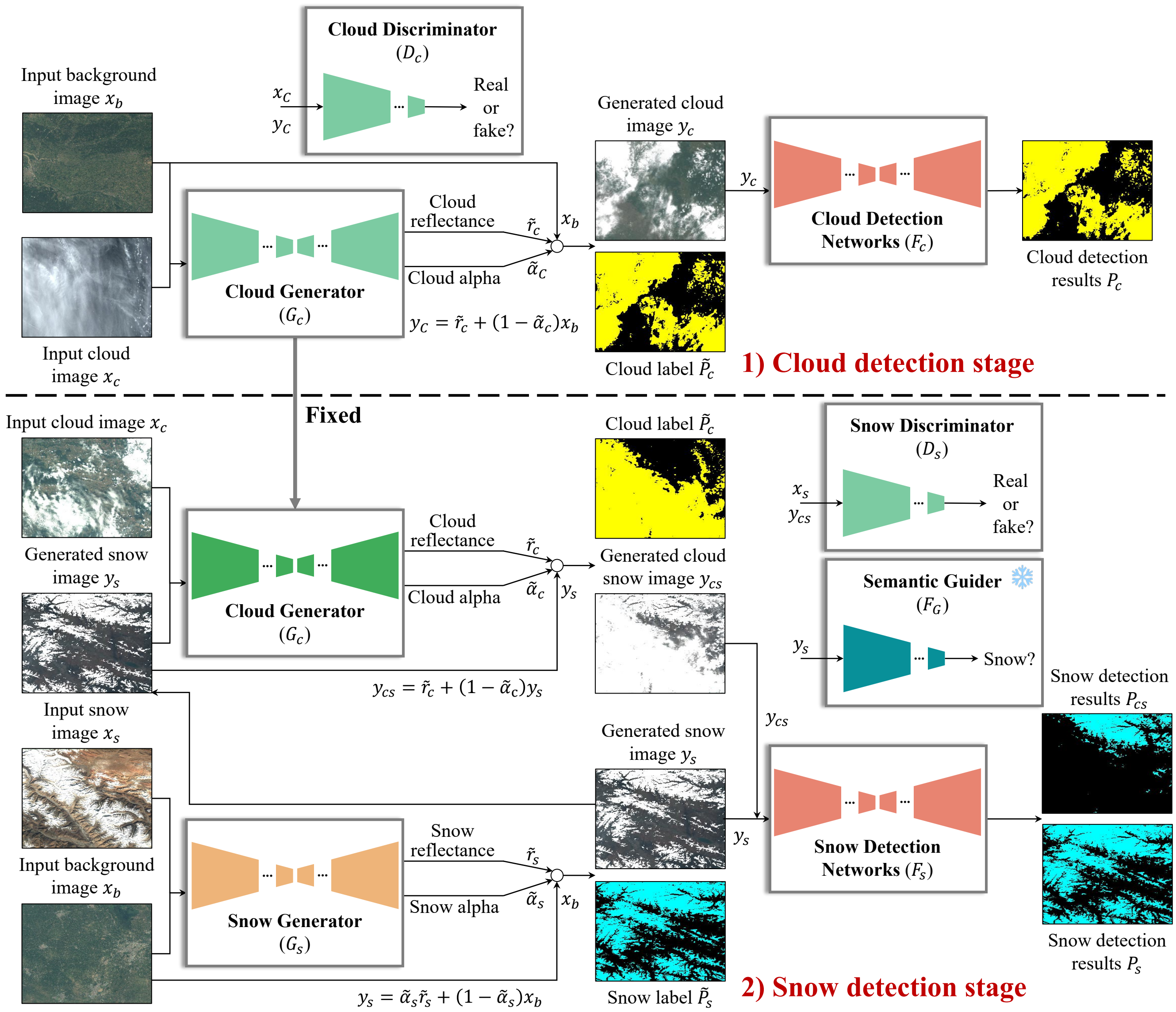

Fig. 2. Overview of the proposed method. It consists of two stages, each of which is responsible for detecting clouds and snow, respectively.

texture details of the cloud regions clearer. Furthermore, under the constraints of the cloud and snow hybrid imaging physical model, we further extract the foreground image of the cloud from the input cloud image. By retaining the original features of the input cloud image, we can avoid color distortion and artifacts that may arise from directly generating cloud images, ensuring the authenticity of the newly generated cloud image in terms of cloud texture and color. This ensures that the synthesized cloud image possesses both visual and physical realism. In addition, the generated cloud reflectivity can be used as a label to guide the training of subsequent cloud detection models.

Like the above cloud generation process, we introduce a snow generator G_s to synthesize snow images and pixel-wise labels. A pair of images consisting of a background remote sensing image x_b and a snow image x_s are fed to the snow generator. The snow reflectance $\tilde{r}_s$ is generated as follows:

$$\tilde{r}_s = x_s r_s. \tag{8}$$

Based on $\tilde{r}_s$, the snow's opacity $\tilde{\alpha}_s$ can be further obtained by setting a threshold η_s:

$$\tilde{\alpha}_s = \begin{cases} 1 & \text{if } \tilde{r}_s \geq \eta_s \\ 0 & \text{else.} \end{cases}. \tag{9}$$

Then, based on the snow imaging model, a new snow remote sensing image y_s can be generated using the input background image x_b and the generated reflectance $\tilde{r}_s$ and opacity $\tilde{\alpha}_s$:

$$\begin{aligned} y_s &= G_s(x_s, x_b) \\ &= \tilde{\alpha}_s\tilde{r}_s + (1 - \tilde{\alpha}_s)x_b. \end{aligned} \tag{10}$$

In addition to adding cloud layers on top of the background image, we also add clouds to snow images. Note that cloud and snow pixels may exist at the same time in the same area, their overlay order cannot be changed arbitrarily, because cloud layers usually overlie snow layers, but not vice versa. Therefore, to generate cloud-snow mixed images, we feed the synthesized snow images y_s and randomly selected cloud

Algorithm 1 Weakly Supervised Cloud and Snow Detection (WCSD)

```
(1) Cloud image generation
Input: Background cloud-free image $x_b$ and cloud image $x_c$
Output: Generated cloud image $y_c$ and label $\widetilde{P}_c$
    $\widetilde{r}_c, \widetilde{\alpha}_c \overset{r_c}{\leftarrow} r_c, G_c(x_b, x_c)$    ▷ $G_c$ is the cloud generator
    $y_c = \widetilde{r}_c + (1 - \widetilde{\alpha}_c)x_b$, $\widetilde{P}_c \leftarrow \widetilde{r}_c$
    $D_c(y_c; x_c) \rightarrow$ Real/Fake    ▷ $D_c$ is the cloud discriminator
    return $y_c, \widetilde{P}_c$
(2) Cloud/snow image generation
Input: Background image $x_b$ and snow image $x_s$, randomly selected cloud image $x_c$, fixed cloud generator $G_c$
Output: Generated cloud/snow image $y_{cs}$, cloud label $\widetilde{P}_c$, snow label $\widetilde{P}_s$
    $\widetilde{r}_s, \widetilde{\alpha}_s \overset{r_s}{\leftarrow} G_s(x_b, x_s)$    ▷ $G_s$ is the snow generator
    $y_s = \widetilde{\alpha}_s \widetilde{r}_s + (1 - \widetilde{\alpha}_s)x_b$
    $\mathcal{L}_A = F_G(y_s)$    ▷ $F_G$ is the pre-trained snow classifier
    $\widetilde{r}_c, \widetilde{\alpha}_c \overset{r_c}{\leftarrow} G_c(y_s, x_c)$
    $y_{cs} = \widetilde{r}_c + (1 - \widetilde{\alpha}_c)y_s$, $\widetilde{P}_c \leftarrow \widetilde{r}_c$, $\widetilde{P}_s = (1 - \widetilde{P}_c)\widetilde{\alpha}_s$
    $D_s(y_{cs}; x_s) \rightarrow$ Real/Fake ▷ $D_s$ is the snow discriminator
    return $y_{cs}, \widetilde{P}_c, \widetilde{P}_s$
(3) Cloud and snow detection
Input: Generated cloud/snow image $y_{cs}$, cloud label $\widetilde{P}_c$, snow label $\widetilde{P}_s$
Output: Cloud detection result $P_c$, snow detection result $P_s$
    $P_c = F_c(y_{cs})$    ▷ $F_c$ is the cloud detection network
    $P_s = F_s(y_{cs})$    ▷ $F_s$ is the snow detection network
    return $P_c, P_s$
```

images x_c into the cloud generator. By following the imaging model Eq. 3 and combining Eq. 7 and Eq. 10, the synthesized images that contain both cloud and snow pixels can be finally expressed as:

$$\begin{aligned} y_{cs} &= \tilde{r}_c + (1 - \tilde{\alpha}_c)y_s \\ &= \tilde{r}_c + (1 - \tilde{\alpha}_c)(\tilde{\alpha}_s \tilde{r}_s + (1 - \tilde{\alpha}_s)x_b). \end{aligned} \tag{11}$$

C. GAN Loss Functions

To train the cloud generator G_c and the snow generator G_s, we introduce two discriminators D_c and D_s respectively. The generators and the discriminators are jointly trained under a Least-Square GAN framework [69].

For the cloud GAN model, during the training process, the cloud generator continuously generates cloud reflectance $\tilde{r}_c$ and opacity $\tilde{\alpha}_c$, which is synthesized to obtain a new cloud image y_c. y_c and input cloud image x_c are fed to the cloud discriminator. On one hand, the discriminator is trained to judge which input images are fake and which are real. The loss function of the discriminator is defined as follows:

$$\mathcal{L}_{D_c} = \frac{1}{2}(D_c(x_c) - 1)^2 + \frac{1}{2}D_c(G_c(x_c, x_b))^2. \tag{12}$$

On the other hand, the generator is trained to make the fake images more similar to the real ones, and thereby "fool" the discriminator. The loss function of the generator is defined as follows:

$$\mathcal{L}_{G_c} = \frac{1}{2}D_c(G_c(x_c, x_b) - 1)^2. \tag{13}$$

To make the generated cloud images more realistic, we apply constraints to the saturation of the generated cloud reflectance: $\beta_s||s||_2^2$, where $s = (\max(R, G, B) - \min(R, G, B))/\max(R, G, B)$ and β_s represents a penalty factor. Therefore, the total loss of the cloud generator can be expressed as:

$$\mathcal{L}_{G_c} = \frac{1}{2}D_c(G_c(x_c, x_b) - 1)^2 + \beta_s||s||_2^2. \tag{14}$$

The cloud generator and cloud discriminator can be trained jointly under a unified training objective. By combing the loss functions Eq. 12 and Eq. 14, the training can be performed under a min-max optimization process, where the objective function is defined as follows:

$$G_c^\star, D_c^\star = \arg \min_{G_c, D_c} \max_{D_c}(\mathcal{L}_{G_c} + \mathcal{L}_{D_c}). \tag{15}$$

For the snow GAN model, the loss functions of the generator G_s and the discriminator D_s are defined similarly to the cloud GAN model:

$$\mathcal{L}_{D_s} = \frac{1}{2}(D_s(x_s) - 1)^2 + \frac{1}{2}D_s(G_s(x_s, x_b))^2, \tag{16}$$

$$\mathcal{L}_{G_s} = \frac{1}{2}D_s(G_s(x_s, x_b) - 1)^2 + \mathcal{L}_A(y_s), \tag{17}$$

where the only difference is that we apply an additional regularization term $\mathcal{L}_A$ to the loss function of the generator. $\mathcal{L}_A$ is the loss value from a pre-trained snow classification network given the input of synthesized snow image y_s. The reason behind this is to prevent the generator from producing "cloud-like" snow images. This ensures that the generated snow images can be classified as snow by the pre-trained classifier, thereby enabling the snow generator to produce semantically more "snow-like" images.

The two networks G_s and D_s can be also jointly trained under a min-max optimization process:

$$G_s^\star, D_s^\star = \arg \min_{G_s, D_s} \max_{D_s}(\mathcal{L}_{G_s} + \mathcal{L}_{D_s}). \tag{18}$$

D. Cloud and Snow Detection

At the cloud and snow detection stage, we train the detectors based on the images and pixel-wise labels generated by the cloud GAN model and the snow GAN model. Both the cloud detector and the snow detector are fully convolutional networks.

The generated cloud/snow image y_{cs} can be used as input for the cloud detection networks f_c^{det}. The labels for training are obtained from the generated cloud reflection image $\tilde{r}_c$:

$$\tilde{p}_c = \begin{cases} 1 & \text{if } \tilde{r}_c \geq \eta_c \\ 0 & \text{else.} \end{cases}, \tag{19}$$

η_c is a predetermined threshold, and the pixel locations in $\tilde{r}_c$ that exceeds η_c indicates the presence of clouds. Suppose f_c^{det} represents the cloud detection network, and the detection output probability map is p_c, we have:

$$p_c = \text{sigmoid}(f_c^{det}(y_{cs})), \tag{20}$$

TABLE I
A DETAILED CONFIGURATION OF OUR NETWORKS.

	Layer	Input	#In	#Out	Stride	$\sigma(\cdot)$
Encoder	conv1	image	img_channel	64	1	ReLu
	ResBlock1	conv1	64	256	2	ReLu
	ResBlock2	ResBlock1	256	512	2	ReLu
	ResBlock3	ResBlock2	512	1024	2	ReLu
	ResBlock4	ResBlock3	1024	2048	2	ReLu
	conv2	ResBlock4	2048	2048	2	ReLu
	conv3	conv2	2048	2048	2	ReLu
Decoder	interp1	conv3	2048	2048	2	None
	conv4	interp1	2048	2048	1	ReLu
	interp2	conv4 + ResBlock4	1024	1024	2	None
	conv5	interp2	1024	1024	1	ReLu
	interp3	conv5 + ResBlock3	1024	1024	2	None
	conv6	interp3	1024	512	1	ReLu
	interp4	conv6 + ResBlock2	512	512	2	None
	conv7	interp4	512	512	1	ReLu
	interp5	conv7 + ResBlock1	256	256	2	None
	conv8	interp5	256	64	1	ReLu
	conv9	conv8+conv1	64	out_channel	1	None

where $f_c^{det}(y_{cs})$ is the output logits of the detection network f_c^{det}, sigmoid$(\cdot)$ is the sigmoid function that converts the output logits to probability values.

The loss function for training the cloud detection network is defined based on the standard cross-entropy loss:

$$\mathcal{L}_{f_c^{det}} = \tilde{p}_c \log p_c + (1-\tilde{p}_c)\log(1-p_c). \tag{21}$$

As for the snow detection branch, its input-output formula and loss function are similar to those of the cloud detection branch. Note that, snow labels may be obscured by clouds. Therefore, cloud cover must be considered when generating snow labels, i.e., only the snow pixels in areas not covered by clouds are used as positive samples for training. The snow label can be obtained using the following formula:

$$\tilde{p}_s = (1-\tilde{p}_c)\tilde{\alpha}_s. \tag{22}$$

After obtaining the label, the snow detection loss function can be also defined similarly to the cloud detection loss:

$$\mathcal{L}_{f_s^{det}} = \tilde{p}_s \log p_s + (1-\tilde{p}_s)\log(1-p_s), \tag{23}$$

where $p_s =$ sigmoid$(f_s^{det}(y_{cs}))$ represents output probability map of snow detection networks.

The pseudocode of the proposed weakly supervised cloud and snow detection framework WCSD is shown in Algorithm 1.

E. Implementation Details

1) Hyperparameter Settings: During the training of the cloud GAN and snow GAN networks, the learning rates of the generator and discriminator are set to $1e^{-5}$ and $1e{-}6$. The cloud detection and snow detection networks have a learning rate of $1e^{-4}$. β_1 is set to 3 and β_2 is set to 0.5. δ is set to 1.0.

2) Network Structure: We use the same network configuration for the cloud/snow detectors and the cloud/snow generators. The details of the network architecture are given in Table I. For the cloud discriminator and snow discriminator, except for the class number of outputs, the network structure is the same as ResNet50 [72]. For the generator and detection networks, more refined features need to be extracted. We, therefore, designed an encoder-decoder architecture based on ResNet50. We remove the initial down-sampling operations of ResNet50 and retain its residual modules. In the feature output part of ResNet50, other convolution operations are involved. In the decoder stage, we use interpolation and convolution to gradually increase the fineness of the output feature maps until it has the same sized output as the input image. For each up-sampling operation, we use a feature fusion strategy to further enrich the features.

The columns "#Out", "#In", "Stride" and "$\boldsymbol{\sigma(\cdot)}$" represent the number of output channels, input channels, stride, and activation functions of the operation respectively. "Layer" represents the name of the operation, where "conv" represents convolution operation and "interp" represents bilinear interpolation. "ResBlock" refers to the residual module in ResNet50.

3) Data Augmentation: We use data augmentation on both cloud/snow images and background images to increase the diversity of the generated data. These data augmentation includes random left-right flips, random up-down flips, and random rotations with $\{0, 90, 180, 270\}$ degrees. In addition, the images are randomly cropped to generate data at various resolutions.

4) Inference Details: During the detection, to improve the accuracy of cloud and snow detection, we first use an image-level classification network that has been pre-trained to determine whether the image to be detected contains clouds or snow. If only clouds are included, only cloud detection is performed. The snow detection part follows a similar logic. If both clouds and snow are presented, cloud detection should precede snow detection, and the cloud detection results should be given the highest priority.

III. EXPERIMENTAL RESULTS

A. Experimental Setup

In the experiments, we compare the proposed method with both weakly supervised segmentation methods and fully su-

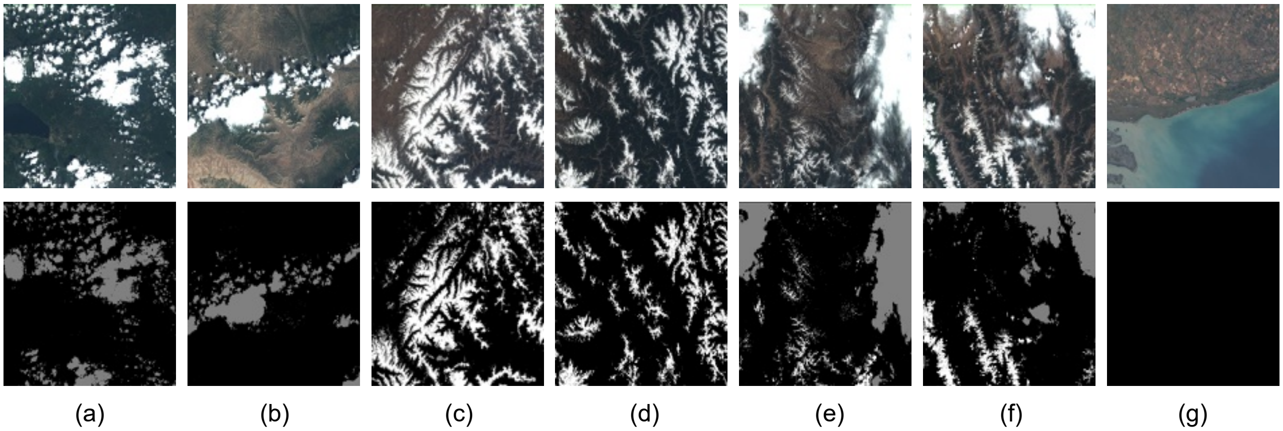

Fig. 3. Examples of remote sensing images from the Levir_CS dataset and their corresponding supervision labels. In the label images, 0 represents background, 128 represents clouds, and 255 represents snow.

pervised segmentation methods. We also compare with some recent methods proposed specifically for the remote sensing image cloud/snow detection tasks. All of the methods are trained and evaluated using the Levir_CS dataset [42]. The Levir_CS dataset contains 4168 GF-1 satellite WFV scenes. Each scene covers $211km \times 192km$, the resolution is $8m$. The multispectral images from the Levir_CS dataset contain four bands. Specifically, the spectral ranges of these bands are $450-520nm$ (Band 1: Blue Band), $520-590nm$ (Band 2: Green Band), $630-690nm$ (Band 3: Red Band), and $770-890nm$ (Band 4: Near-Infrared Band), respectively. In subsequent experiments, we only use the three visible light RGB bands from the Levir_CS dataset. The images in the Levir_CS dataset are collected from a global scale and include various ground types such as plains, plateaus, oceans, deserts, ice, etc. The pixel-level labels for all images are manually annotated into three categories: background, cloud and snow. In subsequent experiments, the training, validation, and test sets are divided in the ratio of 3:1:1. For the weakly supervised learning methods, neither the training set nor the validation set contains pixel-level labels and only image-level labels of whether clouds or snow are contained are provided during the training process. Fig. 3 shows some sample remote sensing images from the Levir_CS dataset and their corresponding cloud and snow pixel-level category labels. Manually annotating such pixel-level supervision labels is expensive and prone to errors, especially when visually similar clouds and snow appear simultaneously in an image. Sample images (a) and (b) in Fig. 3 contain only clouds, (c) and (d) contain only snow, (e) and (f) contain both clouds and snow, and (g) contains neither clouds nor snow.

B. Evaluation Metrics

In our experiment, Precision (P), Recall (R), F1-score (F1), and Overall Accuracy (OA) are used as our evaluation metrics to assess the performance of cloud and snow detection. The metrics are defined as follows:

$$P = \frac{TP}{TP+FP}, \tag{24}$$

$$R = \frac{TN}{TP+FN}, \tag{25}$$

$$F1 = \frac{2*P*R}{P+R}, \tag{26}$$

$$OA = \frac{TP+TN}{TP+TN+FP+FN}, \tag{27}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative pixels, respectively.

C. Compared with Weakly Supervised Learning Methods

We compare our method with the latest weakly supervised learning methods, which all use the same image-level labels for training. Because there are only weakly supervised cloud detection methods currently, we use some latest weakly supervised semantic segmentation methods for cloud and snow detection comparison.

Table II shows the comparison results. Among them, the GradCAM [64] and WDCD method [54] are designed based on the CAM method [63]. GradCAM [64] is an improvement of the CAM method. WDCD method is a cloud detection method based on the CAM method. We extend it to the task of snow detection, with the same network structure as cloud detection. Both GAN-CDM [67] and CloudMatting [53] are weakly supervised cloud detection methods, and both need to generate cloud remote sensing images with the help of GANs. Since they are not easy to extend to snow detection tasks, we only compare their performance in cloud detection tasks.

The proposed method has achieved the best cloud and snow detection effect, as demonstrated by the experimental results. Weakly supervised learning methods based on CAM, such as GradCAM and WDCD methods, can accomplish cloud detection and snow detection simultaneously, but both cloud detection and snow detection have relatively low performance. In terms of index recall, the WDCD method slightly surpasses the proposed WCSD method, indicating that the detection threshold of the WDCD method is relatively lower, allowing it to detect more clouds and snow. However, as can be seen

TABLE II
COMPARISON WITH WEAKLY SUPERVISED LEARNING METHODS ON CLOUD AND SNOW DETECTION TASKS.

Methods	Cloud detection				Snow detection			
	precision	recall	F1	OA	precision	recall	F1	OA
GradCAM [64]	0.2825	0.6976	0.4021	0.6138	0.1598	0.5354	0.2465	0.9349
WDCD [54]	0.2756	**0.8849**	0.4203	0.5455	0.1934	**0.7022**	0.3033	0.9362
GAN-CDM [67]	0.4630	0.8644	0.6030	0.7161	–	–	–	–
CloudMatting [53]	0.7019	0.8345	0.7625	0.9032	–	–	–	–
WCSD (ours)	**0.7353**	0.8292	**0.7838**	**0.9138**	**0.5107**	0.5718	**0.5395**	**0.9807**

from other metrics in Table II, the performance of the WDCD method in metrics other than recall significantly decreases compared to other methods, especially the WCSD method. This suggests that the detection results of the WDCD method contain a large number of false alarms, demonstrating that this method does not balance the detection rate and false alarm rate well. The performance of the GAN-based methods surpasses that of the CAM-based methods, and the performance gap between the GAN-based methods and the proposed WCSD method is relatively small. However, none of the GAN-based methods can detect snow.

Fig. 4 illustrates the results of various cloud and snow detection methods. Different rows in the figure represent various input data. The first column represents the image input, the second column represents the labels, and the remaining columns represent the results from different methods. The black-color pixels represent the ground objects, while the cyan and yellow pixels represent snow and cloud, respectively. According to the visualization results, it is difficult for the CAM-based methods to accurately depict the cloud-snow boundary. Both the GradCAM method and the WDCD method can only approximate the position of clouds or snow, but their accuracy cannot be compared with other methods.

The GAN-based methods, including GAN-CDM and CloudMatting, have significantly improved the accuracy of cloud detection, especially CloudMatting, which can generate pseudo labels pixel-by-pixel and thus achieve extremely precise cloud detection results. However, neither the GAN-CDM nor CloudMatting method can distinguish between cloud and snow. They both incorrectly identify snow as clouds. In the cloud detection task, the proposed WCSD method can produce more accurate cloud detection results than the CloudMatting method. Similarly, in the snow detection task, the WDCD method can achieve more accurate snow detection results, and in some cloud-snow mixed images, it can also distinguish clouds and snow more precisely.

D. Compared with Supervised Learning Methods

We compare our proposed method to several supervised learning methods, such as the well-known semantic segmentation methods FCN-32s and FCN-16s [74], as well as some fully supervised cloud and snow detection methods. All methods of comparison are trained using pixel-level labels. Table III displays their results.

The experimental results demonstrate that our proposed method can achieve comparable results as the fully supervised method (FCN-32s) in cloud and snow detection tasks, but there still remains a gap with CDSNet [49]. Fig. 5 illustrates the results of various methods. Different rows in the figure correspond to distinct input images. The first column represents the image input, the second column represents the labels, and the subsequent columns represent the results of various methods. The color maps are defined in the same way as Fig. 4. Furthermore, it can be seen that some fully supervised methods, especially the FCN method, output smooth clouds and snow masks, while some fine clouds and snow are not detected. This is because the pooling operation used in the feature extraction process of such segmentation methods leads to the loss of image details due to downsampling, and the final prediction mask obtained by upsampling the feature map is not fine enough.

E. Generated Data Analysis

Fig. 6 shows some generated cloud and snow remote sensing images by our method. It can be seen that our method can generate cloud images with different types and thicknesses. The snow images can be generated in various states, and the texture of snow can be restored effectively. In addition, our method can generate images containing both clouds and snow, where the overlay order between clouds and snow is accurately restored. However, in our generated results, there are also cases where the texture of snow and the texture of ground objects are inconsistent. This is because our method has not yet considered the correlation between ground objects and snow when generating snow.

Our method can also generate pseudo labels for the images, where the cloud and snow coverage area can be accurately labeled. Because the labels are generated automatically, the inaccurate labeling of boundaries and thin clouds could be a factor limiting the performance of our method for cloud and snow detection.

F. Ablation Experiments

We conducted ablation experiments where various strategies were eliminated to evaluate the efficacy of our proposed method. The results are shown in Table IV. The evaluated strategies include reflectance stretching, opacity adjustment, and snow semantic guidance, which are described as follows:

reflectance stretching (S). When generating cloud or snow images, our method has two steps: generating reflectance images and synthesizing new images. Considering that for cloud-free areas, the generated reflectivity value should be strictly equal to 0, therefore, we perform a stretching operation on the output reflectance image as described in section II-B.

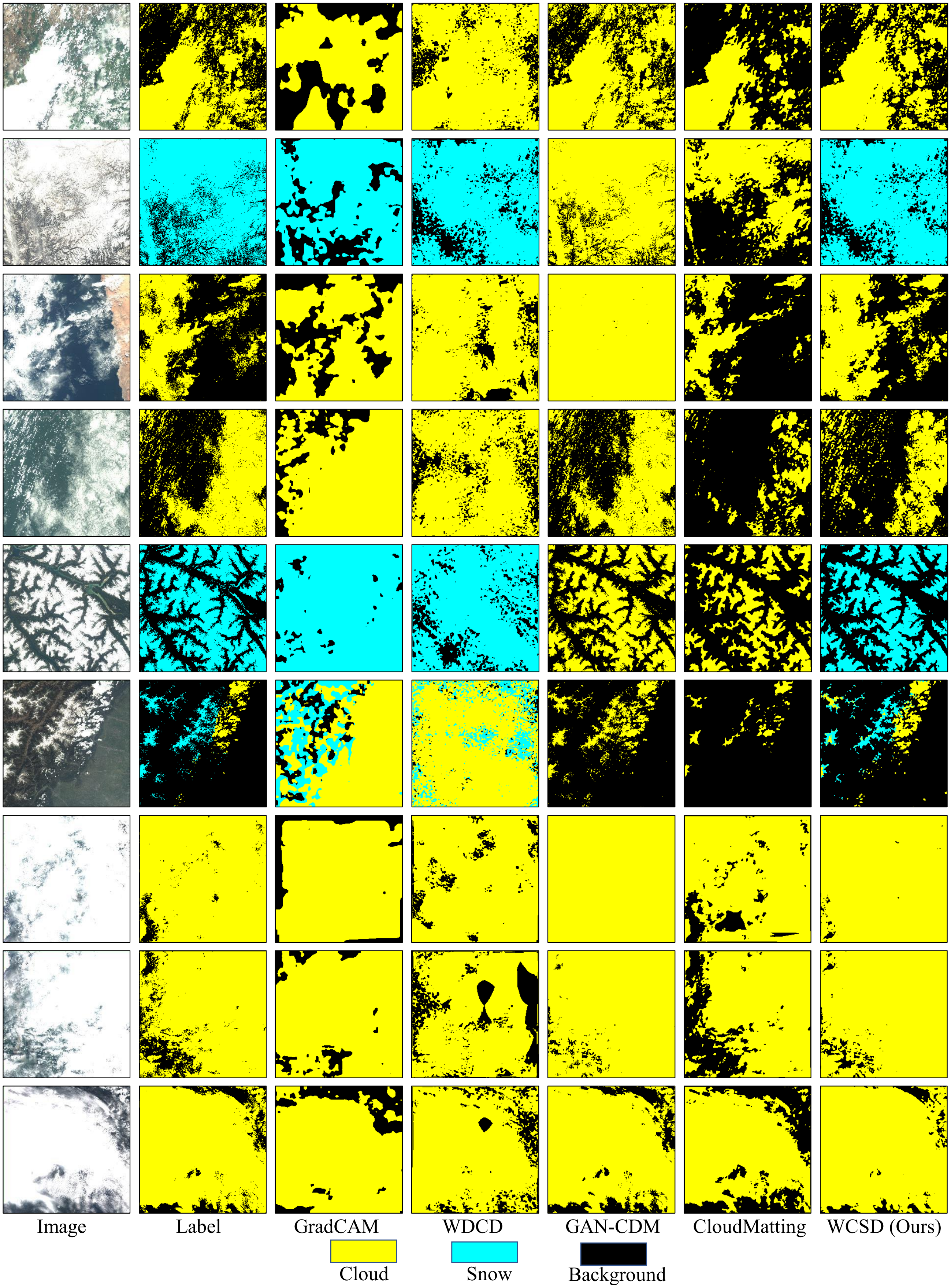

Fig. 4. Detection results of different weakly supervised learning methods on cloud and snow detection task.

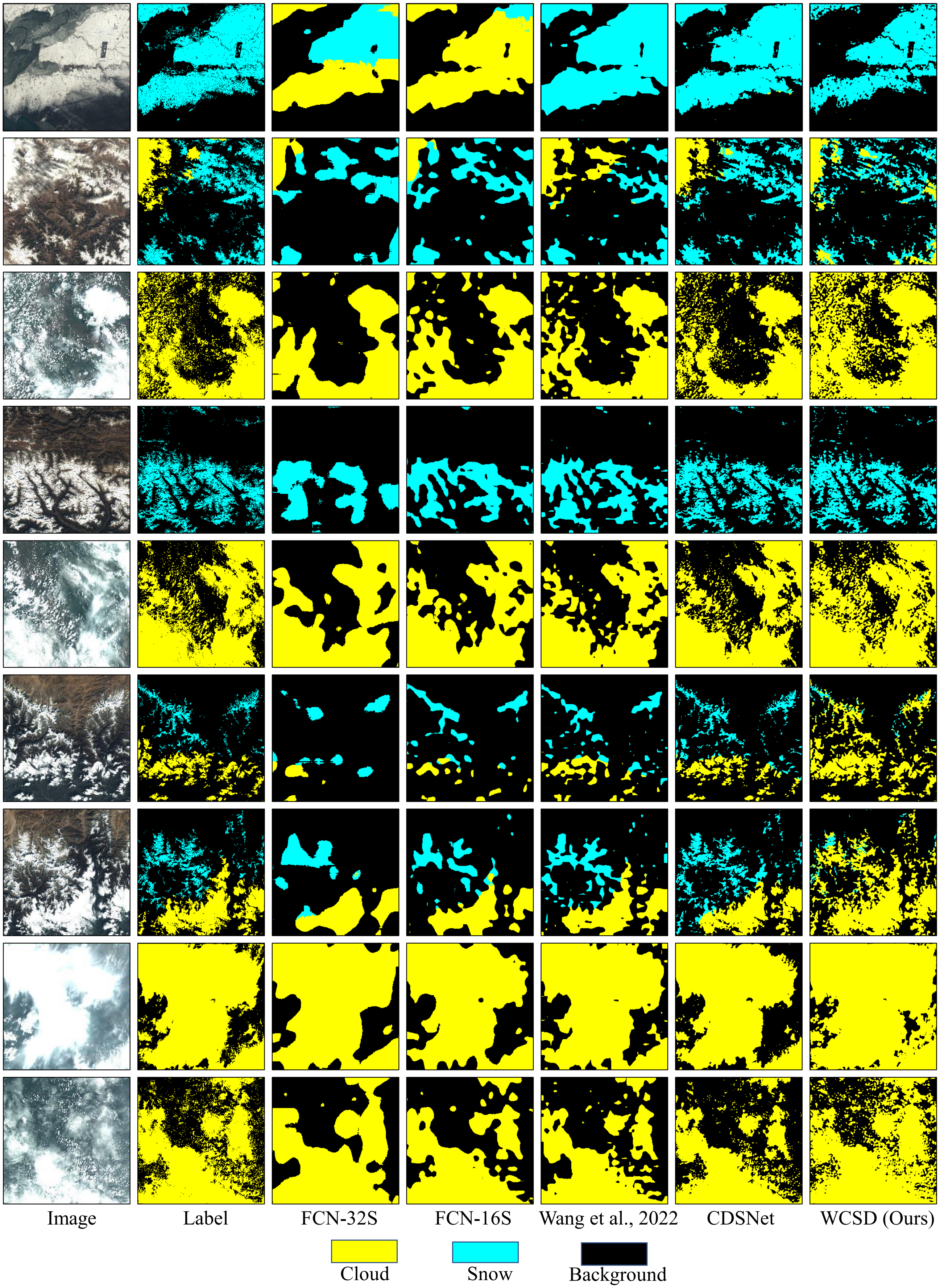

Fig. 5. Detection results of different supervised learning methods on cloud and snow detection task.

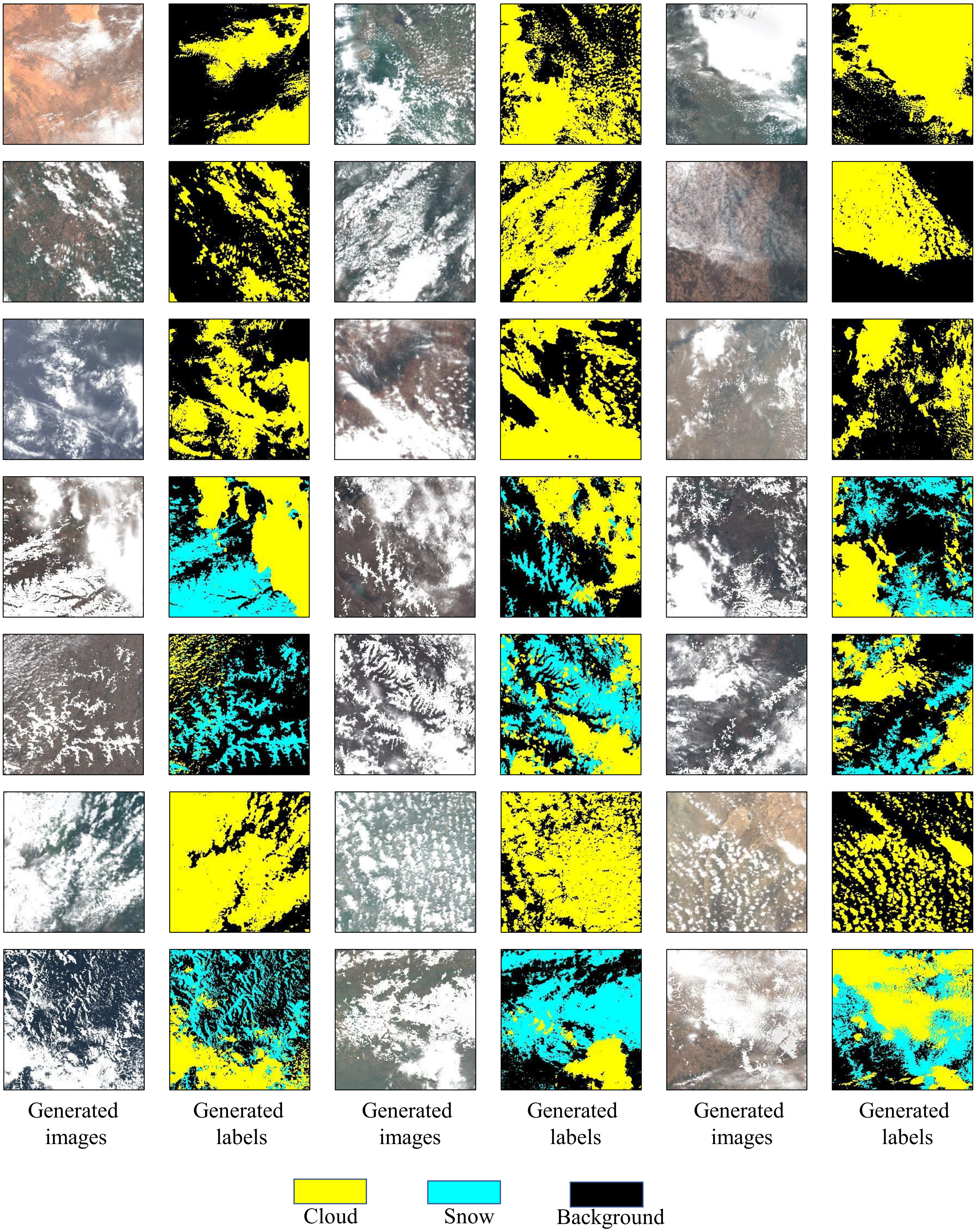

Fig. 6. Some generated cloud and snow remote sensing images and their corresponding pseudo labels.

TABLE III
COMPARISON WITH SUPERVISED LEARNING METHODS ON CLOUD AND SNOW DETECTION TASKS.

Methods	Cloud detection				Snow detection			
	precision	recall	F1	OA	precision	recall	F1	OA
FCN-32s [74]	0.8432	0.7022	0.7663	0.9202	0.5907	0.6230	0.6064	0.9840
FCN-16s [74]	0.8654	0.7220	0.7872	0.9273	0.6132	0.6286	0.6208	0.9848
Wang et al. [50]	0.8803	0.7573	0.8142	0.9352	0.7238	**0.6931**	0.7081	0.9887
CDSNet [49]	**0.9011**	0.8280	**0.8630**	**0.9510**	**0.7320**	0.6886	**0.7097**	**0.9889**
WCSD (ours)	0.7353	**0.8292**	0.7838	0.9138	0.5107	0.5718	0.5395	0.9807

TABLE IV
ABLATION EXPERIMENT RESULTS ON CLOUD AND SNOW DETECTION. ABLATION ITEMS ARE REFLECTANCE STRETCHING (S), OPACITY ADJUSTMENT (A), AND SNOW SEMANTIC GUIDANCE (G).

Ablation			Cloud detection				Snow detection			
S	A	G	precision	recall	F1	OA	precision	recall	F1	OA
×	×	×	**0.8758**	0.5927	0.7070	0.9085	0.3979	0.5622	0.4660	0.9745
✓	×	×	0.7349	0.7876	0.7603	0.9075	0.4391	0.5859	0.5020	0.9770
✓	✓	×	0.7244	0.8120	0.7657	0.9075	0.4088	**0.6587**	0.5045	**0.9807**
✓	✓	✓	0.7353	**0.8292**	**0.7838**	**0.9138**	**0.5107**	0.5718	**0.5395**	**0.9807**

opacity adjustment (A). Considering the correlation between opacity and cloud reflection, we use Eq. 6 to calculate the opacity of the cloud, instead of adding random perturbation to the foreground image like the previous method [53].

snow semantic guidance (G). As discussed in Sec. II-C, to prevent the generator from producing “cloud-like” snow images, we introduce snow classification networks as a semantic guidance in the snow generation process.

According to the results of the ablation experiment, reflectance stretching can effectively improve the accuracy of cloud and snow detection, particularly in cloud detection, which significantly improves recall scores. The generated images with and without reflectance stretching are displayed in Fig. 7. The first three rows represent the results without stretching, while the last three rows represent the results with stretching. Each column, from left to right, depicts the input background image for GANs, the generated cloud image, the generated cloud reflectance, the generated cloud opacity, and the generated cloud labels. Without stretching, the generated cloud remote sensing image may encompass a mask layer resembling thin clouds in the non-cloud region, as depicted in the figure. This is because the generated cloud reflection image has a certain value in the non-cloud region, rather than being close to 0. After stretching is applied, this problem can be effectively avoided, and the generated cloud image is more realistic, thus improving the detection performance.

In addition to reflectance stretching, using additional network branches to predict opacity can increase the diversity of generated cloud images and further improve the performance of cloud and snow detection. The addition of semantic guidance to snow GANs can significantly boost the performance of snow detection. Fig. 8 depicts the snow remote sensing image generated with and without semantic guidance. The figure demonstrates that semantic guidance can result in snow remote sensing images with more pronounced snow texture and a higher level of authenticity. In the absence of semantic guidance, the networks are also capable of generating realistic snow images, but in some cases, they may generate images with subtle textures that resemble cloud images. When these images are used for training the cloud/snow detectors, the detection performance will be affected.

G. Limitation and Discussion

For remote sensing images containing both clouds and snow, distinguishing between cloud and snow areas is a difficult problem. Our method can distinguish between snow and cloud as it can generate more realistic snow and cloud remote sensing images. Nonetheless, the following limits remain in our proposed method.

- The high resemblance of cloud and snow makes the detection challenging, even with pixel-level supervision. There is still room in our method, particularly for snow detection accuracy. Our snow generator currently only considers the texture of the snow itself and ignores the correlation between ground objects and snow textures, resulting in instances where the snow and the ground objects are not well coordinated, thereby affecting the accuracy of the subsequent detection.
- The labels automatically generated by the proposed method may be ambiguous. In many cases, the proposed method can generate visually consistent cloud and snow images, but the labels produced may lack a uniform definition standard in the boundary and thin cloud areas. This results in errors being introduced in the process of training the cloud and snow detector using generated images and labels.

Although clouds and snow have very similar visual effects in remote sensing images, their reflectance characteristics differ in certain specific spectral bands. Currently, our data design is based on the visible RGB three bands, but the WCSD framework is also compatible with other spectral bands. Introducing information from near-infrared and other bands could potentially improve cloud and snow detection accuracy. Addressing these issues by incorporating additional

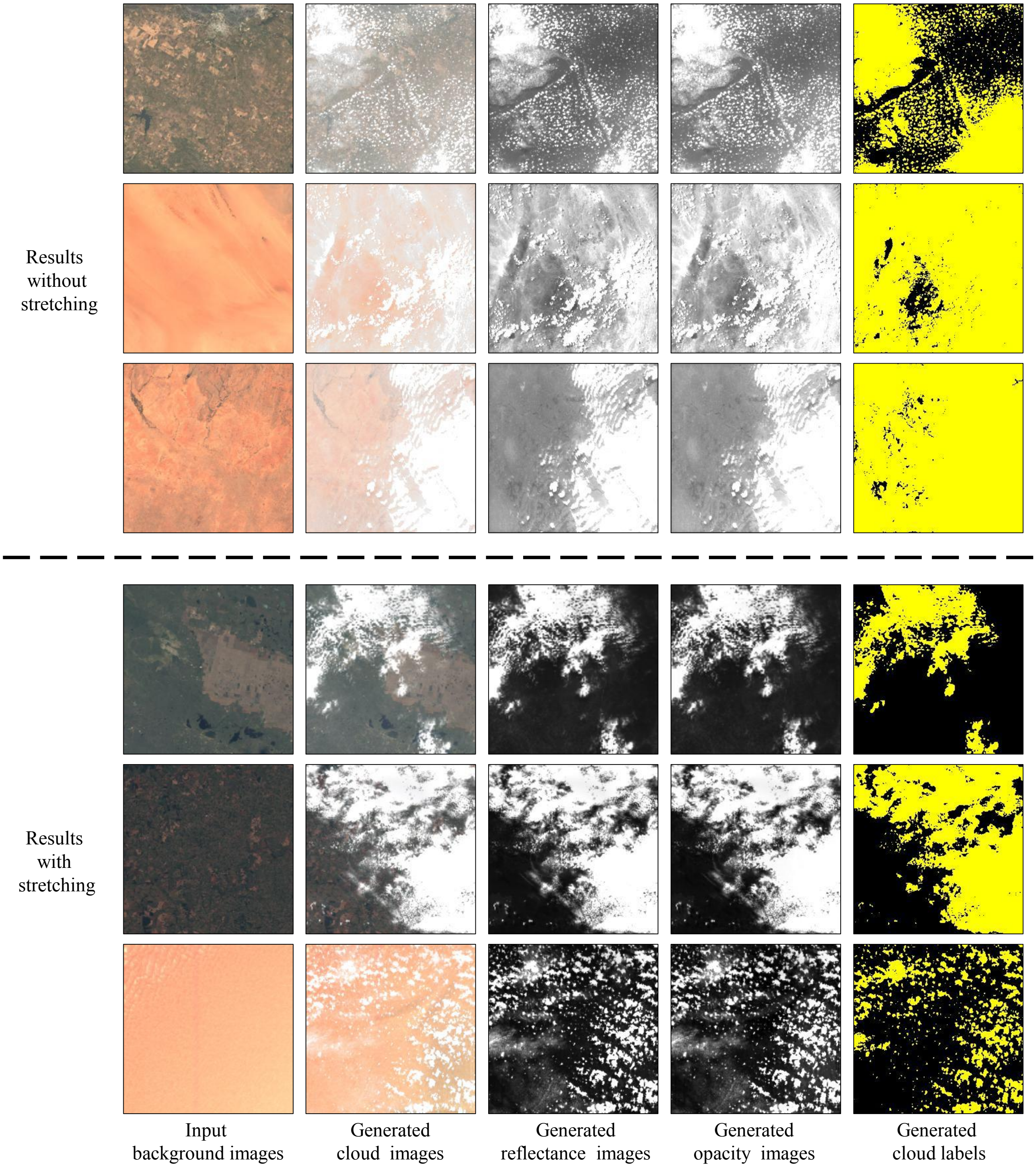

Fig. 7. Generated images with and without reflectance stretching. The figure's first three rows show the results without stretching, while the last three rows show the results with stretching.

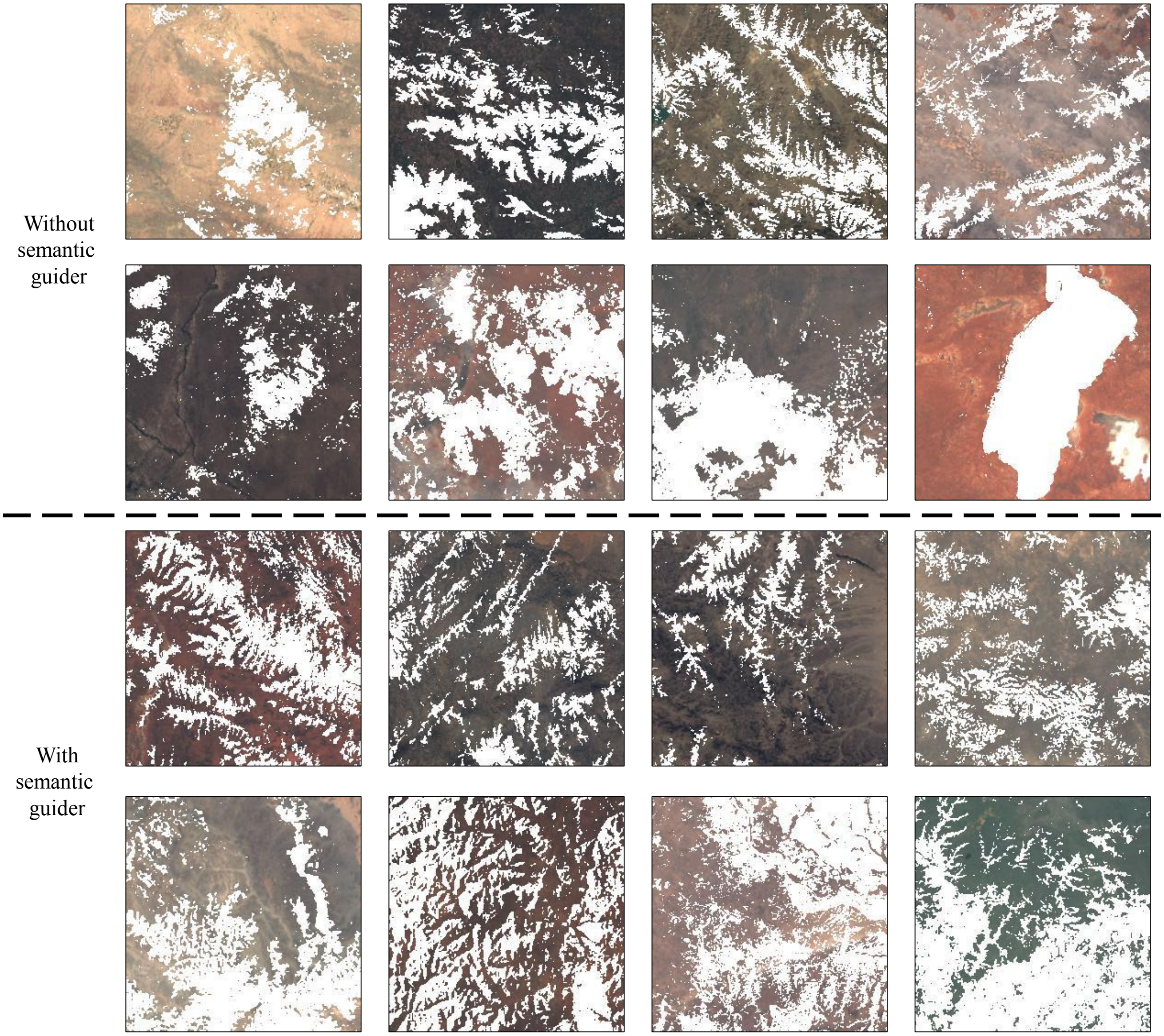

Fig. 8. Snow remote sensing image generated with and without semantic guidance. The figure's first two rows show the results without semantic guidance, while the last two rows show the results with semantic guidance.

spectral information will help enhance detection accuracy and resolve label ambiguities. In addition, to improve the rationality of snow generation, additional meta-information, such as the Digital Elevation Model (DEM) of the image, can be introduced to generate snow images that better match the terrain.

IV. Conclusions

We propose a novel weakly supervised method for remote sensing image cloud and snow detection. Our method employs the cloud and snow imaging mechanism to construct generative adversarial networks that are capable of generating cloud and snow images, as well as their corresponding labels. The generated images and labels are then used for training cloud and snow detection. Compared to other weakly supervised methods, our method achieves superior cloud and snow detection performance. Using only image-level training labels, our method achieves comparable accuracy to fully supervised learning methods trained on pixel-level training labels.

References

[1] Y. Chen, C. He, W. Guo, S. Zheng, and B. Wu, "Mapping urban functional areas using multi-source remote sensing images and open big data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.

[2] X. Liu, H. Zhai, Y. Shen, B. Lou, C. Jiang, T. Li, S. B. Hussain, and G. Shen, "Large-scale crop mapping from multisource remote sensing images in google earth engine," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 414–427, 2020.

[3] B. Yang, S. Hu, Q. Guo, and D. Hong, "Multisource domain transfer learning based on spectral projections for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3730–3739, 2022.

[4] R. Dian, A. Guo, and S. Li, “Zero-shot hyperspectral sharpening,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[5] M. Zhang, W. Li, Y. Zhang, R. Tao, and Q. Du, “Hyperspectral and lidar data classification based on structural optimization transmission,” *IEEE Transactions on Cybernetics*, 2022.

[6] M. Zhang, X. Zhao, W. Li, Y. Zhang, R. Tao, and Q. Du, “Cross-scene joint classification of multisource data with multilevel domain adaption network,” *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[7] W. Li, K. Chen, H. Chen, and Z. Shi, “Geographical knowledge-driven representation learning for remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2021.

[8] Z. LIU, J. YANG, W. WANG, and Z. SHI, “Cloud detection methods for remote sensing images: a survey,” *Chinese Space Science and Technology*, vol. 43, no. 1, p. 1, 2023.

[9] K. Chen, Z. Zou, and Z. Shi, “Building extraction from remote sensing images with sparse token transformers,” *Remote Sensing*, vol. 13, no. 21, p. 4441, 2021.

[10] M. Zhang, W. Li, X. Zhao, H. Liu, R. Tao, and Q. Du, “Morphological transformation and spatial-logical aggregation for tree species classification using hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.

[11] H. Chen, W. Li, S. Chen, and Z. Shi, “Semantic-aware dense representation learning for remote sensing image change detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–18, 2022.

[12] Z. Zhang, W. Xu, Z. Shi, and Q. Qin, “Establishment of a comprehensive drought monitoring index based on multisource remote sensing data and agricultural drought monitoring,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 2113–2126, 2021.

[13] K. Chen, W. Li, S. Lei, J. Chen, X. Jiang, Z. Zou, and Z. Shi, “Continuous remote sensing image super-resolution based on context interaction in implicit function space,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.

[14] C. J. Stubenrauch, W. B. Rossow, S. Kinne, S. Ackerman, G. Cesana, H. Chepfer, L. Di Girolamo, B. Getzewich, A. Guignard, A. Heidinger *et al.*, “Assessment of global cloud datasets from satellites: Project and database initiated by the gewex radiation panel,” *Bulletin of the American Meteorological Society*, vol. 94, no. 7, pp. 1031–1049, 2013.

[15] X. Wu and Z. Shi, “Utilizing multilevel features for cloud detection on satellite imagery,” *Remote Sensing*, vol. 10, no. 11, p. 1853, 2018.

[16] X. Zhang, L. Liu, X. Chen, S. Xie, and L. Lei, “A novel multitemporal cloud and cloud shadow detection method using the integrated cloud z-scores model,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 1, pp. 123–134, 2019.

[17] B. Zhong, W. Chen, S. Wu, L. Hu, X. Luo, and Q. Liu, “A cloud detection method based on relationship between objects of cloud and cloud-shadow for chinese moderate to high resolution satellite imagery,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 11, pp. 4898–4908, 2017.

[18] R. R. Irish, “Landsat 7 automatic cloud cover assessment,” in *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, vol. 4049. SPIE, 2000, pp. 348–355.

[19] R. R. Irish, J. L. Barker, S. N. Goward, and T. Arvidson, “Characterization of the landsat-7 etm+ automated cloud-cover assessment (acca) algorithm,” *Photogrammetric engineering & remote sensing*, vol. 72, no. 10, pp. 1179–1188, 2006.

[20] Z. Zhu and C. E. Woodcock, “Object-based cloud and cloud shadow detection in landsat imagery,” *Remote sensing of environment*, vol. 118, pp. 83–94, 2012.

[21] ——, “Automated cloud, cloud shadow, and snow detection in multitemporal landsat data: An algorithm designed specifically for monitoring land cover change,” *Remote Sensing of Environment*, vol. 152, pp. 217–234, 2014.

[22] S. Qiu, B. He, Z. Zhu, Z. Liao, and X. Quan, “Improving fmask cloud and cloud shadow detection in mountainous area for landsats 4–8 images,” *Remote Sensing of Environment*, vol. 199, pp. 107–119, 2017.

[23] S. Qiu, Z. Zhu, and B. He, “Fmask 4.0: Improved cloud and cloud shadow detection in landsats 4–8 and sentinel-2 imagery,” *Remote Sensing of Environment*, vol. 231, p. 111205, 2019.

[24] Z. An and Z. Shi, “Scene learning for cloud detection on remote-sensing images,” *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 8, no. 8, pp. 4206–4222, 2015.

[25] A. Pérez-Suay, J. Amorós-López, L. Gómez-Chova, J. Muñoz-Marí, D. Just, and G. Camps-Valls, “Pattern recognition scheme for large-scale cloud detection over landmarks,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 3977–3987, 2018.

[26] P. Li, L. Dong, H. Xiao, and M. Xu, “A cloud image detection method based on svm vector machine,” *Neurocomputing*, vol. 169, pp. 34–42, 2015.

[27] S. N. George, “Reconstruction of cloud-contaminated satellite remote sensing images using kernel pca-based image modelling,” *Arabian Journal of Geosciences*, vol. 9, pp. 1–14, 2016.

[28] F. Xie, M. Shi, Z. Shi, J. Yin, and D. Zhao, “Multilevel cloud detection in remote sensing images based on deep learning,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 8, pp. 3631–3640, 2017.

[29] Z. Yan, M. Yan, H. Sun, K. Fu, J. Hong, J. Sun, Y. Zhang, and X. Sun, “Cloud and cloud shadow detection using multilevel feature fused segmentation network,” *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 10, pp. 1600–1604, 2018.

[30] Y. Zhan, J. Wang, J. Shi, G. Cheng, L. Yao, and W. Sun, “Distinguishing cloud and snow in satellite images via deep convolutional network,” *IEEE geoscience and remote sensing letters*, vol. 14, no. 10, pp. 1785–1789, 2017.

[31] A. Francis, P. Sidiropoulos, and J.-P. Muller, “Cloudfcn: Accurate and robust cloud detection for satellite imagery with deep learning,” *Remote Sensing*, vol. 11, no. 19, p. 2312, 2019.

[32] J. H. Jeppesen, R. H. Jacobsen, F. Inceoglu, and T. S. Toftegaard, “A cloud detection algorithm for satellite imagery based on deep learning,” *Remote sensing of environment*, vol. 229, pp. 247–259, 2019.

[33] K. Xu, K. Guan, J. Peng, Y. Luo, and S. Wang, “Deepmask: an algorithm for cloud and cloud shadow detection in optical satellite remote sensing images using deep residual network,” *arXiv preprint arXiv:1911.03607*, 2019.

[34] D. López-Puigdollers, G. Mateo-García, and L. Gómez-Chova, “Benchmarking deep learning models for cloud detection in landsat-8 and sentinel-2 images,” *Remote Sensing*, vol. 13, no. 5, p. 992, 2021.

[35] J. Li, Z. Wu, Z. Hu, C. Jian, S. Luo, L. Mou, X. X. Zhu, and M. Molinier, “A lightweight deep learning-based cloud detection method for sentinel-2a imagery fusing multiscale spectral and spatial features,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2021.

[36] L. Zhang, J. Sun, X. Yang, R. Jiang, and Q. Ye, “Improving deep learning-based cloud detection for satellite images with attention mechanism,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[37] Y. Chen, Q. Weng, L. Tang, Q. Liu, and R. Fan, “An automatic cloud detection neural network for high-resolution remote sensing imagery with cloud–snow coexistence,” *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[38] K. Hu, D. Zhang, and M. Xia, “Cdunet: Cloud detection unet for remote sensing imagery,” *Remote Sensing*, vol. 13, no. 22, p. 4533, 2021.

[39] W. Li, F. Zhang, H. Lin, X. Chen, J. Li, and W. Han, "Cloud detection and classification algorithms for himawari-8 imager measurements based on deep learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022.

[40] Q. He, X. Sun, Z. Yan, and K. Fu, "Dabnet: Deformable contextual and boundary-weighted network for cloud detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2021.

[41] J. Zhang, H. Wang, Y. Wang, Q. Zhou, and Y. Li, "Deep network based on up and down blocks using wavelet transform and successive multi-scale spatial attention for cloud detection," *Remote Sensing of Environment*, vol. 261, p. 112483, 2021.

[42] X. Wu, Z. Shi, and Z. Zou, "A geographic information-driven method and a new large scale dataset for remote sensing cloud/snow detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 174, pp. 87–104, 2021.

[43] K. Hu, D. Zhang, M. Xia, M. Qian, and B. Chen, "Lcdnet: Light-weighted cloud detection network for high-resolution remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 4809–4823, 2022.

[44] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.

[45] H. Fu, Y. Shen, J. Liu, G. He, J. Chen, P. Liu, J. Qian, and J. Li, "Cloud detection for fy meteorology satellite based on ensemble thresholds and random forests approach," *Remote Sensing*, vol. 11, no. 1, p. 44, 2018.

[46] X. Kang, G. Gao, Q. Hao, and S. Li, "A coarse-to-fine method for cloud detection in remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 110–114, 2018.

[47] Y. Yuan and X. Hu, "Bag-of-words and object-based classification for cloud extraction from satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 8, pp. 4197–4205, 2015.

[48] K. Tan, Y. Zhang, and X. Tong, "Cloud extraction from chinese high resolution satellite imagery by probabilistic latent semantic analysis and object-based machine learning," *Remote Sensing*, vol. 8, no. 11, p. 963, 2016.

[49] G. Zhang, X. Gao, Y. Yang, M. Wang, and S. Ran, "Controllably deep supervision and multi-scale feature fusion network for cloud and snow detection based on medium-and high-resolution imagery dataset," *Remote Sensing*, vol. 13, no. 23, p. 4805, 2021.

[50] Z. Wang, B. Fan, Z. Tu, H. Li, and D. Chen, "Cloud and snow identification based on deeplab v3+ and crf combined model for gf-1 wfv images," *Remote Sensing*, vol. 14, no. 19, p. 4880, 2022.

[51] D. Chai, S. Newsam, H. K. Zhang, Y. Qiu, and J. Huang, "Cloud and cloud shadow detection in landsat imagery based on deep convolutional neural networks," *Remote sensing of environment*, vol. 225, pp. 307–316, 2019.

[52] B. Zhang, Y. Zhang, Y. Li, Y. Wan, and Y. Yao, "Cloudvit: A lightweight vision transformer network for remote sensing cloud detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2022.

[53] Z. Zou, W. Li, T. Shi, Z. Shi, and J. Ye, "Generative adversarial training for weakly supervised cloud matting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 201–210.

[54] Y. Li, W. Chen, Y. Zhang, C. Tao, R. Xiao, and Y. Tan, "Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning," *Remote Sensing of Environment*, vol. 250, p. 112045, 2020.

[55] S. Wang, W. Chen, S. M. Xie, G. Azzari, and D. B. Lobell, "Weakly supervised deep learning for segmentation of remote sensing imagery," *Remote Sensing*, vol. 12, no. 2, p. 207, 2020.

[56] Y. Li, T. Shi, Y. Zhang, W. Chen, Z. Wang, and H. Li, "Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 175, pp. 20–33, 2021.

[57] Y. Cao and X. Huang, "A coarse-to-fine weakly supervised learning method for green plastic cover segmentation using high-resolution remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 188, pp. 157–176, 2022.

[58] R. Lian and L. Huang, "Weakly supervised road segmentation in high-resolution remote sensing images using point annotations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2021.

[59] H. Wang, H. Li, W. Qian, W. Diao, L. Zhao, J. Zhang, and D. Zhang, "Dynamic pseudo-label generation for weakly supervised object detection in remote sensing images," *Remote Sensing*, vol. 13, no. 8, p. 1461, 2021.

[60] X. Feng, J. Han, X. Yao, and G. Cheng, "Tcanet: Triple context-aware network for weakly supervised object detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 8, pp. 6946–6955, 2020.

[61] X. Yao, X. Feng, J. Han, G. Cheng, and L. Guo, "Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 675–685, 2020.

[62] C. Fasana, S. Pasini, F. Milani, and P. Fraternali, "Weakly supervised object detection for remote sensing images: A survey," *Remote Sensing*, vol. 14, no. 21, p. 5362, 2022.

[63] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929.

[64] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[65] R. L. Draelos and L. Carin, "Use hirescam instead of grad-cam for faithful explanations of convolutional neural networks," *arXiv e-prints*, pp. arXiv–2011, 2020.

[66] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-cam: Score-weighted visual explanations for convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 24–25.

[67] J. Li, Z. Wu, Q. Sheng, B. Wang, Z. Hu, S. Zheng, G. Camps-Valls, and M. Molinier, "A hybrid generative adversarial network for weakly-supervised cloud detection in multispectral images," *Remote Sensing of Environment*, vol. 280, p. 113197, 2022.

[68] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[69] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.

[70] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 214–223.

[71] Y. Liu, Q. Li, X. Li, S. He, F. Liang, Z. Yao, J. Jiang, and W. Wang, "Leveraging physical rules for weakly supervised cloud detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[72] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[73] W. Li, Z. Zou, and Z. Shi, "Deep matting for cloud detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8490–8502, 2020.

[74] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.