FSC-USNet: Fractional Snow Cover Retrieval on the Tibetan Plateau by Integrating Improved Attention Mechanisms

Xu Liu, *Member, IEEE*, Xi Kan, *Member, IEEE*, Yonghong Zhang, *Member, IEEE*, Linglong Zhu, *Member, IEEE*, Qi Liu, *Senior Member, IEEE*, Zhou Zhou, *Member, IEEE*, and Guangyi Ma, *Member, IEEE*

***Abstract*—Snow on the Tibetan Plateau (TP) is not only a freshwater resource for the major rivers in Asia but also plays a significant role in adjusting temperature by reflecting solar radiation. Fractional snow cover mapping with fine spatial and temporal resolution is of great significance for clarifying snow resources and accurately managing snow water resources. However, due to the complex TP topography, the existing fractional snow cover (FSC) retrieval methods are affected by a variety of disturbance factors, resulting in a decrease in accuracy. In this study, a deep learning-based FSC retrieval method, FSC U-shape net, is proposed to improve snow cover mapping accuracy. The input images of FengYun-4A advanced geosynchronous radiation imager and geographic data are extracted using the proposed spatial-channel feature extraction module to characterize the shallow features with texture information into high-dimensional feature images. Additionally, an attention mechanism is introduced to improve the feature differences between different FSC degrees. Finally, the correlation of the decoded features in the channel direction is mined using the channel refinement module to obtain the final FSC results. In this study, a backpropagation artificial neural network, random forest, ResNet_FSC, and UNet are trained, compared, and validated against the MOA10A1 FSC product. The results show that the proposed method effectively mitigates the problems of unclear texture of snow edges, poor robustness, and underestimation in some areas, which exist in other models. Additionally, the proposed method has higher accuracy, with an R^2 and explained variance score reaching 0.7182 and 0.7332, respectively. Compared to the MOD10A1 snow product, the proposed method exhibits higher detection accuracy in mountainous areas with high snow cover and significantly improves the low snow cover detection rate.**

***Index Terms*—Attention mechanism, deep learning, fractional snow cover, FengYun-4A (FY4A), Tibetan Plateau.**

I. Introduction

SNOW is an important component of the Earth's surface and one of the most active natural factors [1], [2]. Its strong albedo in the visible band, low thermal conductivity, and strong heat absorption during the melting process [3], [4] directly affect the radiation balance and energy exchange at the surface at global and continental scales [2]. Therefore, snow is an important indicator of global climate change [5].

The Tibetan Plateau (TP), referred to as the "Roof of the World" [6], is recognized as one of three principal stable snow accumulation regions in China. The variations in snow cover exhibit a highly prominent predictive value for climate change not only in China but also on a global scale [6]. Existing research has indicated that East Asian atmospheric circulation and summer monsoons are significantly influenced by the dynamic effects of TP snow cover [7], [8]. Fractional snow cover (FSC), as one of the pivotal indicators in snow research [3], can offer more refined and precise snow area parameters for areas with mixed pixels and regions with blurred snow boundaries. Hence, constructing a high-precision and high-efficiency FSC retrieval method is scientifically and practically significant. However, due to the high altitudes, rugged terrain, and vast snow coverage of the TP, manual field techniques for snow detection have become exceedingly challenging. Therefore, remote sensing technology has emerged as the most effective method for monitoring snow cover in the TP region [9]. Due to the unusual variability of the snow on the TP and the presence of cloud interference, high temporal resolution remote sensing data have become an important means of dynamically monitoring snow parameters and filtering out cloud interference.

Currently, the methods for retrieving FSC using remotely sensed data can be broadly categorized into linear regression-based methods [10], [11], hybrid image decomposition methods based on spectral analysis [12], [13], and methods based on machine learning [14], [15]. Salomoson established a linear relationship between snow cover and the normalized difference

Manuscript received 10 November 2023; revised 27 December 2023 and 19 January 2024; accepted 24 January 2024. Date of publication 30 January 2024; date of current version 30 May 2024. This work was supported in part by the National Natural Science Foundation of China, under Grant 42105143 and Grant 42305158, and in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China, under Grant 21KJB170006 and Grant 23KJB170025. *(Corresponding authors: Xi Kan; Yonghong Zhang.)*

Xu Liu, Qi Liu, and Zhou Zhou are with the School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: 202212490498@nuist.edu.cn; qi.liu@nuist.edu.cn; 202212490524@nuist.edu.cn).

Xi Kan and Linglong Zhu are with the School of Internet of Things, Wuxi University, Wuxi 214105, China (e-mail: kanxi@cwxu.edu.cn; llzhu@cwxu.edu.cn).

Yonghong Zhang is with the School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China, and also with the School of Automation, Wuxi University, Wuxi 214105, China (e-mail: zyh@nuist.edu.cn).

Guangyi Ma is with the School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: gyma@nuist.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3360087

snow index (NDSI) [16] through a linear regression method, and this method was later used to produce version 005 of the MODIS global FSC product [17]. However, studies [18], [19] found that MOD10A1 FSC had lower accuracy in the TP region through Landsat8 validation. To further improve FSC retrieval accuracy, Zhang [18] plotted MODIS FSC maps of the TP by means of a four-segment segmentation function, and the standard error of snow cover extracted by this method was reduced from the original 0.35 to 0.22, which significantly improved the accuracy of snow cover mapping. Follow-up studies have shown that the complex topography, diverse land cover types, and scattered snow distribution in the snow region can cause major interference with the linear regression method and spectral decomposition process [15], [20], [21], especially during the transition period between snow accumulation and melting, which affects the retrieval result accuracy [22], [23].

With the development of machine learning technology, researchers have begun to use machine learning to establish the relationship between input data and FSC. Data, such as land cover type and geographic data, are beginning to be used as auxiliary data to provide more comprehensive auxiliary features for FSC retrieval. Dobreva et al. [15] constructed a three-layer backpropagation artificial neural network (BPANN) and achieved comparable results to the MOD10A1 FSC product in the flat Northern Hemisphere using MODIS surface albedo, NDSI, normalized difference vegetation index, and surface cover type as inputs. As subsequent research has advanced, Liu [9] verified that in the TP region, machine learning algorithms exhibit superior precision compared to alternative methods. Hou [20] employed the BPANN method to achieve results surpassing those obtained from the MOD10A1 product. In addition, with the advancement of unmanned aerial vehicle (UAV) technology, UAVs equipped with small cameras have become an effective method for collecting snow data due to their high spatial resolution and flexibility. Liang et al. [1] performed FSC mapping based on UAV data and MODIS data using methods such as linear regression and BPANN. They verified the snow map extracted using Landsat8, which showed that the BPANN method based on MODIS and UAV data was more similar to the Landsat8 snow map. The essence of machine learning is learning to fit a large quantity of input data, which is the fundamental reason why machine learning methods can outperform traditional methods. However, the BPANN and random forest methods only focus on the input features in the channel direction, completely ignoring the features present in the image space or neighboring pixels, which may result in poor modal robustness.

To improve the overall utilization of data information, convolution-based deep learning methods have been applied in FSC retrieval studies. Nihawan et al. [24] combined multispectral imagery captured by satellites with auxiliary information such as geographic elevation and used a deep learning method to obtain satisfactory results in the Himalayan Khiroi region. Zhang et al. [25] introduced the convolution-based deep learning method ResNet into FSC retrieval and drew a high-precision FSC map of the Xinjiang region based on FY4A data. The accuracy was verified to be better than that of the MOD10A1 FSC snow product and other classical methods. Subsequently, based on AMSR2 and MODIS data, Xing et al. [26] proposed a new "area-to-point" depth estimation method by combining a convolutional neural network and residual block, which achieved good estimation accuracy and is expected to be applied in other regions. Although the convolutional operation accounts for both the channel features and the connection between neighboring pixels, the increase in the number of network layers will inevitably cause irreversible information loss, which is very unfavorable to FSC retrieval.

With the operation of advanced geostationary meteorological satellites, such as the FengYun-4A advanced geosynchronous radiation imager (FY4A/AGRI), the data sources that can be used for large-scale FSC mapping have become more reliable and near real-time. Compared with other low temporal resolution satellite data, FY4A can obtain snow information in an all-weather stable manner, ensuring the reliability and continuity of snow monitoring, and capturing snow cover and changes in a more timely manner, which is very useful for emergency decision-making and disaster management. In addition, due to the unusual variability of the snow on the TP and the interference of cloud cover, high temporal resolution remote sensing data have become an effective means to dynamically monitor the snow parameters and to filter out the interference of cloud cover.

In this study, we employ high-resolution Landsat 8 satellite images to extract FSC as reference data and utilize FY4A images in conjunction with geographic elevation data as input to generate a wide-area, fine temporal resolution snow cover map for the TP region. To improve the accuracy of the FSC retrieval method, we compared the efficiency of various machine learning methods in the FSC retrieval process in the TP region and developed a deep learning-based FSC retrieval method with a shallower depth, which can produce more accurate FSC maps based on the FY4A/AGRI imagery data in the TP region. The main features of this article are as follows.

1) Based on the FY4A high temporal resolution satellite data, a snow cover dataset is produced to learn the snow reflectivity characteristics under various topographic features and to improve the retrieval accuracy of snow cover.
2) A new spatial and channel coding module and decoding module are designed to more comprehensively mine the potential high-dimensional features in the input data.
3) An improved attention mechanism and channel refinement module are added to strengthen the snow features to suppress the interference signal, and the correlation between channels is gradually refined by the channel refinement module to output the snow coverage inversion results.

II. Datasets

A. FY4A/AGRI Data

FY4A represents the inaugural satellite within the second-generation cohort of China's geostationary orbit quantitative remote sensing satellites, which carries an advanced geosynchronous radiation imager in the inheritance and development of Fengyun-2 based on developing fourteen data channels, covering the visible, short infrared, mid-wave infrared, and long wave infrared bands. FY4A has a higher temporal resolution of

TABLE I
FY-4A/AGRI RADIOMETER SPECIFIC CHANNEL INFORMATION

Number	Band type	Bandwidths	Use
1	Visible and Near Infrared	0.45–0.49 μm	Small particle aerosols
2		0.55–0.75 μm	Vegetation, Image Alignment
3		0.75–0.90 μm	Vegetation, aerosols over water
4	short-wave infrared	1.36–1.39 μm	cirrus recognition (cloud)
5		1.58–1.64 μm	Low cloud/snow identification
6		2.1–2.35 μm	cirrus (cloud)
7	mid-wave infrared	3.5–4.0 μm	High albedo targets(clouds)

the full disk image observation time of only 15 min. Additionally, FY4A can accomplish a wide range of fine temporal-resolution detections. In this study, the main use of FY4A/AGRI L1 level 1 to 7 channel data and data-specific channel information is shown in Table I.

B. Land Imager Landsat 8 OLI Data

Landsat 8 is the eighth satellite in the Landsat program launched by the National Aeronautics and Space Administration (NASA). Its primary payloads consist of an operational land imager (OLI) for land imaging and a thermal infrared sensor (TIRS) for thermal infrared sensing. The combined wavelength coverage spans from 0.43 to 1.38 μm, with a spatial resolution of 30 m and a revisit period of 16 days. Landsat8 data at a spatial resolution of 30 m are rich in spectral information and can be used to produce reliable FSC maps based on surface reflectance [4]. Many previous studies have used snow FSC maps extracted from Landsat 8 as reference data [20], [27]. In this research, following the example of other studies in obtaining FSC, Landsat 8 images extracted from clear sky conditions with less than 10% cloud cover were selected as the real reference data.

C. MOA10A1 Snow Products

The MODIS daily snow product MOD10A1 V006, provided by NASA, is an internationally recognized and widely accepted leading product in the field of remote sensing for snow, with a spatial resolution of 500 m. MODIS was carried on two satellites, Terrra and Aqua, and the transit times were 10:30 and 13.30 local time per day, respectively, which is a small time difference from the FY4A/AGRI data used in this article to produce the dataset. Therefore, areas with relatively stable snow conditions were selected for comparison and validation. Snow cover products and binary snow cover data are no longer provided in the newly released version (V006) of the MODIS snow cover product, having been replaced with NDSI data. Therefore, this article synthesizes existing studies and produces snow cover based on NDSI data to be used as a comparative validation. It has been shown that MOD10A1 FSC calculated by the recommended equations is less accurate on the TP than in other regions [1], [18]. On one hand, due to the complex topography and land cover of the TP, the global NDSI reference threshold is no longer applicable in the TP region, and Zhang et al. [28] have shown through extensive experiments that an NDSI threshold of 0.29 is better than the global reference threshold of 0.4 in the TP. On the other hand, Liang et al. [1] fitted a linear relationship between MOD10A1 NDSI and FSC in the TP using snow cover extracted from UAV imagery as a label, which implies that the recommended equation may no longer be applicable in the TP region as well. Therefore, in this article, the MOD10A1 FSC is determined based on the fitting relationship proposed by Liang with a selected threshold value of 0.29. The MOD10A1 FSC equation fitted by Liang et al. [1] is shown in the following:

$$\mathrm{FSC}_{\mathrm{MOD}} = 0.8286 \times \mathrm{NDSI} + 0.3941. \tag{1}$$

D. Geographic Data

Elevation data reflect the geographical characteristics of the study area, providing rich features and a foundation for detecting fragmented snow cover. This facilitates the improvement of FSC retrieval accuracy and results in a more realistic distribution pattern in the FSC map. In this research, we use 30-m resolution digital elevation model (DEM) data [29] from the shuttle radar topography mission.

E. Data Processing

In this article, FSC extracted from Landsat 8 imagery is used as label for training the model, but the slope, aspect, and elevation can cause variations in reflectance for similar features with different terrain positions [30], [31], which has an impact on the reliability of extracting FSC with Landsat 8. Therefore, in this article, the SCS + C [32] terrain correction method has been applied to Landsat 8 imagery to eliminate the errors introduced due to terrain factors. The method uses the sun-canopy-sensor (SCS) with a semi-empirical moderator (C) to account for diffuse radiation, comparing to cosine, C, Minnaert, the statistical-empirical method has better calibration results.

While ensuring that the transit time of the Landsat 8 satellite closely aligns with the observation time of the FY4A satellite, we selected areas with cloud cover of less than 10% to create the dataset. Subsequently, the preprocessed Landsat 8 data were employed to generate a binary snow map at a 30-m resolution using the SNOWMAP algorithm [16]. Zhang et al. [28] verified with 201 Landsat8 images of the TP that the optimal Landsat8 NDSI threshold at the regional scale of the TP is about 0.4 for snow cover >50%, which is also not much different from the classical NDSI threshold. Considering that the snow coverage in the snowy region of the Tibetan Plateau is mostly greater than 50%, the NDSI threshold of 0.4 is selected to generate a binary snow map, and the interference of water bodies on snow identification is eliminated by determining whether the reflectivity in the fifth band is greater than 0.11 [30]. Finally, the binary snow map was aggregated into a 2000-m resolution FSC map by the definition method, which was calculated as shown in the following:

$$\mathrm{FSC} = \frac{N_S}{N_t} \times 100\% \tag{2}$$

where N_S denotes the number of snow-covered pixels in each grid and N_t denotes the total number of pixels included in each grid.

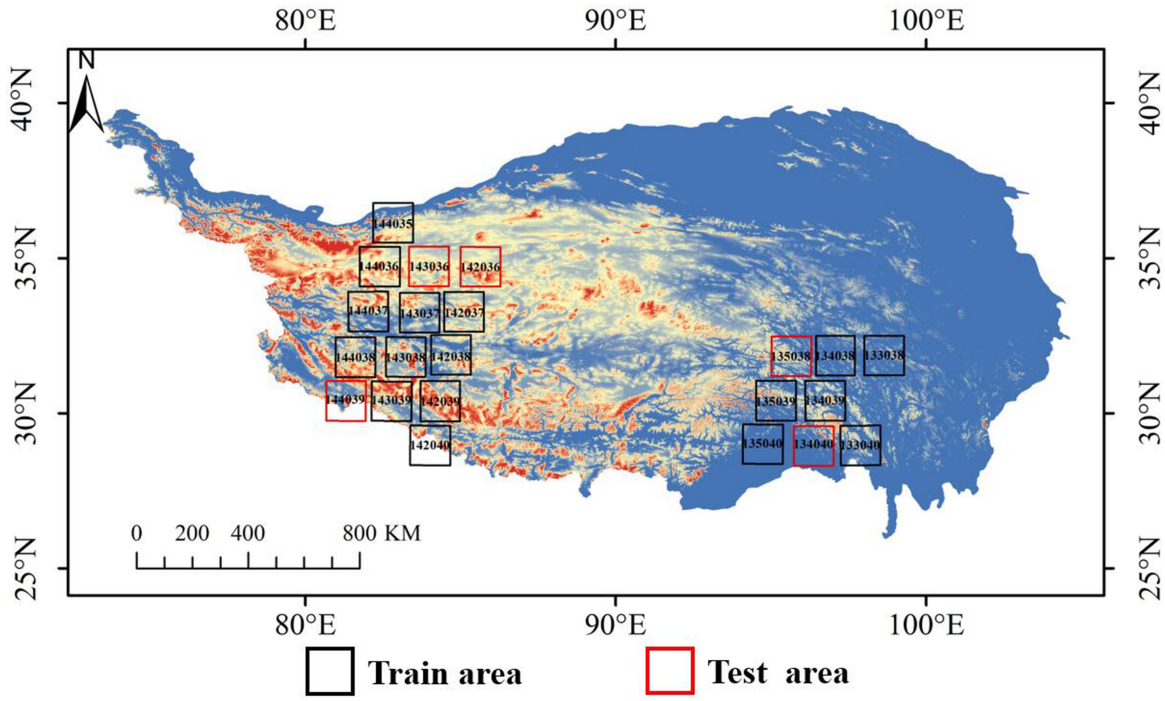


Fig. 1. Map of the study area.

The original FY4A/AGRI data underwent several steps, including radiometric calibration, geometric correction, and vector clipping, to obtain TP region images. Radiometric calibration is the conversion of satellite scanning radiometer output counts into reflectance physical quantities. The pixel values of the image data layer "NOMChannelX" (X is the channel number) are first read from the officially downloaded raw data, and then the data is calibrated according to the calibration table "CALChannelX" (X is the channel number) to convert the raw digital signal into reflectance. Geometric correction is to eliminate or correct geometric errors in remote sensing images. First of all, the region of interest is established in the original data and the official longitude and latitude lookup table, and then according to the conversion relationship given by the China Satellite Meteorological Center, the data row number and longitude and latitude are converted to each other to obtain the latitude and longitude information. Subsequently, DEM data were integrated as the eighth channel data and merged with the FY4A/AGRI data. Finally, the FY4A/AGRI data with DEM channels were cropped to a size of 64×64 based on the FSC maps extracted from Landsat 8 OLI data. To ensure the balance of the training samples, the images from the west to the east of the TP, which included the gradual transition from the snow-free region to the high snow-covered region, were selected as the dataset. The selected areas of the training and test sets are shown in Fig. 1.

III. Methodology

UNet [33] is a classical network model based on an encoder–decoder structure, which has been successfully applied to glacier identification in remote sensing images [34], [35]. Its unique U-shaped structure makes it accurate when trained using a small number of samples, making it particularly suitable for use in this FSC retrieval study. Therefore, this study proposes a deep learning-based FSC retrieval method; the overall flow is shown in Fig. 2. Raw data were processed and fed into the model for training. Subsequently, model evaluation was performed to select the best model. Finally, validation was conducted using MOD10A1 snow products. The fractional snow cover U-shape net (FSC-USNet) method proposed in this study mainly consists of the spatial and channel feature extraction module (SCFE module), decoding module, improved attention module (ICA module) and channel refinement module (CRF module); the overall network structure is shown in Fig. 3. The encoding module mines the relationships between various channels and spatial information from the input image and characterizes the shallow features into high-dimensional feature images. The decoding module gradually inverts the sublinear relationship between each channel and FSC from the deep features and gradually recovers the spatial information. The ICA module associates the encoding and decoding modules, allocating different weight information to different pixels and channels to enhance the distinctiveness of various FSC ranges. Finally, the final FSC retrieval result was obtained by gradually refining the number of output channels through the output header with residuals.

A. Spatial-Channel Feature Extraction Module

In previous research, substantial FSC retrieval was accomplished by linearly combining different spectral bands of multispectral images. The relationships existing between different bands are crucial for FSC extent retrieval. Therefore, this study

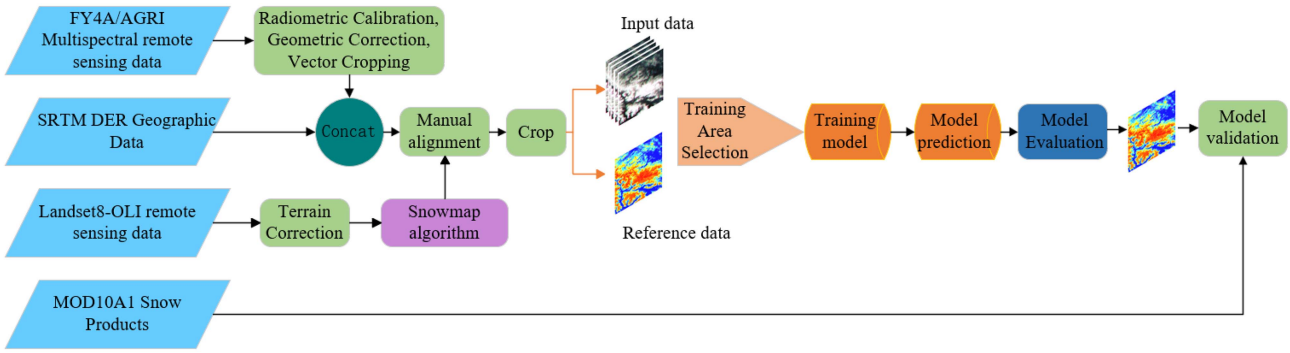


Fig. 2. Overall flowchart of the study.

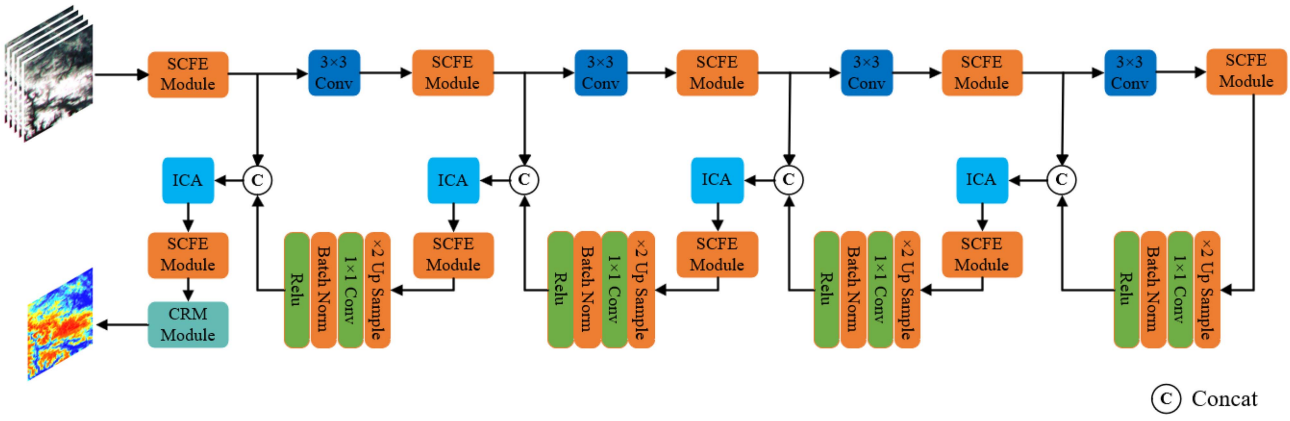


Fig. 3. Structure of the FSC-USNet network.

incorporates a channel feature extraction branch composed of 1×1 convolutional kernels into the feature extraction module to explore the relationships between various channels in the input image. To obtain finer snow boundary information and spatial information, we add a spatial feature extraction branch composed of 3×3 convolutional kernels. To improve the generalization ability of the model to prevent overfitting, we added dropout for random deactivation of the convolution kernel after the convolution kernel in the above two branches. After mapping the activation function, the spatial features are concatenated in the direction of the channels, and the spatial features are fused with the 1×1 convolutional kernel to reduce the number of channels by half to avoid redundancy. To solve the underestimation problem in the UNet FSC retrieval process, we introduce the classical residual structure. In the residual branch, the number of channels is adjusted by a 1×1 convolution kernel, which is then summed with the fused main branch feature map to obtain the final spatial and channel features. The overall module structure is shown in Fig. 4. It is worth noting that the network architecture proposed here employs 3×3 convolutions as the downsampling method between each SCFE module. The traditional UNet model is designed for image segmentation, which rarely needs to account for the numerical differences in the neighboring pixels; thus, using max-pooling as a downsampling method in the image segmentation task has

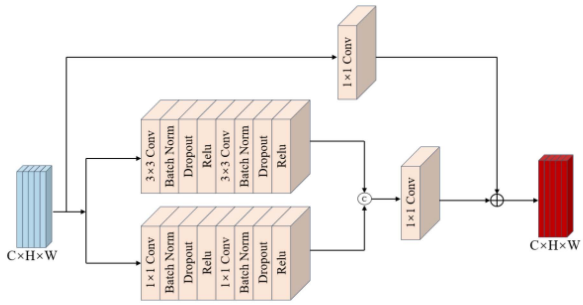


Fig. 4. Structure of the SCFE module. Extract and encode snow information from input images.

little effect on the overall results. However, FSC is a continuous value, and due to resolution constraints, significant numerical differences can occur between neighboring pixels. Using 3×3 convolutional kernels not only allows for thoroughly extracting spatial features but also facilitates further adjusting relationships between different channels.

B. Improved Coordination Attention Module

Due to limited resolution, the input image size is only 64×64; thus, the dataset is unsuitable for very deep networks. Simply

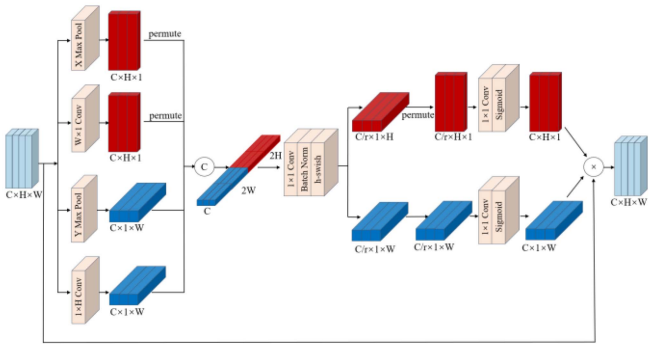


Fig. 5. Structure of the ICA module. Deep fusion of coding features and decoding features after skip connection.

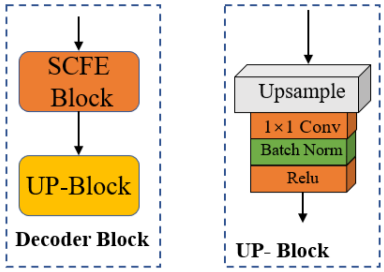


Fig. 6. Structure of the decoding module. Feature decoding of encoded features and gradual recovery to original resolution.

concatenating the shallow features extracted by the encoder and the deep features restored by the decoder makes it difficult to distinguish highly similar features and perfectly integrate two-stage features. To further enhance retrieval performance, an improved coordinated attention mechanism is added after fusing encoded and decoded features. The coordinated attention mechanism [36] is an attention mechanism that simultaneously accounts for spatial and channel dimensions and considers position-related information in the spatial direction. The coordinated attention mechanism assigns different weights to various channels in the channel dimension to enhance the discrimination between snow-cover and snow-free pixels, which is an outstanding contribution to snow-pixel recognition. Spatially, the classical coordinated attention mechanism encodes the input features using average pooling along the height and width directions, which weakens feature information in regions with highly fragmented snow coverage due to the averaging effect. Additionally, because of the high similarity between cloud and snow classes, average pooling can lead to combining feature information, making it ineffective in distinguishing highly similar categories. Therefore, in this study, an enhancement to the coordinated attention mechanism is implemented, replacing avg-pooling with max-pooling to prevent feature loss in high FSC extent areas and enhance sensitivity to texture features in the feature maps. Additionally, H×1 and 1×W convolutional branches are added to learn feature map texture features to capture remote spatial interactions with precise location information and to learn texture features for each snow-covered area. Then, the attention vectors in the horizontal and vertical directions are obtained by F_1 variation. Finally, the attention vectors are multiplied by the input features to obtain the final feature map with weights. The ICA module is shown in Fig. 5.

For a given feature map $X \in R^{C\times H\times W}$, with maximum pooling along the height and width directions using the $1, W$ and $(H, 1)$ pooling kernels, its output at the Cth channel at height h can be expressed as follows:

$$Z_c^{h\_\max} = \max_{0\le i\le w} \left[x_c\left(h,i\right)\right] \tag{3}$$

$$Z_c^{h\_\text{conv}} = \text{Conv}_{1\times w} \left[x_c\left(h,i\right)\right]. \tag{4}$$

Similarly, the output of its C channel at width w can be expressed as follows:

$$Z_c^{2\_\max} = \max_{0\le i\le h} \left[x_c\left(w,i\right)\right] \tag{5}$$

$$Z_c^{w\_\text{conv}} = \text{Conv}_{h\times 1} \left[x_c\left(w,i\right)\right]. \tag{6}$$

Then, the output tensor is spliced along the channel direction and F_1 is transformed by a shared 1×1 convolution to obtain the coded intermediate feature mapping of spatial information in the horizontal and vertical directions $f \in R^{\frac{c}{r}\times(H+W)}$, where r is the downsampling multiplicity, which is computed as shown in the following:

$$f = \delta\left(F_1\left(\left[Z^{h\_\max}, Z^{h_{\text{conv}}}, Z^{w\_\max}, Z^{w\_\text{conv}}\right]\right)\right) \tag{7}$$

where δ is the nonlinear activation function. Immediately after that, f is sliced into two separate tensors $f^h \in R^{\frac{c}{r}\times H}$ and $f^w \in R^{\frac{c}{r}\times W}$ along the spatial dimension. Then, f^h and f^w are transformed into tensors with the same number of channels as the input X by two 1×1 convolutional transforms, F_h and F_w. g^h and g^w are obtained by the sigmoid activation function. The equations are shown in the following:

$$g^h = \text{Sigmoid}\left[F_h\left(f_h\right)\right] \tag{8}$$

$$g^w = \text{Sigmoid}\left[F_w\left(f_w\right)\right]. \tag{9}$$

Finally, the attention weight matrix is obtained by tensor multiplication; the output is represented as follows:

$$y_c\ (i,j) = x_c\ (i,j) \times g_c^h\ (i) \times g_c^w\ (j). \tag{10}$$

C. Decoding Module

The decoding module, shown in Fig. 6, consists of two parts: the SCFE module and the upsampling module. After the last level of feature extraction is completed, deep feature decoding begins. First, the features are upsampled by a factor of 2 to recover the spatial texture features. Then, a 1×1 convolution is applied to mine the relationship between the channels and halve the number of channels. Finally, the nonlinear mapping is completed by batch normalization and an activation function. It is undeniable that upsampling improves the spatial dimension of feature information, but it does not bring additional information in essence. Therefore, in this study, coded features are aggregated for feature information filling through the ICA module, while effective features are extracted through a spatial-channel feature extraction module and redundant features are eliminated.

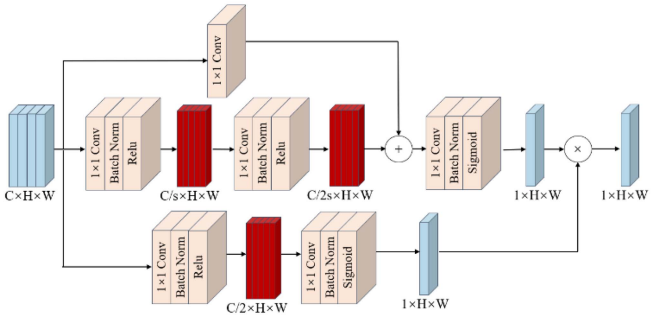


Fig. 7. Structure of the CRF module. Suppresses non-snow pixels and outputs FSC maps.

D. Channel Refinement Module

After going through the final decoding module, deep features are restored to 32 channels with rich texture information. However, experimental results have shown that directly mapping the 32 channels to the output through fully connected layers leads to less-than-ideal retrieval results. This issue is reflected in underestimated FSC in the fragmented snow area, and blurred texture features in areas with high FSC. Therefore, this study adds the CRF module, shown in Fig. 7, to gradually refine the output channels and achieve the suppression of nonsnow image elements before outputting the results. First, the number of channels is gradually reduced twice in the main branch by a 1×1 convolution kernel while recombining all channel information. In the residual branch, a 1×1 convolution kernel is used to reduce the number of channels in line with the main branch. Then, the results of the two branches are added together for the superposition of information to supplement the possible information loss in the main branch. Finally, we replace the fully connected layers with 1×1 convolutions, followed by batch normalization and a sigmoid activation function to map the output to the [0, 1] range. The advantage is that the fully connected layer destroys the spatial structure of the image, whereas the convolutional layer does not. The computational process of the 1×1 convolution is equivalent to the fully connected layer [37]. In the snow accumulation detection branch, we convolve the decoded features with rich texture information twice to reduce the number of channels and complete the nonlinear mapping to obtain the classification results of snow and nonsnow pixels, where we use 0 to denote nonsnow pixels. Finally, it is multiplied with the FSC map obtained from the main branch to achieve the suppression of nonsnow pixels.

IV. Experiments

A. Experimental Setup

The deep learning framework used in the proposed method is PyTorch 1.12.0, the GPU acceleration tool is CUDA 11.3, and the programming language is Python 3.8. The hardware configuration is an i7-12700H CPU and an NVIDIA GeForce RTX 3060 GPU. The dataset is 300 TIF images, the size is 64×64, and the training set and test set are divided according to 8:2. When training the network, the batch size was set to 8. Since the experiment almost converged after 150 iterations, the number of iterations was set to 200, the initial learning rate was set to 0.01, the decay coefficient was set to 0.1, and the learning rate was updated every 50 iterations using the Adam optimizer.

B. Evaluation Indicators

FSC is an indication of the degree of FSC within a given image element and is a continuous value belonging to the range between [0, 1]. In this study, the mean squared error (MSE) is employed as the loss function, and the evaluation metrics include the R2_score (R^2), root mean square error (RMSE), correlation coefficient, kappa coefficient (kappa), and explained variance score (EVS). For the FSC retrieval model, R^2 reflects the proportion of the total variation in the dependent variable that can be explained by the variable through the regression relationship. EVS characterizes the degree of the model's explanation of the variation in FSC, and the closer R^2 and EVS are to 1, the better it indicates that the model can predict the variation in FSC well. RMSE measures the average difference between the model's predicted coverage and the true coverage, with a lower RMSE indicating a more accurate model prediction. The correlation coefficient reflects the strength and direction of the linear relationship between true and model-predicted FSC. The kappa coefficient evaluates FSC retrieval accuracy and is commonly used in discrete classification tasks. In this study, FSC is divided into discrete data to calculate the kappa coefficient. The formula is shown as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{y_i})^2 \tag{11}$$

$$R^2 = 1-\frac{\sum_{i=1}^{n}\left(y^i-\widehat{y_i}\right)^2}{\sum_{i=1}^{n}\left(y_i-\overline{y_i}\right)^2} \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{y_i})^2} \tag{13}$$

$$P = \frac{\sum_{i=1}^{n}\left(y_i-\overline{y_i}\right)\left(\widehat{y_i}-\overline{\widehat{y_i}}\right)}{(n-1)\times\sigma_{y_i}\times\sigma_{\widehat{y_i}}} \tag{14}$$

$$\mathrm{Kappa} = \frac{p_o-p_e}{1-p_e} \tag{15}$$

$$\mathrm{EVS} = 1-\frac{\sigma^2\left(y_i-\widehat{y_i}\right)}{\sigma^2\left(\widehat{y_i}\right)} \tag{16}$$

where y_i denotes the true value, $\widehat{y_i}$ denotes the predicted value, $\overline{y_i}$ denotes the mean of the true value, $\overline{\widehat{y_i}}$ denotes the mean of the predicted value, and n denotes the sample size, p_o denotes the observational accuracy, which is the amount by which the predicted values match the number of classifications of the true value, p_e denotes the stochastic accuracy, which is the proportion of the model calculated according to the probability of stochastic classifications that match the actual category, σ denotes the standard deviation, and σ^2 denotes the variance.

TABLE II
COMPARISON TABLE OF DIFFERENT INPUT INVERSION METRICS

Method	Input	R^2	RMSE	Kappa	Correlation	EVS
Unet	Band 1-7	0.5449	0.1301	0.3641	0.7672	0.5756
	Band 1,4,5,6,7	0.5112	0.1324	0.3604	0.7584	0.5515

TABLE III
COMPARISON OF RETRIEVAL METRICS FOR EACH METHOD

Method	Input	R^2	RMSE	Kappa	Correlation	EVS
BPANN	FY4A	0.6157	0.1296	0.3635	0.7787	0.6301
	FY4A+DEM	0.6557	0.1254	0.3741	0.7987	0.6601
UNet	FY4A	0.5449	0.1301	0.3641	0.7672	0.5756
	FY4A+DEM	0.5786	0.1267	0.3702	0.7791	0.5941
UNet++	FY4A	0.6222	0.1258	0.3698	0.7921	0.6354
	FY4A+DEM	0.6431	0.1211	0.3789	0.8095	0.6638
Resnet_FSC	FY4A	0.5550	0.1319	0.3632	0.7798	0.5603
	FY4A+DEM	0.5778	0.1286	0.3762	0.7853	0.5815
RF	FY4A	0.6227	0.1258	0.3785	0.7893	0.6444
	FY4A+DEM	0.6559	0.1247	0.3892	0.7912	0.6714
SMA	FY4A	0.5297	0.1565	0.3327	0.7135	0.5294
Our	FY4A	0.6871	0.1241	0.3883	0.7912	0.7114
	FY4A+DEM	**0.7182**	**0.1100**	**0.4060**	**0.8178**	**0.7332**

The bold values indicate optimal indicator data.

C. Prelaboratory

The FY4A/AGRI remote sensing data provide reflectance data in seven bands, of which NDSI can be obtained by the operation of band 1 with band 5, so it is reasonable to assume that they are helpful for the retrieval of snow cover. As can be seen in Table II, the data in the other bands help to avoid interference from feature signals such as clouds and bodies of water, and only band 2 and band 3 are targeted at the role of vegetation. To further verify the contribution of band 2 and band 3 to the retrieval of snow parameters, we conducted experiments with all bands and data excluding band 2 and third band 3 as inputs to the base model Unet, respectively. The experimental results are shown in Table II, and the model trained with all bands as inputs shows better performance in terms of various metrics, with the largest fluctuations in R^2, EVS, and the smallest increase in Kappa. In addition, since the texture characteristics of snow on remote sensing images are related to the nature of the subsurface [38], adding the bands with vegetation information to the input data is also useful for improving the inversion accuracy of snow cover. Therefore, bands 1 to 7 are used as input data in the subsequent experiments in this article.

V. RESULTS

A. Comparison of Experimental Results

To validate and compare the performance of the FSC retrieval method proposed in this study, it was tested against several classical retrieval methods on the test dataset. The comparative methods include BPANN, random forest, UNet [33], UNet++ [39], ResNet_FSC [25], and spectral mixture analysis (SMA) [41] methods. The comparison of the retrieval indices of each method on the test set is given in Table III, in which the average value of R^2 of the proposed method on the test set is 0.7182, the average value of RMSE is 0.1100, the average value of kappa is 0.4060, the average value of correlation is 0.8178, and the average value of EVS is 0.7332. Table II indicates that SMA methods can obtain relatively accurate results within snow-rich image elements. However, there are large errors and even a large number of missed detections within the image elements with small snow areas and weak signals, which is the main source of errors in the spectral mixture analysis method. The BPANN method focuses on the relationship between each channel of the input image, and can achieve relatively accurate point-to-point FSC retrieval, with R^2 and EVS reaching 0.6557 and 0.6601, respectively. However, it is more sensitive to interference in snow-free areas, and there is the phenomenon of recognizing snow-free areas as low snow-covered areas, which is the main reason for its relatively high RMSE (0.1254). In addition, since BPANN mines features in a pointwise manner, the retrieval results have an overall discontinuous scatter distribution.

UNet, UNet++ adopt convolution as the way to extract features, which on the one hand expands the sensory field during feature extraction and takes into account the relationship between neighboring pixels. On the other hand, the relationship of each channel is taken into account, so the robustness is improved compared to the BPANN method. Compared with UNet++, UNet only uses a simple concat process when skip connection, which results in less in-depth fusion of encoded and decoded features. This directly leads to the underestimation of FSC by UNet in some areas, which is the main reason for the relatively large RMSE of 0.1267 for the UNet. In contrast, UNet++ performs fusion of features by introducing a dense jump connection mechanism. This dense jump connection helps to enhance the perception of information at different scales and levels, making the spatial distribution of snow closer to the real situation, with R^2 and EVS reaching 0.6431 and 0.6638, respectively. Resnet_FSC uses Resnet as the backbone network for feature extraction, and the feature extraction capability is greatly enhanced. However, due to the network is too deep, the loss of texture detail information is severe and excessive attention is paid to the areas with high snow coverage, resulting in the results presenting patchy features with lack of texture, which is the main reason why the Resnet_FSC error (RMSE) is as high as 0.1286.

Random forest, in contrast, maintains high efficiency even with relatively high-dimensional inputs in this study. It benefits from its random sampling approach for individual pixels, which results in strong generalization capabilities and excellent performance. With an R^2 score as high as 0.6599, it is second only to the proposed model. It also achieves favorable metrics in terms of EVS (0.6197). The proposed model focuses on extracting more comprehensive feature maps by leveraging the SCFE module, which emphasizes correlations between adjacent pixels and channels, dropout randomized disable convolution

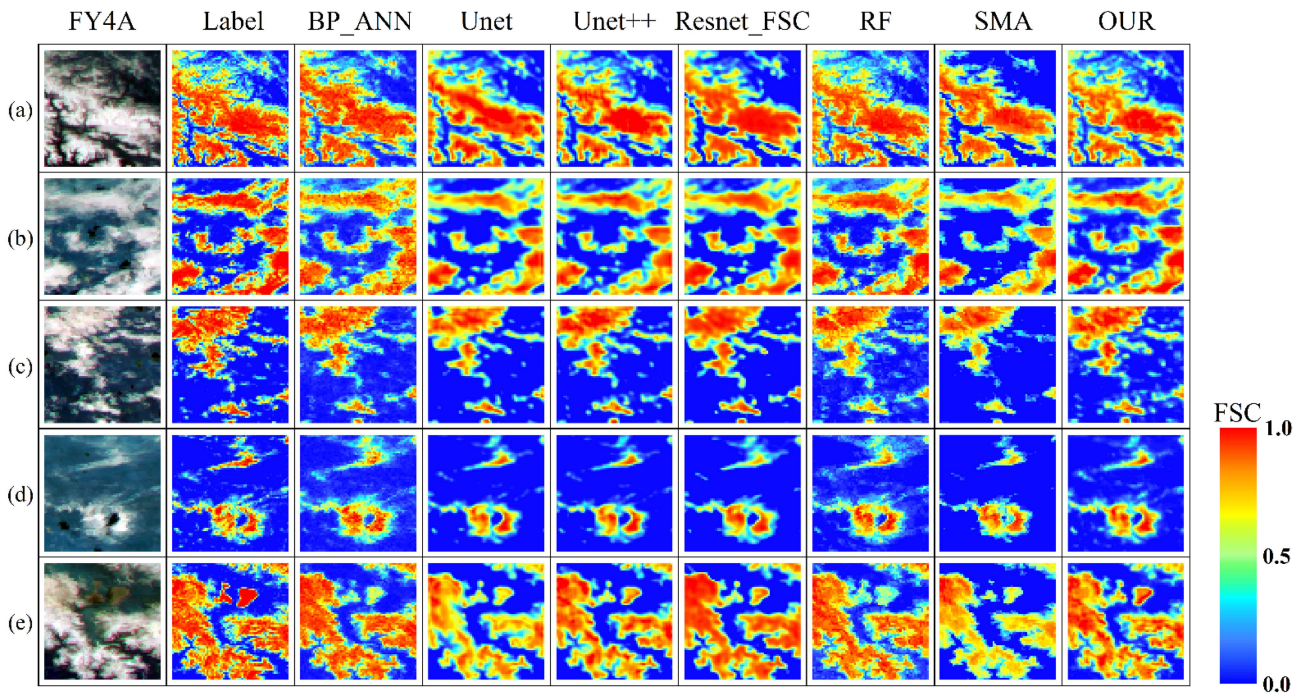


Fig. 8. Comparison of retrieval results of each method. FY4A denotes FY4A true color image.

kernel essentially achieves the same effect as random forests. Additionally, it employs convolution as a downsampling method to alleviate the spatial information loss caused by max-pooling. The introduced ICA mechanism aggregates features from different stages, supplying the decoder with weighted supplementary information. Ultimately, the relationship between the individual channels of the decoded feature map is mined by the CRF module and the non-snow pixels are suppressed. Thus, the retrieval results have a small RMSE of only 0.1100 deviations from the true label, and the kappa and correlation coefficients are also 0.4060 and0.8152, respectively.

In addition, from Table III, it can be seen that the addition of DEM data makes an outstanding contribution to improving FSC retrieval accuracy. Notably, the proposed method achieved an approximately 4.5% increase in R^2 and a 4.6% increase in the kappa coefficient (kappa) after incorporating elevation data. Other evaluation metrics also exhibited varying degrees of improvement.

The FSC retrieval results obtained by different methods are depicted in Fig. 8. Fig. 8(b) shows that the BPANN retrieval results exhibit weaker spatial coherence. The transition zones between snowy and snow-free regions appear significantly elongated, and the method demonstrates higher sensitivity to disturbances in snow-free areas, resulting in a substantial number of misclassifications of snow-free areas with FSC percentages below 10%. In the region where FSC is higher than 70%, the BPANN inversion results show an underestimation and this underestimation is distributed in a point-like manner. In the convolution-based approach, the excessive depth of the ResNet_FSC network diminishes the distinctions between high snow-cover areas and loses information about low FSC and fragmented snow-covered regions. This directly results in excessive smoothing of the FSC results in areas with over 70% coverage, with small differences in neighboring pixels. It also leads to the common issue of misclassifying regions with FSC percentages below 20% and fragmented snow areas as background. These factors are the primary reasons for the high RMSE observed in the retrieval results of ResNet_FSC, as illustrated in Fig. 8. The proposed method fuses the shallow features encompassing rich edge information and the upsampled features encompassing high-dimensional information through the ICA module to strengthen the feature differences between background and different FSC types and to enhance the attention on snowy elements, which realizes the fine-loading judgment of fragmented and low FSC image elements. In comparison, the UNet network structure has a shallower depth, preserving shallow structural features such as boundaries while also exploring more high-dimensional features. However, limited by its oversimplified skip connection approach, the inversion results not only lose information about the fragmented snow region but also show a huge underestimation problem in the high FSC region. UNet++ is an upgraded version of UNet, which mitigates to some extent the underestimation of UNet in areas where the FSC is higher than 70%, but still has some errors.

To address this underestimation issue, the proposed method employs a dual-pronged approach. First, it enhances feature extraction and diminishes information loss through the SCFE module. Second, it gradually reduces the number of channels using the CRM module and supplements any potential information loss within the main branch through the residual branch. Fig. 8 shows that the proposed method is effective in alleviating the underestimation problem in areas with FSC higher than 70% and is rich in hierarchical and textural information. In transition areas with approximately 20% FSC between high-snow-covered

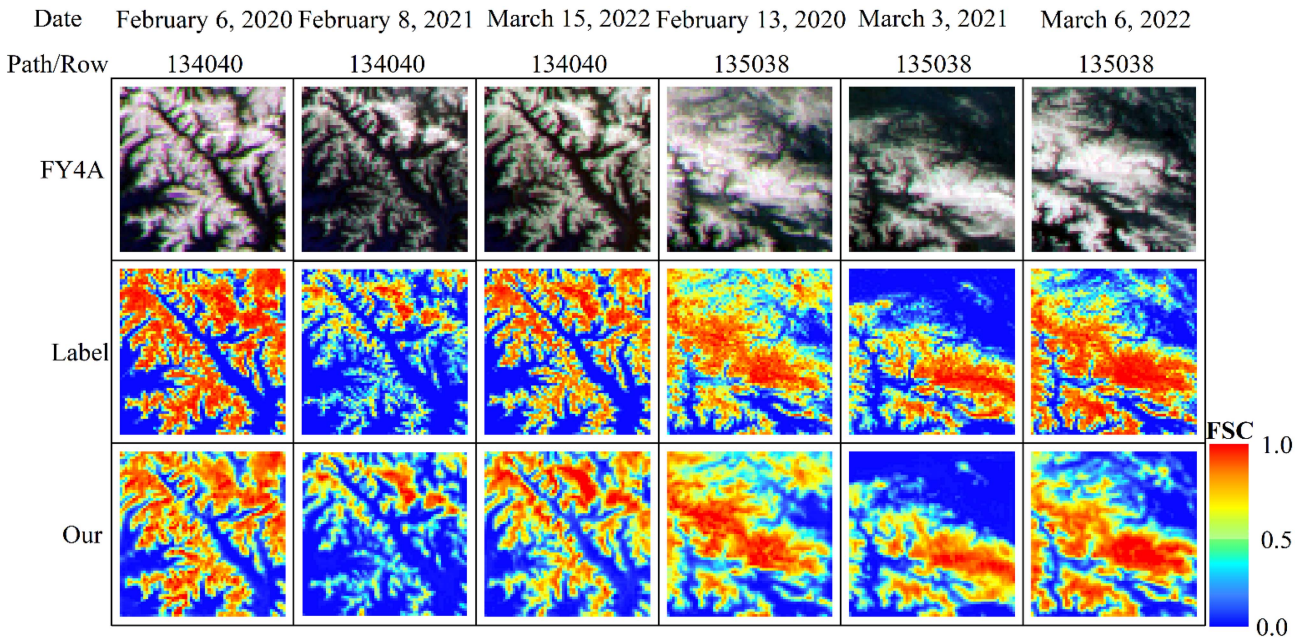


Fig. 9. Results of FSC retrieval based on the proposed method at different times in the same region.

regions, the proposed method demonstrates accurate classification without producing adhesion or excessive segmentation.

In addition, the convolution-based method is more effective in segmenting the edges of the region with and without snow, but there may be a phenomenon that such overly clear edges are enlarged. Comparing the retrieval results of the different methods in Fig. 8(a) and (e), it can be seen that for such narrow boundaries UNet++, Resnet_FSC, and other methods detected significantly more snow-free regions than labels. The BP-ANN and RF methods produce different degrees of misjudgment at the boundary, resulting in the blurring of the boundary, which is because it is difficult to shield the interference of the background information only by the characteristics of a single image element. The proposed method, on the other hand, not only combines the relationship between the channel and the surrounding pixels, but also suppresses the nonsnow pixels in the CRF module, which achieves a relative balance between this boundary blurring and expansion.

B. Extensibility Validation

To further verify the generalization performance of the method in this article, we conducted two sets of comparisons of FSCs in the same region at different times, and the results are shown in Fig. 9. From the figure, it can be seen that for the same area of the spatial distribution of snow even at different times show obvious similarity, but the specific FSC does not show a certain pattern, only closely related to the environment at each time. The proposed method is not greatly affected by this, not only the FSC retrieval results are consistent with the label, but also presents obvious spatial texture information in the mountainous areas with complex terrain. This proves that the proposed method possesses the ability to be extended to other times or regions.

TABLE IV
TABLE OF ABLATION EXPERIMENT RESULTS

Number	SCFE	Improved CA	CRM	R^2	RMSE	Kappa	Correlation	EVS
1	×	×	×	0.5786	0.1267	0.3702	0.7791	0.5941
2	✓	×	×	0.6397	0.1219	0.3928	0.7818	0.6497
3	✓	✓	×	0.6479	0.1211	0.3993	0.7827	0.6609
4	✓	×	✓	0.6953	0.1174	0.3969	0.7925	0.7174
5	✓	✓	✓	**0.7182**	**0.1100**	**0.4060**	**0.8178**	**0.7332**

The bold values indicate optimal indicator data.

C. Ablation Experiments

To validate the effectiveness of the proposed modules, we conducted a series of ablation experiments to reveal the impact of each module in the FSC retrieval process. First, we replaced the feature extraction module with the SCFE module for the experiments. Table IV shows that the R^2 is greatly improved to 0.6397 after replacing the SCFE module. The kappa and the EVS also improved by more than 5%. This suggests that compared to the original model, the SCFE module effectively mitigates the underestimation problem in regions with high FSC. This is because whether the encoder sufficiently extracts the features from the input images fundamentally determines the retrieval performance. The SCFE module accounts for the relationships between space and channels and introduces a residual branch to alleviate the underestimation problem. As a result, the retrieval accuracy is significantly improved. Subsequently, we added the ICA module, and we can see from the various metrics that the ICA module has slightly improved RMSE, but both R^2 and EVS

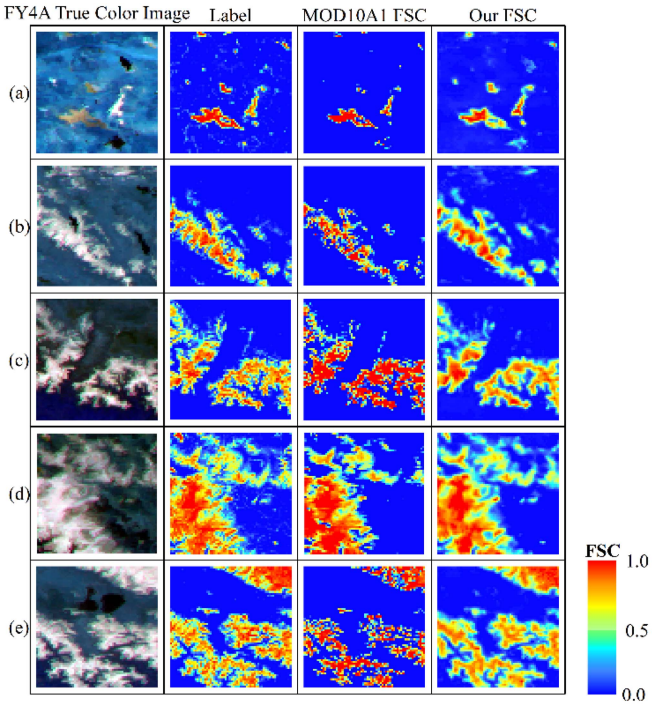


Fig. 10. Comparative validation plot of retrieval results.

have gained a modest improvement. However, the primary role of the ICA module lies not only in the modest optimization of the retrieval results but also in the comprehensive aggregation of shallow-level features with upsampled features, thereby selectively supplementing information loss in the decoding stage. As a result, after adding the CRM module to gradually refine the decoding features, the R^2 obtains an enormous improvement of 0.0703, and the RMSE and EVS are also optimized by 0.0111 and 0.0732, respectively.

To verify that the proposed ICA module is more effective than the classical coordination attention mechanism, we specifically set up a fourth set of experiments in Table III. In the fourth set of ablation experiments, we added the classical coordination attention mechanism, and in comparison, the proposed ICA module gained a 0.0229 improvement in R^2 and achieved better results in other metrics by a narrow margin.

D. Snow Product Comparison

The MOD10A1snow product is internationally recognized as the predominant snow product. In this study, the retrieval results of the proposed method are compared with the MOD10A1 product to validate the results and calculate the indicators, as shown in Fig. 10 and Table V. Overall, in comparison to the MOD10A1 snow product, the proposed method demonstrates superior retrieval capabilities in predicting FSC. Both R^2 and RMSE exhibit significant improvements, but there is still a certain degree of underestimation problem for some high snow cover regions. From Fig. 10, it can be seen that Liang's [1] proposed equation for fitting the snow cover of MOD10A1 on the TP is in high agreement with the snow cover computed from

TABLE V
COMPARISON OF RETRIEVAL METRICS BY METHOD

Method	R^2	RMSE	Kappa	Correlation	EVS
Our	0.6743	0.1395	0.3619	0.7566	0.6894
MOD10A1	0.5877	0.2021	0.3358	0.6916	0.5727

Landsat 8 high-resolution image, with an average R^2 and correlation coefficient of 0.5877 and 0.6916, respectively. However, the phenomenon of overestimation in mountainous areas with high snow cover and underestimation in small patchy areas [13], [20], [41] still exists. On one hand, a larger solar angle leads to an overestimation of snow cover in MOD10A1 [42], and other environmental factors are also an important contributor to the overestimation of FSC. The retrieval results of the proposed method in this paper are closer to the true value in the region with snow cover higher than 80%, and the error is smaller relative to the overestimation of MOD10A. The shortcoming is that the underestimation occurs in the edge area of the high snow cover region in Fig. 10. On the other hand, MOD10A1 is more susceptible to fragmented snow [42]. This may be highly correlated with the resolution of the MOD10A1, where snow cover extracted through the MOD10A1 has been missed or underestimated in scattered small, low snow cover areas. The snow cover extracted by high-resolution satellites accurately extracts these weak signals, and the model trained with this as truth value also understands this information in depth, so that the proposed method in this article has a better detection ability for small, low snow cover areas. In addition, the spatial texture information of snow cover extracted by MOD10A1 is weak, and in the vast majority of cases, it is scattered, which is particularly obvious in Fig. 10(e). In contrast, the snow cover inverted by the method in this article is overall coherent and rich in texture information. From the graph, it is evident that the simple linear regression method employed by MOD10A1 results in a substantial overestimation exceeding 5% in mountainous areas with FSC exceeding 70%, while underestimation occurs in patchy snow-covered regions. This is not only reflected in Fig. 9 but also highly consistent with previous studies [13], [20], [41]. Furthermore, the MOD10A1 product shows a cliff-like drop in FSC at the junction of snowy and snow-free areas. In other words, the MOD10A1 product is insensitive to the junction of snowy and low FSC areas, and there are instances of FSC misclassification in Fig. 9. The reason for this could be attributed to the fact that the MOD10A1 snow product only provides daily snow data, while on the TP, snow accumulation and melting rates are rapid, especially in areas with low FSC. Therefore, the MOD10A1 snow product cannot accurately observe low FSC areas. In this study, we utilize FY4A data with a temporal resolution as fine as 15 min as the data source, which effectively meets the requirements of real-time observations and allows for a clear observation of temporal changes in FSC. In addition, the proposed method fully considers the characteristics of the neighboring pixels, so it has certain texture characteristics. A clear transition zone is also present in the junction area between

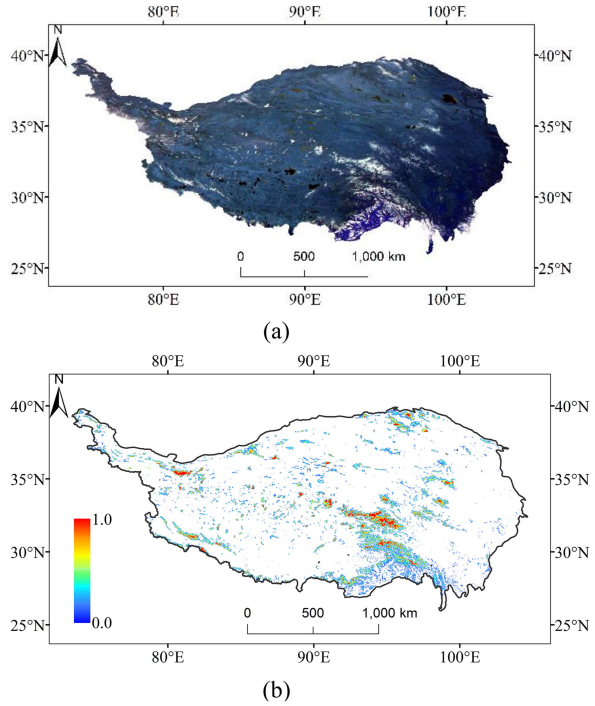


Fig. 11. Imagery of the TP region on 23 December, 2022. (a) Is FY4A/AGRI true color imagery. (b) Is the map of declouded snow cover by the proposed method.

snow-covered and snow-free regions, which is more in line with the actual situation. Table IV shows that the proposed method achieves better results than the MOD10A1 product, both in the area with abundant FSC and in the area with relatively fragmented FSC.

FSC mapping was performed for the TP region using the method proposed in this study. We selected the FY4A/AGRI image at 4:00 on 23 December, 2022, which has less cloud cover, as the input data. The high temporal resolution of FY4A was utilized to filter out the cloud cover in the northwestern and southeastern parts of the TP. Specifically, we preprocessed five images around 04:00 on 23 December, 2022, using the same vector crop to produce the FY4A remote sensing image of the TP region. Then the clouds are identified based on cloud movement and snow immobility and filled with data from the proximity time, resulting in a multitemporal phase filtered cloud effect. The final map of snow cover on the TP was produced, as illustrated in Fig. 11. From Fig. 11(b), it can be seen that the method proposed in this article is sensitive to the fragmented and low snow-covered areas, and the image elements are arranged continuously and compactly with rich texture information. Combined with the high temporal resolution FY4A remote sensing data, it can comprehensively and accurately complete the inversion of snow cover.

VI. Uncertainties and Limitations

In this article, snow cover on the TP is estimated based on FY4A remote sensing data by deep learning method, which improves the accuracy of snow cover retrieval and faces some problems at the same time. The convolution-based feature extraction method comprehensively takes into account the long-term association of image elements in both channel and spatial dimensions. However, it will weaken the feature differences of neighboring image elements to a certain extent, especially at high frequencies such as at the edge of snow, which will cause information smoothing, and this is the main reason for the unclear boundary between snow-covered and snow-free areas. Therefore, it is necessary to construct a more effective feature extraction method in feature extraction, so that it can better capture the subtle differences between neighboring image elements and also extract rich spatial information.

In addition, the reference FSC used in this study was derived from Landsat8 according to the SNOWMAP algorithm, and the choice of the NDSI threshold has a decisive role in the selection of the snowpack image elements. However, study [28] showed that the optimal NDSI threshold generally increases with slope and decreases as the aspect varies from southeast to northwest in the TP. Harer [3] has also found through his research that in some cases, fixing NDSI thresholds to estimate snow cover introduces some error, and he suggests that seasonal variations in NDSI thresholds be considered. Therefore, the fixed use of the optimal NDSI threshold for the TP region in this study is subject to errors due to the spatial and temporal distribution of the dataset. On the other hand, while controlling for the Landsat8 transit time versus the FY4A/AGRI shot time, it is still possible for shallow snow to melt or ablate within this time difference. Because of the relatively small snow depths on the TP [44], it is still possible for latent snow to melt or sublimate within this time difference, which creates some error in the estimation of FSC in the latent snow region.

Finally, Zhang et al. [28] found that FSC has significant accuracy differences between different land cover types, and this study did not fully consider such differences on different land types when preparing the truth labels and training the model. Many studies [9], [24] have found that FSC in forested areas tends to be significantly underestimated due to the forest canopy and the accuracy is significantly lower than that in basins and grasslands [28]. Such a problem can be alleviated by considering view angle effects [45] and multi-index technique [46], and a separate study can be carried out for forested areas to improve the accuracy of FSC.

VII. Conclusion

In this study, a deep learning-based FSC retrieval method is proposed for FSC detection in the TP region. The network architecture follows the classical encoder–decoder structure, with the addition of the SCFE module and the decoding module for extracting spatial features and channel relationships. Furthermore, the ICA module is incorporated to fuse shallow features with upsampled features, providing the decoder with comprehensive and detailed abstract information. Combined with the channel refinement module, a high temporal resolution FSC map based on FY4A/AGRI data was generated.

The experimental results show that compared with the traditional FSC retrieval method, the proposed method can obtain

highly accurate FSC retrieval results. The SCFE module and the decoding module comprehensively provide the spatial texture structure of the FSC map. The ICA mechanism assigns higher weights to snow information, enhancing the model's resistance to interference. Furthermore, the CRF module comprehensively refines the decoding features, effectively alleviating the underestimation issues in FSC estimation in some regions and improving the efficiency of detecting fragmented FSC.

However, in this study, cloud interference is avoided in the dataset production, and cloud-snow confusion caused by fragmented clouds still needs to be solved further. In addition, the higher resolution FY4A/AGRI images are only available in a few bands, which directly limits the spatial resolution of FSC estimation in this study. In the future, we expect to improve the spatial resolution of FY4A/AGRI images using super-resolution techniques to produce more accurate snow maps.

REFERENCES

[1] H. Liang, X. D. Huang, Y. H. Sun, Y. L. Wang, and T. G. Liang, "Fractional snow-cover mapping based on MODIS and UAV data over the Tibetan Plateau," *Remote Sens.*, vol. 9, no. 12, Dec. 2017, Art. no. 1332, doi: 10.3390/rs9121332.

[2] J. Hou and C. Huang, "Improving mountainous snow cover fraction mapping via artificial neural networks combined with MODIS and ancillary topographic data," *IEEE Trans. Remote Sens.*, vol. 52, no. 9, pp. 5601–5611, Sep. 2014, doi: 10.1109/TGRS.2013.2290996.

[3] Y. F. Zou et al., "Snow cover in the three stable snow cover areas of China and spatio-temporal patterns of the future," *Remote Sens.*, vol. 14, no. 13, Jul. 2022, Art. no. 3098, doi: 10.3390/rs14133098.

[4] G. X. Wang et al., "Snow-covered area retrieval from Himawari-8 AHI imagery of the Tibetan Plateau," *Remote Sens.*, vol. 11, no. 20, Oct. 2019, Art. no. 2391, doi: 10.3390/rs11202391.

[5] D. E. Rupp et al., "Detection and attribution of observed changes in northern hemisphere spring snow cover," *J. Climate*, vol. 26, pp. 6904–6914, Sep. 2013, doi: 10.1175/JCLI-D-12-00563.1.

[6] H. Wang, S. L. Wang, X. J. Yu, M. X. Wang, and X. Y. Han, "Spatiotemporal variation of snow cover in Xinjiang based on ground observation from 1961 to 2017," *J. Glaciol. Geocryol.*, vol. 42, pp. 72–80, Jun. 2020.

[7] Z. X. Xiao and A. M. Duan, "Impacts of Tibetan Plateau snow cover on the interannual variability of the East Asian summer monsoon," *J. Climate*, vol. 29, pp. 8549–8514, Dec. 2016, doi: 10.1175/JCLI-D-16-0029.1.

[8] W. K. Li et al., "Influence of tibetan plateau snow cover on East Asian atmospheric circulation at medium-range time scales," *Nature Commun.*, vol. 9, Oct. 2018, Art. no. 4243, doi: 10.1038/s41467-018-06762-5.

[9] C. Y. Liu, X. D. Huang, X. B. Li, and T. G. Liang, "MODIS fractional snow cover mapping using machine learning technology in a mountainous area," *Remote Sens.*, vol. 12, no. 6, Dec. 2020, Art. no. 962, doi: 10.3390/rs12060962.

[10] S. Metsämäki et al., "An optical reflectance model-based method for fractional snow cover mapping applicable to continental scale," *Remote Sens. Environ.*, vol. 123, pp. 508–521, Aug. 2012, doi: 10.1016/j.rse.2012.04.010.

[11] L. L. Liu, Q. Xu, Y. Hu, and L. Huang, "Estimating fractional snow cover based on nonlinear NDSI model," *Wuhan Univ. J.*, vol. 37, pp. 534–536, Apr. 2012, doi: 10.13203/j.whugis2012.05.029.

[12] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012, doi: 10.1109/JSTARS.2012.2194696.

[13] K. Rittger, T. H. Painter, and J. Dozier, "Assessment of methods for mapping snow cover from MODIS," *Adv. Water Resour.*, vol. 51, pp. 367–380, Jan. 2013, doi: 10.1016/j.advwatres.2012.03.002.

[14] I. D. Dobreva and A. G. Klein, "Artificial neural networks approach to fractional snow cover mapping," in *Proc. 66th Eastern Snow Conf.*, 2009, pp. 9–11.

[15] I. D. Dobreva and A. G. Klein, "Fractional snow cover mapping through artificial neural network analysis of MODIS surface reflectance," *Remote Sens. Enviroment*, vol. 115, pp. 3355–3366, Dec. 2011, doi: 10.1016/j.rse.2011.07.018.

[16] D. K. Hall, G. A. Riggs, and V. V. Salomonson, "Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data," *Remote Sens. Environ.*, vol. 54, pp. 127–140, Nov. 1995.

[17] V. V. Salmonson and I. Appel, "Development of the aqua MODIS NDSI fractional snow cover algorithm and validation results," *IEEE Trans. Remote Sens.*, vol. 44, no. 7, pp. 1747–1756, Jul. 2006, doi: 10.1109/TGRS.2006.876029.

[18] Y. Zhang, X. D. Huang, W. Wang, and T. G. Liang, "MODIS daily snow cover coverage product validation and algorithm reconstruction," *Arid Zone Res.*, vol. 30, pp. 808–814, Sep. 2013, doi: 10.13866/j.azr.2013.05.026.

[19] X. H. Hao, J. Wang, and H. Y. Li, "Evaluation of the NDSI threshold value in mapping snow cover of MODIS-A case study of snow in the middle Qilian mountains," *J. Glaciol. Geocryol.*, vol. 30, no. 1, pp. 132–138, Feb. 2008.

[20] J. Hou, C. Huang, Y. Zhang, and J. Guo, "On the value of available MODIS and Landsat8 OLI image pairs for MODIS fractional snow cover mapping based on an artificial neural network," *IEEE Trans. Remote Sens.*, vol. 58, no. 6, pp. 4319–4334, Jun. 2020, doi: 10.1109/TGRS.2019.2963075.

[21] J. Zhu, J. C. Shi, and Y. H. Wang, "Subpixel snow mapping of the Qinghai-Tibet Plateau using MODIS data," *Int. J. Appl. Earth Observation Geoinformation*, vol. 18, pp. 251–262, Aug. 2012, doi: 10.1016/j.jag.2012.02.001.

[22] D. K. Hall and G. A. Riggs, "Accuracy assessment of the MODIS snow products," *Hydrol. Process.*, vol. 21, no. 12, pp. 1534–1547, Jun. 2007.

[23] T. Masson et al., "An assessment of existing methodologies to retrieve snow cover fraction from MODIS data," *Remote Sens.*, vol. 10, no. 4, Apr. 2018, Art. no. 619.

[24] R. Nihawan, J. Das, and B. Raman, "A hybrid of deep learning and hand-crafted features based approach for snow cover mapping," *Int. J. Remote Sens.*, vol. 40, pp. 759–773, Jan. 2019, doi: 10.1080/01431161.2018.1519277.

[25] Y. H. Zhang, F. Xu, and X. Kan, "Estimation of snow cover in Xinjiang based on multi-scale feature fusion network," *Comput. Eng.*, vol. 48, pp. 288–295, Jul. 2021.

[26] D. Xing, J. L. Hou, C. L. Huang, and W. M. Zhang, "Estimation of snow depth from AMSR2 and MODIS data based on deep residual learning network," *Remote Sens.*, vol. 14, no. 20, Oct. 2022, Art. no. 5089, doi: 10.3390/rs14205089.

[27] S. Kuter, Z. Akyurek, and G.-W. Weber, "Retrieval of fractional snow covered area from MODIS data by multivariate adaptive regression splines," *Remote Sens. Environ.*, vol. 205, pp. 236–252, Feb. 2018.

[28] H. B. Zhang, F. Zhang, G. Q. Zhang, W. Yan, and S. E. Li, "Enhanced scaling effects significantly lower the ability of MODIS normalized difference snow index to estimate fractional and binary snow cover on the Tibetan Plateau," *J. Hydrol.*, vol. 592, Jan. 2021, Art. no. 125795, doi: 10.1016/j.jhydrol.2020.125795.

[29] National Tibetan Plateau/Third Pole Environment Data Center, "SRTM DEM dataset in China," National Tibetan Plateau, 2013. [Online]. Available: https://data.tpdc.ac.cn

[30] D. Riano, E. Chuvieco, J. Salas, and I. Aguado, "Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types (2003)," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1056–1061, May 2003, doi: 10.1109/TGRS.2003.811693.

[31] J. D. Shepherd and J. R. Dymond, "Correcting satellite imagery for the variance of reflectance and illumination with topography," *Int. J. Remote Sens.*, vol. 24, pp. 3504–3514, Nov. 2010, doi: 10.1080/01431160210154029.

[32] S. A. Soenen, D. R. Peddle, and C. A. Coburn, "SCS+ C: A modified sun-canopy-sensor topographic correction in forested terrain," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2148–2159, Sep. 2005, doi: 10.1109/TGRS.2005.852480.

[33] O. Ronneberger, P. Fischer, and P. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.

[34] H. Y. Fan, "Semantic segmentation of glacier remote sensing images based on deep learning," M.S. thesis, Dept. Computer., Northwest Univ., ShanXi, China, 2021.

[35] D. Q. Zhang, Y. H. Fan, B. S. Kang, J. Gao, and T. J. Li, "Glacier remote sensing image detection in complex background based on improved U-Net network," *J. Appl. Basic Eng. Sci.*, vol. 30, pp. 806–818, Dec. 2022, doi: 10.16058/j.issn.1005-0930.2022.04.002.

[36] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13708–13717.

[37] B. Y. Chen, M. Xia, and J. Q. Huang, "MFANet: A multi-level feature aggregation network for semantic segmentation of land cover," *Remote Sens.*, vol. 13, no. 4, Feb. 2021, Art. no. 731, doi: 10.3390/rs13040731.
[38] K. Y. Chen, "Research on the method of extracting snow information in forested areas based on multi-source remote sensing data," M.S. thesis, Northwest Normal Univ., Lanzhou, China, 2016.
[39] Z. Zhou, S. M. M. Rahman, and N. Tajbakhsh, "Unet++: A nested u-net architecture for medical image segmentation," in *Proc. Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support*, 2018, pp. 3–11.
[40] X. H. Hao, G. J. Wang, J. Wang, X. D. Huang, H. Y. Li, and Y. Liu, "Comparison of spectral feature observation and demixing methods for snow-mixed images," *Spectrosc. Spectral Anal.*, vol. 10, pp. 2753–2758, Oct. 2012.
[41] C. J. Crawford, "MODIS terra collection 6 fractional snow cover validation in mountainous terrain during spring snowmelt using Landsat TM and ETM," *Hydrological Processes*, vol. 29, pp. 128–138, Jan. 2015, doi: 10.1002/hyp.10134.
[42] S. Hao, L. Jiang, J. Shi, G. Wang, and X. Liu, "Assessment of MODIS-based fractional snow cover products over the Tibetan Plateau," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 533–548, Feb. 2019, doi: 10.1109/JSTARS.2018.2879666.
[43] S. Härer, M. Bernhardt, M. Siebers, and K. Schulz, "On the need for a time- and location-dependent estimation of the NDSI threshold value for reducing existing uncertainties in snow cover maps at different scales," *Cryosphere*, vol. 12, pp. 1629–1642, May 2017, doi: 10.5194/tc-12-1629-2018.
[44] H. B. Zhang, F. Zhang, T. Che, and S. S. Wang, "Comparative evaluation of VIIRS daily snow cover product with MODIS for snow detection in China based on ground observations," *Sci. Total Environ.*, vol. 724, pp. 138–156, Jul. 2020, doi: 10.1016/j.scitotenv.2020.138156.
[45] Q. C. Xin et al., "View angle effects on MODIS snow mapping in forests," *Remote Sens. Environ.*, vol. 118, pp. 50–59, Mar. 2012, doi: 10.1016/j.rse.2011.10.029.
[46] X. Wang, J. Wang, T. Che, X. Huang, X. Hao, and H. Li, "Snow cover mapping for complex mountainous forested environments based on a multi-index technique," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1433–1441, May 2018, doi: 10.1109/JSTARS.2018.2810094.

Xu Liu (Member, IEEE) received the B.S. degree in automatization from the Binjiang College of Nanjing University of Information Science and Technology, Nanjing, China, in 2022. He is currently working toward the master's degree in electronic information with the School of Nanjing University of Information Science and Technology, Nanjing.

His research interests include remote sensing image processing and deep learning. He has published two research papers on international journals.

Xi Kan (Member, IEEE) received the Ph.D. degree in meteorological information technology from the Nanjing University of Information Science and Technology, Nanjing, China, in 2019.

He is a Lecturer with Wuxi University, Wuxi, China. He has authored and coauthored more than 20 research papers on international conferences and journals. His research interests include quantitative remote sensing, pattern recognition, deep learning etc.

Dr. Kan is a member of CCF and CAAI.

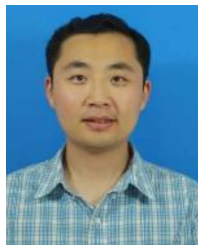
Yonghong Zhang (Member, IEEE) received the Ph.D. degree in mechanical and power engineering from Shanghai Jiao Tong University, Shanghai, China, in 2005.

He is currently a Professor with the Wuxi University and Nanjing University of Information Science and Technology now. He has authored and coauthored more than 50 research papers on international conferences and journals. His research interests include pattern recognition and intelligent systems, remote sensing Big Data analysis and deep learning, integration of intelligent equipment, and IoT systems.

Linglong Zhu (Member, IEEE) received the Ph.D. degree in information and communication engineering from the Nanjing University of Information Science and Technology, Nanjing, China, in 2022.

He is a Lecturer with Wuxi University, Wuxi, China. He has authored and coauthored more than 20 research papers on international conferences and journals. His research interests include quantitative remote sensing, data analysis, deep learning on climate, etc.

Dr. Zhu has been serving as a Reviewer at STOTEN and RS etc.

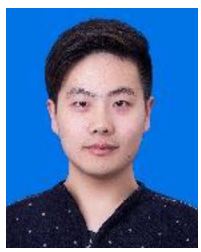


Qi Liu (Senior Member, IEEE) received the Ph.D. degree in data communication and networking from Salford University, Manchester, U.K., in 2010.

He is currently a Professor with the Nanjing University of Information Science and Technology, Nanjing, China. He has published more than 70 research papers on international conferences and journals. His research interests include disaster monitoring and early warning, wireless communication network, distributed system optimization, etc.

Zhou Zhou (Member, IEEE) received the B.S. degree in measurement and control technology and instrumentation from the Nanjing Institute of Technology, Nanjing, China, in 2022. She is currently working toward the master's degree in electronic information from the Nanjing University of Information Science and Technology, Nanjing.

Her research interests include snow depth inversion and machine learning. She has authored and coauthored two research papers on international journals.



Guangyi Ma (Member, IEEE) received the master's degree in control science and engineering from the Nanjing University of Information Science and Technology, Nanjing, China, in 2020. He is currently working toward the Ph.D. degree in information and communication engineering in the School of Nanjing University of Information Science and Technology, Nanjing.

His research interests include remote sensing image processing and satellite precipitation. He has authored and coauthored seven research papers on international conferences and journals.