

## Research Article

# Improving Cloud/Snow Detection in Remote Sensing Image with Spatiotemporal Information Fusion

Jianfeng Wen,<sup>1</sup> Hao Zhang,<sup>1</sup> Changxian He,<sup>2</sup> and Gang Xu <sup>1,2</sup>

<sup>1</sup>Zhejiang College of Security Technology, Ouhai District, Wenzhou 325016, China

<sup>2</sup>School of Geosciences and Info-Physics, Central South University, Changsha 410083, China

Correspondence should be addressed to Gang Xu; 20096342@zjcst.edu.cn

Received 5 August 2022; Revised 27 August 2022; Accepted 1 September 2022; Published 17 October 2022

Academic Editor: Feng Ding

Copyright © 2022 Jianfeng Wen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The clouds and snow in optical remote sensing images always interfere with the interpretation of remote sensing images, which even makes an entire image unavailable. In general, the proportion of cloud/snow cover in remote sensing images needs to be clarified to improve the utilization of remote sensing images. The metadata of remote sensing image products contains prior knowledge of spatiotemporal information, such as imaging time, latitude and longitude, and altitude. This paper proposes a remote sensing image cloud/snow detection method that fuses spatial and temporal information. The proposed method can combine spatiotemporal information for feature extraction and stitching, thus improving the accuracy of remote sensing image cloud/snow detection. In this study, the proposed method is trained and tested with a large-scale cloud/snow image dataset. The experimental results show that both the temporal or spatial information alone and the fused temporal and spatial information can improve the cloud/snow detection accuracy in remote sensing images. The easy-to-obtain imaging time information can also significantly improve the detection accuracy for cloud/snow. The proposed method can be used to improve the cloud/snow detection effect of any remote sensing image product containing prior knowledge of spatiotemporal information and has a good application prospect.

## 1. Introduction

Optical remote sensing images have been widely used for their advantages, such as large information capacity and stable geometric properties. However, the optical imaging process is susceptible to interference from clouds and snow. In recent years, with the development of telemetry technology, such as the improvement of UAV endurance in [1], a large number of remote sensing images have been put into use. On the one hand, clouds form an occlusion of the ground cover, with about 1/3 of the ground surface covered by clouds [2], and the occlusion of clouds greatly limits the application of optical images. On the other hand, the spectral characteristics of clouds and snow have many similarities, which makes them prone to misclassification in the image classification process. The above-given two factors will greatly hamper the use and analysis of remote sensing images. In addition, the detection of clouds and

snow still has deep-seated requirements; for example, the accurate detection of clouds and snow can serve the inversion of atmospheric aerosols [3, 4] and reconstruct missing image information [5, 6]. Therefore, accurately and automatically detecting clouds and snow from remote sensing images is of great importance. For example, achieving high-precision cloud detection before satellite image transmission can discover and eliminate high-cloud images in time, thus saving storage space and improving time efficiency.

To reject high-cloudiness images with low information density and improve the efficiency of image utilization, accurate identification of clouds/snow in remote sensing images has been a basic and important research topic. The current cloud/snow automatic detection algorithms include three main types: (1) physical model-based methods, (2) statistical model-based methods, and (3) deep learning-based methods.

Traditional cloud/snow detection methods are mainly based on physical models. The methods are to use the reflectance of a particular image band or calculation of the ratio between bands to identify cloud/snow. One of the most typical methods is the automatic cloud coverage assessment [7, 8], which performs the calculation based on bands 2 to 6 of Landsat-7 ETM+ imagery by using the spectral differences that exist between clouds and snow in different geographical environments to distinguish them. The normalized difference snow index [9], which separates cloud/snow image elements by calculating the ratio of the difference between the green band G and the short IR band IR to the sum, has also been proposed. In addition, an enhanced multitemporal cloud detection algorithm [10] was designed to solve the problem of the over-detection of cloud in snow-covered regions. The physical model-based approach, while effective in obtaining cloud/snow masks of images without relying on pixel-level labels for training, is strongly influenced by the reflectance of the image bands and predefined thresholds.

In recent years, big data has driven the development of statistical models [11]. The statistical model-based approaches typically handle cloud/snow detection tasks in a classification paradigm. Current popular classifiers, such as support vector machines (SVMs) [12] and random forests [13], are commonly used for such tasks. In [14], Amato applied principal component analysis to remote sensing image cloud detection based on statistical theory. In [15], Merchant proposed a cloud detection algorithm based on full probability Bayesian theory. In [16, 17], Bai and Li performed SVM classification based on multiple texture features. Methods based on statistical models can utilize a priori knowledge, such as the spatial pattern information and improve the utilization of remote sensing image information via segmentation models. Moreover, the use of image features greatly reduces the dependence of traditional cloud/snow detection methods on the spectral characteristics. In recent years, deep learning-based methods have shown powerful capabilities in remote sensing feature extraction and classification tasks [18]. The methods treat the task of cloud/snow detection as an image pixel-by-pixel classification process (the pixels are usually divided into three categories: “cloud,” “snow,” and “background”) [19]. In [20], Lei segmented an image into a set of superpixels and then used a neural network to classify these superpixels. In [21], Long and Shelhamer introduced Fully convolutional networks (FCNs) for cloud/snow detection. In [22], Nie and Xu, respectively, introduced DeepLabV3+ for cloud/snow detection. All these approaches implement cloud/snow detection through a deep learning-based semantic segmentation task. For the characteristics and difficulties of the cloud/snow detection task itself, the multiscale convolutional feature fusion method [23] and the multiscale fusion gated cloud detection model MFGNet [24] have also been proposed. They couple the multiscale features of remote sensing images, which effectively improve the cloud detection accuracy. At present, methods based on deep learning can mine the multidimensional and deep-level

features of clouds and snow, thereby enhancing the distinguishability of clouds and snow.

In addition to the spectral properties of the cloud/snow itself, the spatiotemporal information carried in images also helps in cloud/snow detection. In fact, geographic information (such as elevation, latitude and longitude, and imaging time) is often present in remote sensing images as the basic metarecords. In cloud/snow detection, elevation and altitude are important a priori information. For example, in some low altitude or low-latitude regions, snow is unlikely to exist, and the visual appearance of clouds generated in different geographic regions may also differ. To use spatial information to guide the detection of clouds and snow in remote sensing imagery, a geographic information-driven method (GeoInfoNet) for remote sensing cloud/snow detection was proposed in [25]. Unlike previous methods of detection based solely on image data, this method encodes the elevation, and spatial location of the image into a set of geographic knowledge-aided maps and then integrates these maps containing spatial information into the feature extraction network to assist in cloud and snow detection. In [26], Wu and Shi proposed the scene aggregation network, which fuses scene information with remote sensing images to perform scene classification while achieving cloud detection. Although the above models confirm the effectiveness of spatial prior knowledge to drive cloud/snow detection, they ignore the effect of temporal prior knowledge on the cloud/snow distribution. The imaging time, as the time information carried by almost all remote sensing images, is able to synergize with the spatial information. For example, snow at low altitudes occurs more often in winter and rarely in summer; as latitude rises, snow occurs longer each year at higher latitudes; clouds occur more often in the rainy season. The temporal information should be fully utilized to reduce false detection of cloud/snow effectively. Therefore, the fusion of spatiotemporal information can provide richer and more reliable a priori knowledge for cloud/snow detection tasks.

In view of this, this study supervises the cloud/snow detection by incorporating temporal information based on existing studies. Specifically, this study constructs a cloud/snow detection model integrating temporal and spatial information to provide a feasible solution for cloud/snow detection in remote sensing image processing. We conduct ablation experiments for different spatiotemporal information combined into prior knowledge to verify the effects of various types of spatiotemporal prior knowledge on cloud/snow detection. In this paper, we will introduce the specific structure of the model in Section 2; in Section 3, we will introduce the data set and evaluation indicators, and analyze and evaluate the results; in Section 4, we draw a conclusion and look forward to the future research direction.

## 2. Methodology

The model in this paper is a two-branch feature extraction network, as shown in Figure 1. First, spatiotemporal information is encoded into a geographic knowledge-aided map by using a geographic information encoder. Then, the

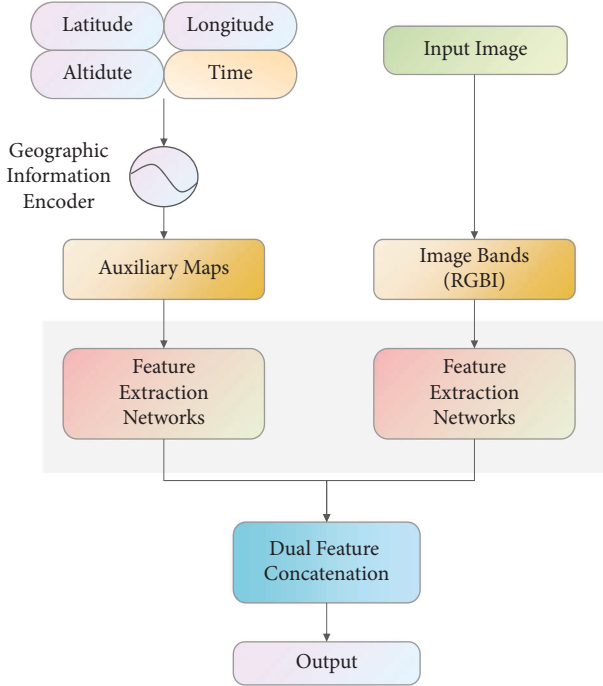


FIGURE 1: Framework of the research model.

original image and the geographic knowledge-aided map are fed into the two branches of the feature extraction network to obtain hierarchical features. Finally, the cloud/snow detection results are obtained by fusing the features of the two branches through a double-feature splicing module.

**2.1. Geographic Information Encoder.** Specifically, to make full use of the a priori geographic knowledge from satellite remote sensing images, the spatial a priori knowledge is encoded into a spatial knowledge-aided map by using a geographic information encoder [25]. To investigate the effect of temporal prior knowledge on cloud/snow distribution, we improve the cloud/snow detection network by incorporating spatiotemporal information by encoding and integrating the imaging temporal information into the geographic knowledge-aided map for feature extraction. In this way, the imaging time can be integrated into the network as a priori knowledge, enabling the model to couple the spatiotemporal a priori knowledge to aid cloud/snow detection from remote sensing images.

The geographic information encoder can be considered the preprocessing module of the model. It corresponds the longitude and latitude information of the remote sensing image to the spatial resolution of the image through affine transformation and obtains a longitude and latitude map of the same size as the image. Then, the longitude and latitude maps are combined with the DEM upsampled to a consistent resolution to obtain a spatial knowledge-aided map of this satellite remote sensing image. The time information is also able to be converted into a global feature by the idea of spatial information encoding. And finally, the imaging time is normalized by dividing the total number of days in the

year, and the time information encoded by this method can better represent the season imaged by the image. In this way, we can integrate the temporal parameters as a channel with the spatial knowledge-aided map to obtain a geographic knowledge-aided map integrating spatiotemporal information. Furthermore, the cloud/snow detection of remote sensing images is assisted by mining the spatiotemporal a priori knowledge of geographic knowledge-aided maps.

Specifically, in each channel of the geographic knowledge-aided map, for a pixel in row  $y$  and column  $x$ , the corresponding elevation is, while the corresponding longitude, latitude, and time are calculated as follows:

$$A_{\text{Long}}(y, x) = A_{\text{Long}}(0, 0) + y * r_{1,1} + x * r_{1,2}, \quad (1)$$

$$A_{\text{Lat}}(y, x) = A_{\text{Lat}}(0, 0) + y * r_{2,1} + x * r_{2,2}, \quad (2)$$

$$A_{\text{Time}}(y, x) = \frac{(A_{\text{month}} * 30 + A_{\text{day}})}{360}. \quad (3)$$

The final coded auxiliary map  $A$  for each input satellite remote sensing image is obtained by stitching the four geographic knowledge auxiliaries in the channel dimension as described above.

To exploit the spatiotemporal correlation knowledge embedded in the geographic knowledge-aided map fused with spatiotemporal information fully, we first use a two-branch network with DenseNet121 [27] structure as the backbone to extract scale-dense features from the input image and the geographic knowledge-aided map separately and then fuse the features extracted from both branches. The two basic modules are “DenseNet-based feature extraction” and “dual-feature concatenation.” The role of the former is to mine the deep features of the original image and the geographic knowledge-aided map, and the role of the latter is to stitch and fuse the features of the two branches and generate a high-precision cloud/snow mask segmentation result.

**2.2. DenseNet-Based Feature Extraction.** To exploit the spatiotemporal correlation knowledge embedded in the geographic knowledge-aided map that incorporates spatiotemporal information fully, the original image and the geographic knowledge-aided map encoding four spatiotemporal prior knowledge are input to the dense feature extraction module. The module is capable of mining deep features of the image and spatiotemporal prior knowledge and using them for the prediction of cloud/snow masks.

Considering the balance of computational efficiency and GPU memory cost, the dense feature extraction module uses DenseNet121 (structure is shown in Table 1) as the backbone network for feature extraction, which consists of multiple dense blocks. In each block, the feature maps from all previous convolutional layers are concatenated. Formally, the feature map  $Ml + 1$  in the  $(l + 1)$ th layer can be expressed as (4), where  $(\cdot)$  represents the nonlinear transformation on the feature:

TABLE 1: The configuration of the dense-base feature extraction module.

Layers	Layer setting
Conv_0	7×7 conv, stride=1
Pool_0	3×3 max_pool, stride=2
Dense_block_1	6 bottlenecks
Transition_1	1×1 conv, 2×2 avg_pool, stride=2
Dense_block_2	12 bottlenecks
Transition_2	1×1 conv, 2×2 avg_pool, stride=2
Dense_block_3	24 bottlenecks
Transition_3	1×1 conv, 2×2 avg_pool, stride=2
Dense_block_4	16 bottlenecks

$$M_{l+1} = \delta(\text{concat}(M_l + M_{l-1} + \dots + M_1)). \quad (4)$$

The module receives an input feature map and first performs feature extraction via an convolutional layer (“Conv\_0”) and an pooling layer (“Pool\_0”) and then sequentially fed into the four dense blocks and three transformation blocks for processing. In the above-given feature extraction process, the number of output feature maps increases as the number of layers deepens. The intensive feature extraction module enables the full exploitation of feature maps, i.e., spatiotemporal a priori knowledge.

**2.3. Dual Feature Concatenation.** The DenseNet-based feature extraction module enables the extraction of fine-grained feature representations required for cloud/snow mask segmentation. To apply the spatiotemporal a priori knowledge extracted from the auxiliary map branch to the cloud/snow detection task, the image features need to be fused with the spatiotemporal a priori knowledge. Therefore, this module is connected to the DenseNet-based feature extraction module (as shown in Figure 2) and fuses spatiotemporal prior knowledge with high-level features of cloud/snow by merging feature maps from different blocks to obtain segmentation results of high-precision cloud/snow masks.

Given that the image feature extraction branch and the auxiliary map feature extraction branch use blue, green, red, and infrared images and coded auxiliary maps as inputs, respectively, the resolution of the feature maps extracted by different blocks is not consistent. Thus, the spatial feature maps of each block are upsampled to the size of the input image through bilinear interpolation, which in turn connects the upsampled feature maps along their channel dimensions. Before stitching,  $1 \times 1$  convolution is used to adjust the channel size of each block’s features so that they have the same number of channels.

All blocks in both branches are concatenated to obtain the concatenation feature  $M$  which can be expressed as  $M = \text{concat}\{M_{\text{img}0}, \dots, M_{\text{img}4}, M_{\text{aux}0}, \dots, M_{\text{aux}4}\}$ . The subscripts 0–4 are the upsampled features of the feature maps in each dense feature extraction module, respectively, and the spatiotemporal knowledge is fused through the stitching of the feature maps. Finally, The feature maps incorporating spatiotemporal information are fed into a convolutional layer with a  $1 \times 1$  filter to generate a pixel-level fractional map of three classes: background S1, cloud S2, and snow S3.

**2.4. Loss Settings.** The output score maps are normalized by using a softmax function and convert the pixel scores to probabilities  $[0, 1]$ . The probability map  $P_t$  of each class  $t = \{1, 2, 3\}$  can be expressed as follows:

$$P_t = \frac{\exp(S_t)}{\sum_{m=1}^3 \exp(S_m)}. \quad (5)$$

Finally, the cross-entropy loss is used as the loss function of the network. Suppose  $y_m \in \{0, 1\}$  represents the ground truth label of the class  $m$ . The loss function is expressed as follows:

$$L = - \sum_{m=1}^3 y_m \log(P_m). \quad (6)$$

### 3. Results and Analysis

**3.1. Experimental Design and Dataset.** To investigate the role of spatiotemporal prior knowledge in remote sensing image cloud/snow detection, this study conducts experiments on how the model affects the cloud/snow mask extraction accuracy when different spatiotemporal information is added. The effect of cloud/snow mask extraction on remote sensing images assisted by spatiotemporal a priori knowledge is verified through comparative experiments.

In this study, we use the large-scale cloud/snow detection dataset “Levir\_CS” [25] for model training, which contains a total of 4168 GF-1 WFV scenes. The scenarios in the dataset are distributed across the globe, as shown in Figure 3. Various types of topographical features are taken into account, such as plains, plateaus, bodies, deserts, and glaciers. Complex landforms formed by different combinations of landforms also exist. Figure 4 shows example scenarios. Given that these scenarios are globally distributed, they may include multiple types of climatic conditions, such as desert climate or ocean climate. All scenes were imaged from May 2013 to February 2019 and are available for download at <http://www.cresda.com/>.

In this dataset, level-1A product data that are radiometrically calibrated but not geometrically corrected by the system are used for each scenario to improve time efficiency, as needed for cloud/snow detection in real-world situations. The pixel-level label masks of all images in the dataset are divided into three categories: “background” (labeled 0), “cloud” (labeled 128), and “snow” (labeled 255). The number of pixels of the three types is as follows: the background occupies the largest number of pixels (79.2%), and the snow occupies the smallest number of pixels (2.2%). Cloud pixels account for 18.6% of the total number of pixels. There were 3068 images for training and 1100 for testing. The following observations can be obtained through visual interpretation:

- (1) The three types of pixels are remarkably different in various locations. In particular, the background varies a lot.
- (2) Clouds are common in different geographical locations.

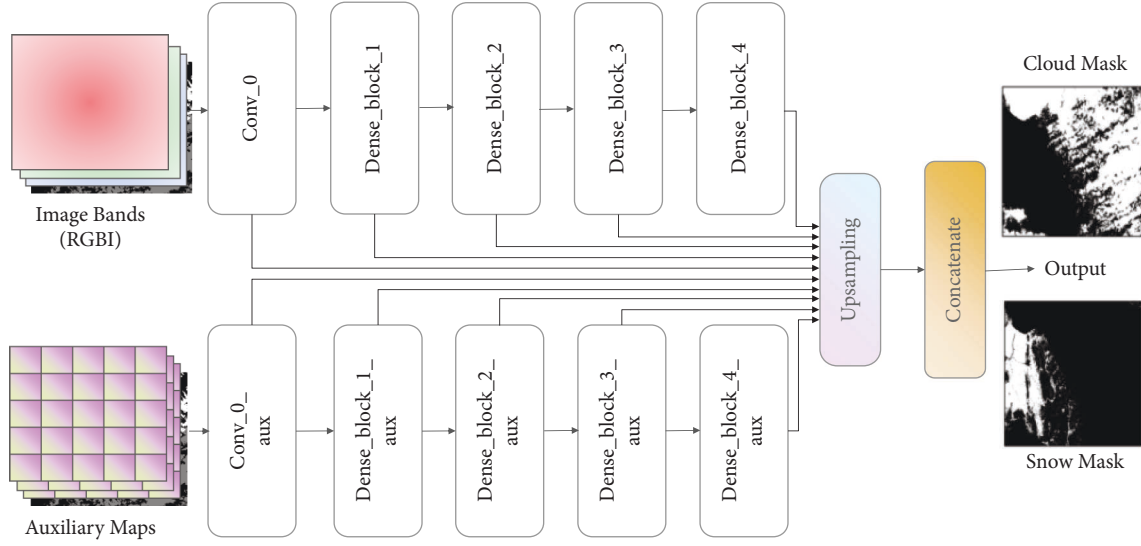


FIGURE 2: DenseNet-based feature extraction.

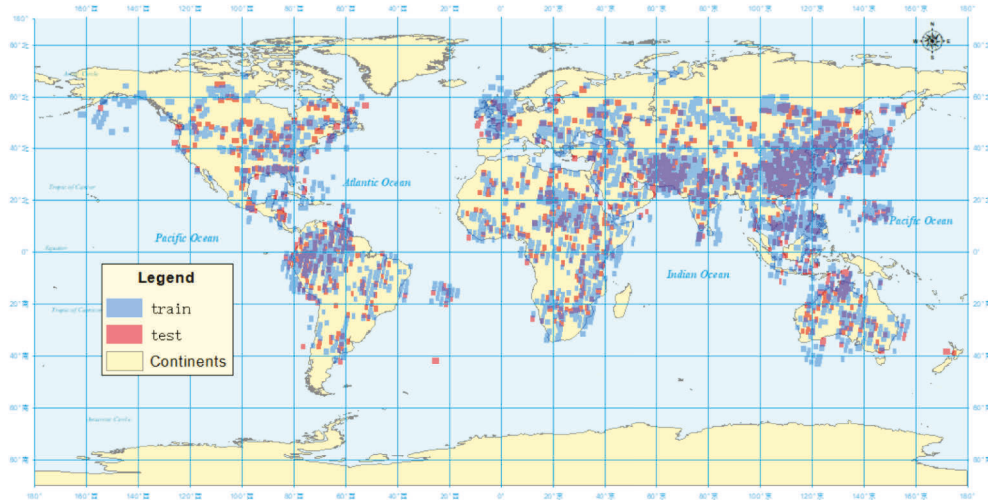


FIGURE 3: Distribution of data sets.

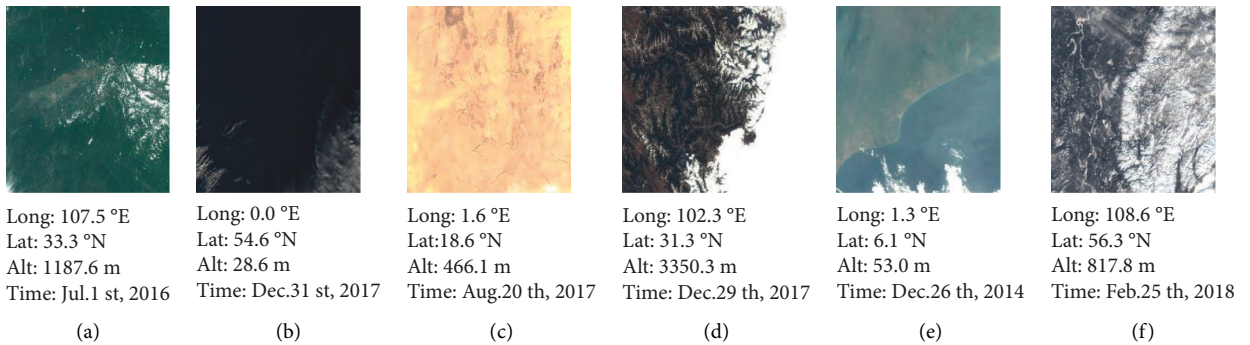


FIGURE 4: Different landscapes in the dataset. (a) Plain. (b) Water. (c) Dessert. (d) Plateau. (e) Plain&amp;Water. (f) Ice&amp;Water.

(3) In terms of latitude, most of the snow is concentrated in high latitudes (see Figure 4(f) for an example). There is almost no snow in the equatorial regions (see Figures 4(a), 4(c), 4(e) for examples).

(4) In terms of altitude, the cloud amount is higher in the region below 500 m (see Figures 4(a), 4(b), 4(e) for example), and the snow accumulation gradually increases with increasing altitude (see Figures 4(d),

4(f)) for example). High-altitude areas are usually mountainous, while the snow accumulation varies regularly with the seasons. From the above statistics.

- (5) In terms of imaging time, snow at low altitudes and low latitudes often occur in winter (see Figure 4(f) for an example), but rarely in summer (see Figure 4(c) for an example), the use of spatial and temporal information is necessary for cloud/snow detection.

**3.2. Parameters and Evaluation Indexes.** In the experiments, different prior knowledge is encoded, combined and input to the auxiliary branch for training to explore the role of spatiotemporal prior knowledge in remote sensing image cloud/snow detection. The model is trained for 200 epochs to reach convergence.

The scenes in the above-given dataset are randomly divided into two datasets: 3068 scenes for a training set and 1100 scenes for a test set.

In this study, the prediction results are evaluated quantitatively by using recall, precision, F1-score, accuracy and IoU. The definition and calculation formula of each parameter are as follows:

True positives (TP): the number of samples that are actually positive and correctly classified as positive by the classifier

False positives (FP): the number of samples that are actually negative but are incorrectly classified as positive by the classifier

False negatives (FN): the number of samples that are actually positive cases but are incorrectly classified as negative cases by the classifier

True-negatives (TN): the number of samples that are actually negative and correctly classified as negative by the classifier

IoU defines the overlap between the labeled and predicted regions.

$$\text{recall} = \frac{TP}{TP + FN}, \quad (7)$$

$$\text{precision} = \frac{TP}{TP + FP}, \quad (8)$$

$$\text{F1score} = ((2 * \text{recall} * \text{precision}) / (\text{recall} + \text{precision})), \quad (9)$$

$$\text{accuracy} = (TP + TN) / (TP + TN + FP + FN), \quad (10)$$

**3.3. Results Evaluation.** The results show that the inclusion of different auxiliary information in the cloud/snow detection task is effective. Compared with the advanced cloud/snow detection methods, our method also shows excellent performance and achieves the best accuracy and IoU, and the time cost is acceptable as shown in Table 2.

TABLE 2: Quantitative comparisons of different methods.

method	Accuracy	IoU	Time
DeepLapV3+	0.9665	0.6259	5.76 s
CDnetV1 [28]	0.9506	0.7883	0.48 s
CDnetV2 [29]	0.9462	0.7523	<b>0.42 s</b>
GeoInfoNet [25]	0.9758	0.8598	2.50 s
ST+(ours)	<b>0.9812</b>	<b>0.8653</b>	2.86 s

The bold values represents the proposed method which achieve the best accuracy.

Tables 3–5 show the accuracy of cloud/snow detection using different spatiotemporal a priori knowledge-aided remote sensing images. The background occupies most of the area in the remote sensing images, and there is a large gap between the background and the clouds and snow. Thus, the background has achieved a high detection accuracy without introducing prior knowledge. The introduction of prior knowledge can slightly improve the IoU of background detection, in which the model incorporating spatiotemporal information can achieve the highest detection accuracy, as shown in Table 3.

For the detection of clouds in images, temporal or spatial information alone is of limited use to improve the accuracy of cloud detection. The model that introduces DEM as a priori knowledge has the highest accuracy improvement, with an IoU improvement of 0.55%. Meanwhile, the model incorporating spatiotemporal information improves IoU by 1.16%. Table 3 is the detection accuracy of clouds in images. The results in Table 4 demonstrate that the joint spatiotemporal information can effectively improve the detection accuracy for clouds in remote sensing images.

Given that snow accounts for the least amount of the image, introducing any separate spatiotemporal prior knowledge can effectively improve the snow detection accuracy. That is, a strong correlation exists between snow distribution and both spatial and temporal properties, which is in line with the perception. The model incorporating spatiotemporal information obtains the highest accuracy, with an 11-point improvement in IoU (from 60% to 71%). As shown in Table 5, this finding further confirms the effectiveness of the fusion of spatiotemporal information to improve the accuracy of cloud/snow detection.

Based on the above-given quantitative analysis of the experimental results, training with a single spatiotemporal information-aided map can improve cloud/snow detection compared with using images alone. The use of the deep convolutional neural networks can mine the knowledge of correlation between spatiotemporal a priori information and the distribution of surface clouds and snow. From Table 4, the use of temporal information alone is less effective for cloud detection, although some improvement is realized. By contrast, the addition of temporal information alone can improve the snow IoU by 7 points (from 60% to 67%), allowing for a very significant improvement. The extent of using some single spatial information to improve the detection accuracy for the cloud is similar to the effect when using temporal information alone. The reason may be that the accuracy of the IoU for cloud detection is already high enough

TABLE 3: Detection accuracy of background in images.

Knowledges	Recall	Precision	IoU	F1-score
Null	0.9746	0.9944	0.9693	0.9844
DEM + Spatial + Time	<b>0.9843</b>	<b>0.9893</b>	<b>0.9739</b>	<b>0.9868</b>
DEM + Spatial [25]	0.9805	0.9912	0.9721	0.9858
DEM	0.9788	0.9927	0.9719	0.9857
Spatial	0.9780	0.9945	0.9728	0.9862
Time	0.9853	0.9868	0.9725	0.9860

The bold values represents the proposed method which achieve the best accuracy.

TABLE 4: Detection accuracy of clouds in images.

Knowledges	Recall	Precision	IoU	F1-score
Null	0.9711	0.9233	0.8986	0.9466
DEM + Spatial + Time	<b>0.9523</b>	<b>0.9537</b>	<b>0.9102</b>	<b>0.9530</b>
DEM + Spatial [25]	0.9625	0.9384	0.9054	0.9503
DEM	0.9662	0.9336	0.9041	0.9496
Spatial	0.9723	0.9258	0.9019	0.9485
Time	0.9410	0.9557	0.9017	0.9482

The bold values represents the proposed method which achieve the best accuracy.

TABLE 5: Detection accuracy of snow in images.

Knowledges	Recall	Precision	IoU	F1-score
Null	0.9672	0.6149	0.6024	0.7518
DEM + Spatial + Time	<b>0.9494</b>	<b>0.7399</b>	<b>0.7118</b>	<b>0.8317</b>
DEM + Spatial [25]	0.9363	0.7371	0.7020	0.8248
DEM	0.9512	0.7025	0.6781	0.8081
Spatial	0.9471	0.7269	0.6986	0.8225
Time	0.9058	0.7199	0.6698	0.8022

The bold values represents the proposed method which achieve the best accuracy.

and stable at around 90%, with limited room for improvement. However, Table 4 indicates that the addition of any of the spatiotemporal information components improves the accuracy of snow detection by about 7 points. Specifically, simply adding latitude and longitude information can even boost 10 points. Accordingly, the occurrence of snow shows a strong correlation with the latitude and longitude where it is located, and it can be speculated that the climatic zones of different latitudes cause this difference.

From the comparison experiments with different combinations of spatiotemporal information-aided components, the model with fused spatiotemporal information obtains the highest precision for cloud/snow detection. This result shows that we can further improve the results of cloud/snow detection by introducing inexpensive and easily accessible temporal information into the results with only spatial information. In summary, our proposed remote sensing image cloud/snow detection method incorporating spatiotemporal information achieves the highest accuracy in cloud/snow detection, with IoU reaching 91% for cloud and 71% for

snow, which is a stable performance improvement compared with the method incorporating only spatial information knowledge in Literature [25].

The results presented in Figure 5 further corroborate the conclusions of the quantitative analysis. On the one hand, individual temporal or spatial information can be useful for cloud/snow detection. The most accessible temporal information can effectively improve cloud detection, while the spatial information contributes most to the improvement of snow detection accuracy. On the other hand, the models incorporating spatiotemporal prior knowledge have the best results for cloud/snow detection. In particular, the introduction of temporal information can effectively improve the detection of clouds in images, as shown in the red box in Figure 5. In the high-altitude region shown in the blue box, the clouds have a high probability of being misdetected as snow by introducing a temporal prior or a spatial prior alone. The fusion of temporal and geographical can form a complementary effect and effectively alleviate the problem of cloud/snow false detection.



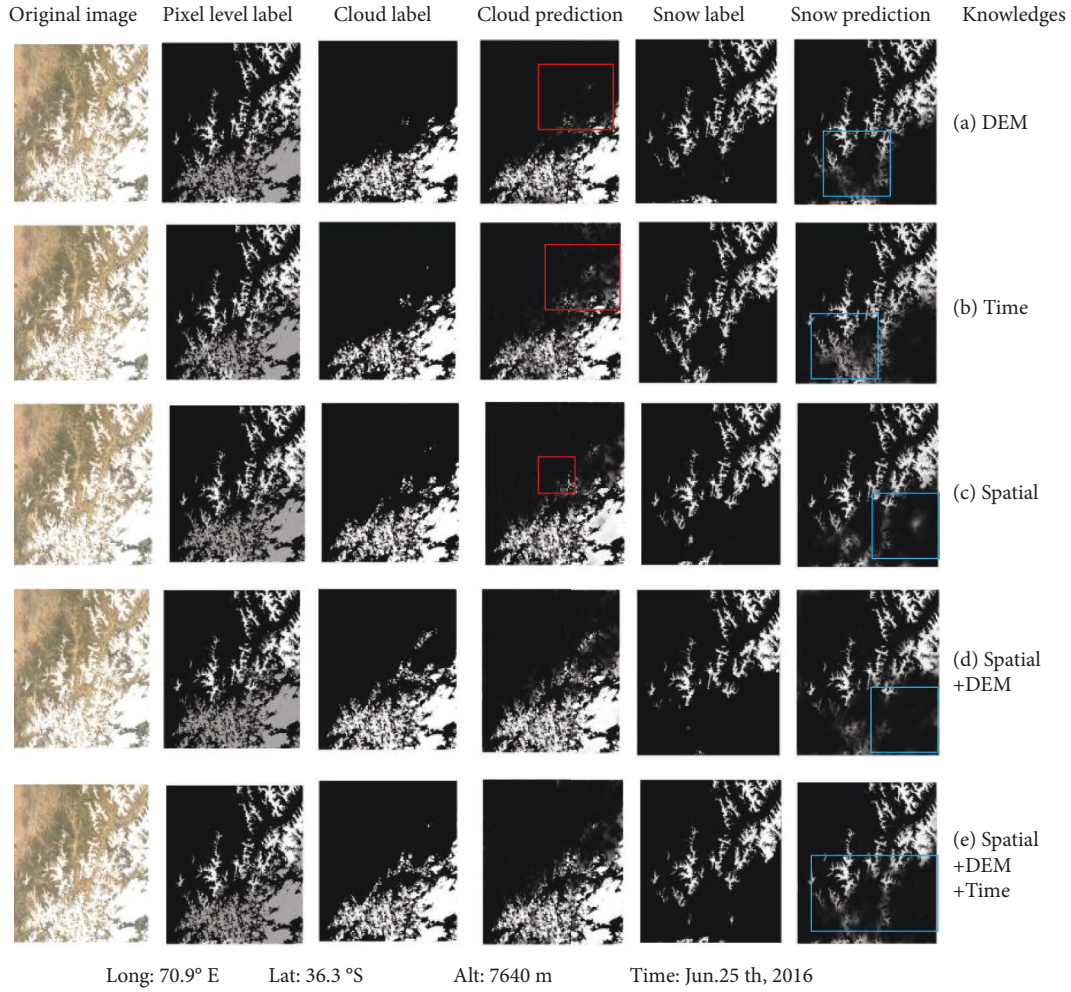


FIGURE 5: Results of cloud snow detection with the introduction of different prior knowledge.

#### 4. Conclusion

Inspired by the fact that cloud/snow has a strong seasonal a priori, this paper proposes a cloud/snow detection model integrating temporal and geographical information based on existing research on geographic knowledge-driven cloud/snow detection. Among all the a priori knowledge, temporal information contributes the most to improving cloud detection accuracy, while latitude and longitude information contributes the most to improving snow detection accuracy, and fusing temporal and spatial information at the same time can obtain the highest accuracy. This research can effectively complement the existing algorithms that do not fully utilize the a priori knowledge of imaging time. This method is expected to be used for the fast detection of high cloud cover images driven by temporal information and the detection of cloud snow essence fused with spatiotemporal information. Our method also has some limitations, for example, it only performs simple coupling splicing of spatiotemporal features, and does not dig into their correlations.

In the future, research can be further improved from several perspectives, including incorporating richer a priori

knowledge, such as scene information, to assist remote sensing images for cloud/snow detection, as well as introducing an attention mechanism to explore the more essential and deeper relationship between cloud/snow and a priori knowledge in remote sensing images.

#### Data Availability

The image dataset used to support the findings of this study can be downloaded from <https://github.com/permanentCH5/GeoInfoNet>. The dataset is cited at relevant places within the text as references [25].

#### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

#### Acknowledgments

This work was supported in part by the Natural Science Foundation of Hunan Province, China (Grant no. 2020JJ4691).



## References

- [1] C. Cai, J. Wang, F. Zhang, X. Liu, P. Zhang, and Y.-G. Zhou, "A multichannel wireless UAV charging system with compact receivers for improving transmission stability and capacity," *IEEE Systems Journal*, vol. 16, no. 1, pp. 997–1008, March 2022.
- [2] Z. Zhu and C. E. Woodcock, "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.
- [3] E. Carboni, "GOME aerosol optical depth retrieval over ocean: correcting for the effects of residual cloud contamination," *Atmospheric Environment*, vol. 40, no. 36, pp. 6975–6987, 2006.
- [4] J. Bi, J. H. Belle, Y. Wang, A. I. Lyapustin, A. Wildani, and Y. Liu, "Impacts of snow and cloud covers on satellite-derived PM<sub>2.5</sub> levels," *Remote Sensing of Environment*, vol. 221, pp. 665–674, 2019.
- [5] P. Singh and N. Komodakis, "Cloud-Gan: cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks," in *Proceedings of the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1772–1775, IEEE, Valencia, Spain, 22–27 July 2018.
- [6] X. Li, L. Wang, Q. Cheng, P. Wu, W. Gan, and L. Fang, "Cloud removal in remote sensing images using nonnegative matrix factorization and error correction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 148, pp. 103–113, 2019.
- [7] R. R. Irish, "Landsat 7 automatic cloud cover assessment," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 4049, pp. 348–355, 2000.
- [8] 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, pp. 1179–1188, 2006.
- [9] D. J. Selkowitz and R. R. Forster, "An automated approach for mapping persistent ice and snow cover over high latitude regions," *Remote Sensing*, vol. 8, no. 1, p. 16, 2015.
- [10] X. Chen, X. Zhang, L. Liu, and X. Wang, "Enhanced multi-temporal cloud detection algorithm for optical remote sensing images," *Remote Sensing*, vol. 10, no. 11834, pp. 280–290, 2019.
- [11] R. Jiang, Y. Xin, Z. Chen, and Y. Zhang, "A medical big data access control model based on fuzzy trust prediction and regression," *Applied Soft Computing*, vol. 177, pp. 1–20, 2022.
- [12] P. Li, L. Dong, H. Xiao, and M. Xu, "A cloud image detection method based on SVM vector machine," *Neurocomputing*, vol. 169, no. 2, pp. 34–42, 2015.
- [13] N. Ghasemian and M. Akhoondzadeh, "Introducing two Random Forest based methods for cloud detection in remote sensing images," *Advances in Space Research*, vol. 62, no. 2, pp. 288–303, 2018.
- [14] U. Amato, A. Antoniadis, V. Cuomo et al., "Statistical cloud detection from SEVIRI multispectral images," *Remote Sensing of Environment*, vol. 112, no. 3, pp. 750–766, 2008.
- [15] C. J. Merchant, A. R. Harris, E. Maturi, and S. Maccallum, "Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval," *Quarterly Journal of the Royal Meteorological Society*, vol. 131, no. 611, pp. 2735–2755, 2005.
- [16] T. Bai, D. Li, K. Sun, Y. Chen, and W. Li, "Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion," *Remote Sensing*, vol. 8, no. 9, p. 715, 2016.
- [17] Z. Li, H. Shen, H. Li, G. Xia, P. Gamba, and L. Zhang, "Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery," *Remote Sensing of Environment*, vol. 191, pp. 342–358, 2017.
- [18] Y. Yu, M. Rashidi, B. Samali, M. Mohammadi, T. N. Nguyen, and X. Zhou, "Crack detection of concrete structures using deep convolutional neural networks optimized by enhanced chicken swarm algorithm," *Structural Health Monitoring*, vol. 21, no. 5, pp. 2244–2263, 2022.
- [19] T. Shi, Q. Xu, Z. Zou, Z. Shi, P. Gamba, and L. Zhang, "Automatic raft labeling for remote sensing images via dual-scale homogeneous convolutional neural network," *Remote Sensing*, vol. 10, no. 7, p. 1130, 2018.
- [20] L. Wang, Y. Chen, L. Tang, R. Fan, and Y. Yao, "Object-based convolutional neural networks for cloud and snow detection in high-resolution multispectral imagers," *Water*, vol. 10, no. 11, p. 1666, 2018.
- [21] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.
- [22] Z. Nie, J. Xu, and S. Zhang, "Analysis on DeepLabV3+ performance for automatic steel defects detection," *Advances in Intelligent Systems and Computing*, vol. 789, pp. 141–149, 2020.
- [23] Z. Li, H. Shen, Q. Cheng, Y. Liu, S. You, and Z. He, "Deep learning-based cloud detection for medium and high resolution remote sensing images of different sensors," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 197–212, 2019.
- [24] J. Yu, Y. Li, X. Zheng, Y. Zhong, and P. He, "An effective cloud detection method for gaofen-5 images via deep learning," *Remote Sensing*, vol. 12, no. 13, p. 2106, 2020.
- [25] 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.
- [26] X. Wu and Z. Shi, "Scene aggregation network for cloud detection on remote sensing imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.
- [27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.
- [28] J. Yang, J. Guo, H. Yue, Z. Liu, H. Hu, and K. Li, "CDnet: CNN-Based Cloud Detection for Remote Sensing Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 6195–6211, 2019.
- [29] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou, and K. Li, "CDnetV2: CNN-Based Cloud Detection for Remote Sensing Imagery With Cloud-Snow Coexistence," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 700–713, 2021.