https://doi.org/10.5194/tc-2022-61
Preprint. Discussion started: 11 April 2022

The Cryosphere
Discussions
Open Access

Glacier extraction based on high spatial resolution remote sensing images using a deep learning approach with attention mechanism

Xinde Chu[1], Xiaojun Yao[1], Hongyu Duan[1], Cong Chen[2], Jing Li[1], Wenlong Pang[3]

[1]College of Geography and Environmental Science, Northwest Normal University, Lanzhou, 730070, China
[2]Key Laboratory of Western China's Environmental Systems (Ministry of Education), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, 730000, China
[3]Xining Center of Natural Resources Comprehensive Survey, China Geological Survey, Xining, 810000, China

Correspondence to: Xiaojun Yao (xj_yao@nwnu.edu.cn)

Abstract. Accurate and quick extraction of glacier boundaries plays an important role in studies of glacier inventory, glacier change and glacier movement, and it faces great opportunities and challenges due to the increasing availability of high-resolution remote sensing images with larger data volume and richer texture informations. In this study, we improved the DeepLab V3+ as Attention DeepLab V3+ and designed a complete solution based on the improved network to automatically extract glacier outlines from the Gaofen-6 PMS images with a spatial resolution of 2 m. In the solution, the Test-Time Augmentation (TTA) was adopted to increase model robustness, and the Convolutional Block Attention Module (CBAM) was added into the Atrous Spatial Pyramid Poolin (ASPP) structure in DeepLab V3+ to enhance the weight of the target pixels and reduce the impact of useless features. The results show that the improved model effectively improves the robustness of the model, enhances the weight of target image elements and reduces the influence of non-target elements. Compared with deep learning models such as FCN, U-Net and DeepLab3+, the improved model performs better, with OA and Kappa coefficients of 99.58% and 0.9915 for the test dataset, respectively. Moreover, our method achieves the highest OA and Kappa of 99.40% and 0.9846 for glacier boundary extraction in parts of the Tanggula Mountains and Kunlun Mountains based on Gaofen-6 PMS images, showing its excellent performance and great potential.

1 Introduction

The cryosphere is one of the five major circles of climate system (Li et al., 2008), of which mountain glaciers are an important part whose changes are closely related to regional climate and are regarded as natural indicators and early warner of climate change (Oerlemans, 1994; Pfeffer et al., 2008; Azam et al., 2018). Since the second half of the 20th century, climate warming has led to the rapid shrinkage of glaciers around the world (Yao et al., 2012), causing important impact on the utilization of regional water resources and rise of sea level (King et al., 2012; Grinsted, 2013; Schrama et al., 2014). Accurate extraction of glacier boundaries can help to detect the status of glacier areas (Racoviteanu et al., 2015) and understand the response pattern of glaciers to climate change (Bishop et al., 2004; Sun et al., 2018). As a large-range and long-range sensing technology for ground exploration, remote sensing can rapidly achieve glacier information including their boundary, velocity, etc (Robson et al., 2015; Zhang et al., 2019). However, most of the existing researches of glacier changes and glacier inventories are based on remote sensing images of low- to medium-resolution such as Landsat series and Aster images (Liu et al., 2020; Zhao et al., 2020), which may result in an inaccurate estimation of the global glacier resource to some extent (e.g., the glacier area threshold for Randolph Glacier Inventory (RGI) is 0.01 km^2). The available remote sensing data are increasingly abundant with the successive launches of high-resolution remote sensing satellites, and the efficient and rapid acquisition of glacier boundaries based on these data is currently a frontier issue in glacier remote sensing research.

Glacier boundaries are generally extracted by manual visual interpretation and semi-automatic or automatic methods. The former can yield a relatively accurate results, but the applications based on high-resolution imagery over a wide area are time-consuming and laborious (Yan and Wang, 2013). The latter extracts glaciers based on its spectral differences from other features that snow/ice has a strong absorption in the short-wave infrared band (1.55-1.75 μm) and a robust reflection in visible to near-infrared band (0.45-0.90 μm) (Guo et al., 2017), mainly including ratio method (Ji et al., 2020), snow cover index (Wang et al., 2021a), supervised classification and unsupervised classification (Nie et al., 2010), etc. However, the absence of short-wave infrared band in some high-resolution optical remote sensing images (such as QuickBird satellite images, WorldView-2 satellite images, SPOT-6 NAOMI and Gaofen-6 PMS) limits the application of the ratio method and Normalized Difference Snow Index (NDSI) which have a better extraction effect.

Deep learning has been widely adopted in the field of computer vision and image processing in recent years (Girshick, 2015), which can automatically obtain mid- and high-level abstract features from images due to its powerful feature learning and characterization capabilities compared with the traditional classification methods (Redmon et al., 2016). At present, numerous typical deep learning models, such as Full Convolutional Network (FCN) (Long et al., 2015), Segnet (Badrinarayanan et al., 2017), U-Net (Ronneberger et al., 2015), and DeepLab series (Chen et al., 2018) have been successfully applied to the semantic segmentation task of remote sensing images (Huang et al., 2018; Tong et al., 2020) including the cryosphere domain. Zhang et al. (2019) automatically delineated the calving front of Jakobshavn Isbræ Glacier using a deep learning method; Robson et al. (2020) combined deep learning and object-based image analysis to extract rock glacier boundaries in the La Laguna catchment in Northern Chile and the Poiqu catchment in Central Himalaya; He et al. (2021) extracted the glacial lakes of the Alatau Mountains of Tianshan through deep learning; Marochov et al. (2021) segmented Sentinel-2 image covered marine-terminating outlet glaciers in Greenland into seven classes using Convolutional Neural Network; Baumhoer et al. (2019) extracted Antarctic glacier and ice shelf fronts at nine locations using deep learning based on Sentinel-1 images. However, the attention mechanism has not been used in the glacier extraction, and the most studies were limited to extract small-scale glaciers in some regions. Therefore, the combination of deep learning and attention mechanism has the potential to provide an effective and powerful technique for the automatic extraction of mountain glaciers.

The main objective of this study is to propose a new method for automatic extraction of glacier boundaries from high resolution Gaofen-6 PMS images based on DeepLab V3+ network and attention mechanism. Then, to ascertain the accuracy and robustness of the proposed method by comparing with the reference outlines of glaciers based on manual interpretation of orthorectified images taking parts of the Tanggula Mountain and Kunlun Mountain as the test region. Meanwhile, we assess our result by comparing with GAMDAM glacier inventory (GGI) (Nuimura et al., 2015) and the glacier coverage data on the Tibetan Plateau in 2017 (TPG2017) (Ye et al., 2017; Ye, 2019).

2 Model structure and data process scheme

In this study, we used the Deeplab V3+ in combination with the attention mechanism (section 2.1) to explore the glacier extraction method based on high spatial resolution images. The Gaofen-6 images were preprocessed and divided into two groups, firstly, the former was used to make samples by data augmentation to train the improved Deeplab V3+, then the latter was used to test the trained model by performing the classification (section 2.3). Meanwhile, the Test-time augmentation (TTA) was added into this classification to improve model accuracy, which is described in Section 2.4. Figure 1 shows the overall flow of the method.

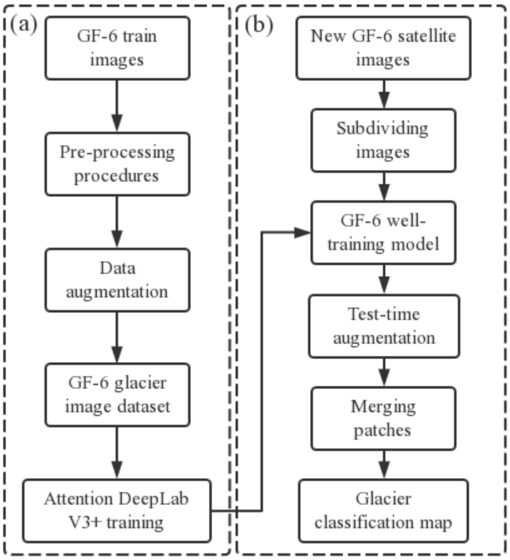
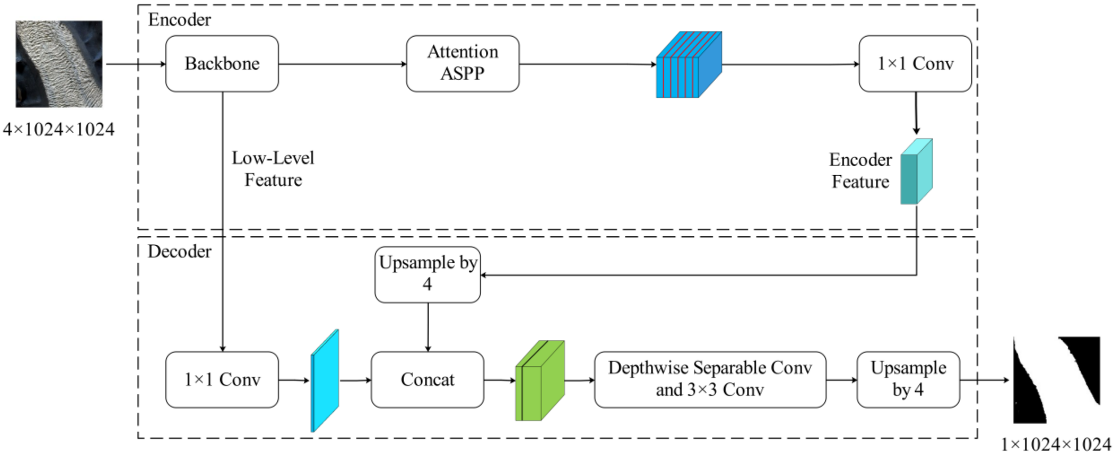

(a)
GF-6 train images
Pre-processing procedures
Data augmentation
GF-6 glacier image dataset
Attention DeepLab V3+ training
(b)
New GF-6 satellite images
Subdividing images
GF-6 well-training model
Test-time augmentation
Merging patches
Glacier classification map

Figure 1. The overall flow of glacier extraction based on deep learning, (a) the model training, (b) the predict process.

2.1 Network structure

DeepLab V3+, an encoding-decoding architecture proposed by Chen et al. (2018), is one of the most advanced semantic segmentation algorithms. In this paper, we improved the DeepLab V3+ model by adding an attention mechanism to the encoding-decoding structure (Attention DeepLab V3+). In the encoder, the ResNet 34 (He et al., 2016) was used as the backbone network to extract semantic information to obtain the low-level feature and then the Atrous Spatial Pyramid Pooling

(ASPP) module with attention mechanism was connected to obtain the encoder feature. In the decoder, the low-level feature and the encoder feature maps were input. The encoder feature performed a upsampling with a factor of four and then fused with the low-level feature. After that, a depthwise separable convolution and a bilinear interpolation upsampling with a factor of four were executed to extract the expected features and output them at the same size as the input image (Fig. 2).



85

Figure 2. Architecture of Attention DeepLab V3+.

2.1.1 Attention mechanism

The attention mechanism can learn contextual information and capture the internal correlation, its basic idea is to ignore irrelevant information and focus on key information in operations (Woo et al., 2018). In this paper, we used the Convolutional Block Attention Module (CBAM) (Fig. 3) (Woo et al., 2018) including a channel attention module and a spatial attention module to obtain the attention weights along both channel and spatial dimension in turn, and then multiplied with the original feature map to adaptively adjust the features and increase the weights of target features.

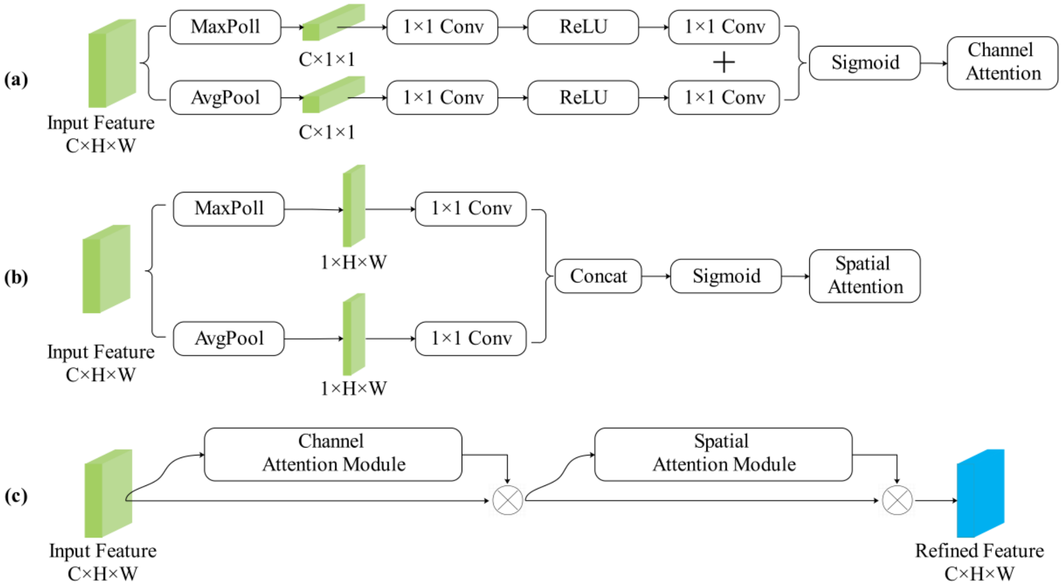


Figure 3. Attention module, (a)-(c) Channel attention module, Spatial attention module, CBAM, respectively.

2.1.2 Attention ASPP

The ASPP structure replaces the partial volumes of the deep neural network with a dilated convolution (Yu and Koltun, 2015), which expands the perceptual field without increasing the parameters, thus obtaining more feature information. However, the dilated convolution may cause discontinuity of spatial information, which was addressed by incorporating CBAM into the juxtaposition structure of the extracted features in this paper. The role of the attention mechanism was to focus on the noticed target pixel and enhance its weight, and the dilated convolution in the ASPP can obtain contextual information at different scales. Adding the attention mechanism could make the features of related categories more aggregated, thus effectively reduce the phenomenon of empty features. The five branches of the juxtaposition structure obtained more contextual information by collecting features from different sensory domains, and by combining this information with refined feature maps, dependencies between pixels and differences between categories were gained (Fig. 4).

https://doi.org/10.5194/tc-2022-61
Preprint. Discussion started: 11 April 2022

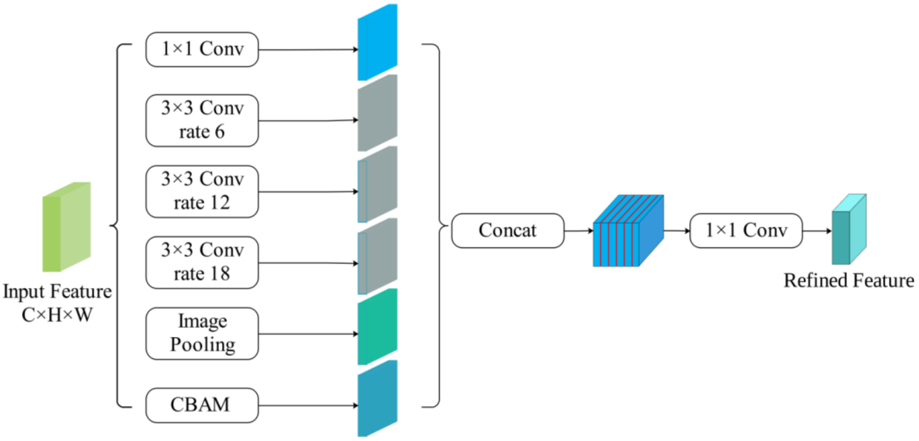


Figure 4. The structure of Attention ASPP.

2.1.3 Backbone

The ability of CNN to retrieve relevant information from images is enhanced with the increase of the network depth (Telgarsky, 2016), but the too deep network could lead to the gradient explosion and network degradation. Residual connections (He et al., 2016) solved this problem by feeding a given layer into the previous one where a building block of residual learning was included (Fig. 5), and by which, the depth of the network and learning capacity can be dramatically increased. ResNet has many branches with different numbers of blocks, and we adopt ResNet-34 with 16 blocks and 33 convolutional layers in total in the present study (He et al., 2016).

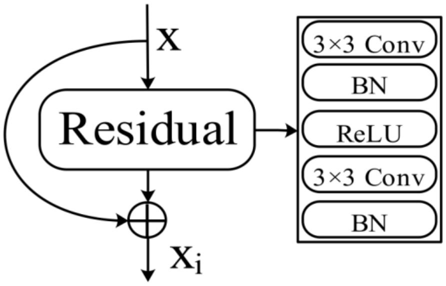


Figure 5. Residual connection. Each residual block is composed by convolutions (Conv), batch normalizations (BN) and rectified linear units (ReLU).

2.1.4 Depthwise separable convolution

In this paper, we added a depthwise separable convolution at the end of semantic segmentation after combining high-level and low-level features (Chollet, 2017; Howard et al., 2017). This convolution can decompose the traditional convolution into a depthwise convolution and a 1 × 1 pointwise convolution, improving the efficiency of operation without losing too much

https://doi.org/10.5194/tc-2022-61
Preprint. Discussion started: 11 April 2022

accuracy compared with normal convolution. In addition, the layers of neural network with depthwise separable convolution can be deeper for the same number of parameters.

2.1.5 Loss function

Due to the high resolution of Gaofen-6 PMS images and the large size of some glaciers, only a small part of glaciers or other features may be present in the sample, which has the potential to cause the sample imbalance. In this paper, the dice loss (Milletari et al., 2016) was used as the loss function to mitigate this phenomenon. The dice loss is expressed as:

$$L_d = 1 - \frac{I+\varepsilon}{U+\varepsilon} \tag{1}$$

where *I* is the number of intersection between sample labels and predicted pixels, *U* is the sum of sample labels and predicted pixels, and *ε* is a constant mainly used to prevent the denominator from being zero and smooth the loss operation.

The use of dice loss generally produces severe oscillations during the training of the model when the positive sample is a small target because a large loss change and a consequently gradient drastic change will occur once some pixels of the small target are predicted incorrectly in the case of only foreground and background. Therefore, we introduced a combination of cross entropy loss and dice loss to make the network training more stable, and the cross-entropy loss is as follows:

$$L_c = \frac{1}{N}\sum_i -[y_i \times \log(p_i) + (1-y_i)\times\log(1-p_i)] \tag{2}$$

where p_i is the probability that the pixel is predicted as glacier, and y_i is the sample label which takes the value of 1 if the sample pixel being a glacier, and 0 otherwise.

Therefore, these above two losses were fused as the final loss function.

$$L = 0.5\times(L_c + L_d) \tag{3}$$

2.2 The pre-processing procedures of Gaofen-6 PMS image and datasets production

Deep learning requires a large amount of labeled data related to the classification target for training the model. However, the currently available open-source datasets cannot meet the requirements of the classification in this paper. Therefore, we collected Gaofen-6 PMS images with the less cloudy and snowy from China High-resolution Earth Observation System (https://www.cheosgrid.org.cn/) (Tab. 1) as a training and validation dataset of the model, some of which were selected to test the accuracy of the glacier extraction method. The Gaofen-6 satellite, officially operational since March 21, 2019, is a low-orbiting optical remote sensing satellite with high spatial resolution, featuring wide coverage, high quality and efficient imaging. A two-meter panchromatic/eight-meter multispectral high-resolution camera and a 16-meter multispectral medium-resolution wide field camera are boarded on Gaofen-6 satellite, the former with an observation width of 90 km and the latter with that of 800 km.

Table 1. Descriptions of Gaofen-6 PMS images used in this paper.

ID	Date	Resolution(m)	Utilization
Gaofen6_PMS_E90.9_N35.8	2020-09-21	2	Training and validation
Gaofen6_PMS_E91.1_N36.5	2020-09-21		
Gaofen6_PMS_E91.5_N35.8	2020-09-25		
Gaofen6_PMS_E91.7_N36.5	2020-09-25		
Gaofen6_PMS_E91.5_N35.8	2020-11-05		
Gaofen6_PMS_E91.0_N33.6	2020-12-16		
Gaofen6_PMS_E91.4_N32.8	2020-12-20		
Gaofen6_PMS_E91.9_N32.8	2020-08-19		Test
Gaofen6_PMS_E92.1_N35.8	2020-09-29		
Gaofen6_PMS_E91.6_N33.6	2020-12-20		

Pre-processing, including fusion, orthorectification, geometric alignment and some other operations were performed prior to using the original defective images. Glaciers in the images were extracted as true values by manual visual interpretation. Considering the high spatial resolution of Gaofen-6 PMS images, the larger scale of glaciers and other features, and the inclusion of two kinds of features (glaciers and other features) in the samples whenever possible, the images were cropped into roughly 1024×1024 size and used as the input for deep learning training. It is important to prepare a sufficiently diverse dataset to ensure that the model can be adapted to different scenarios of glacier extraction. In this paper, the data augmentation including randomly clip the Gaofen-6 PMS images, vertical flipping, horizontal flipping and diagonal flipping of the samples as well as clockwise 90° and counterclockwise 90° rotations were conducted to expand the sample library, improve the model accuracy and enhance its generalization performance. Finally, a training set and a validation set containing 3600 well-annotated images of 1024×1024 size with blue, green, red and near infrared bands were obtained. Meanwhile, we kept a test set containing 400 images without data augmentation. An example of the sample is shown in Fig. 6.

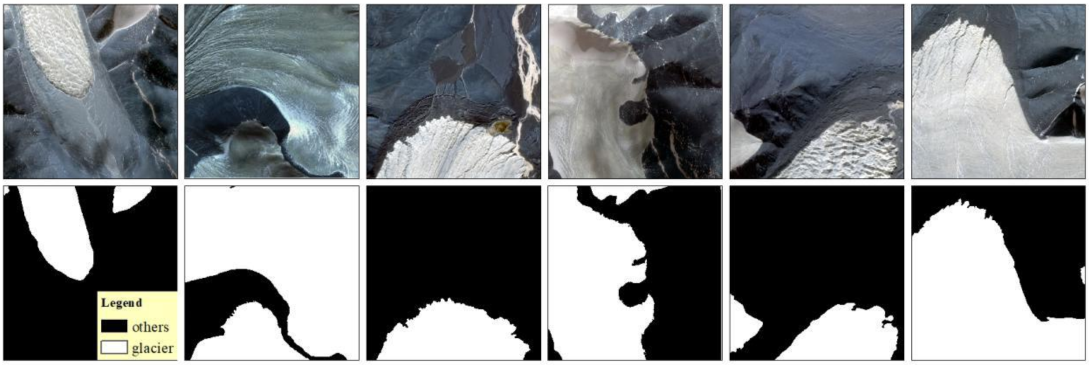
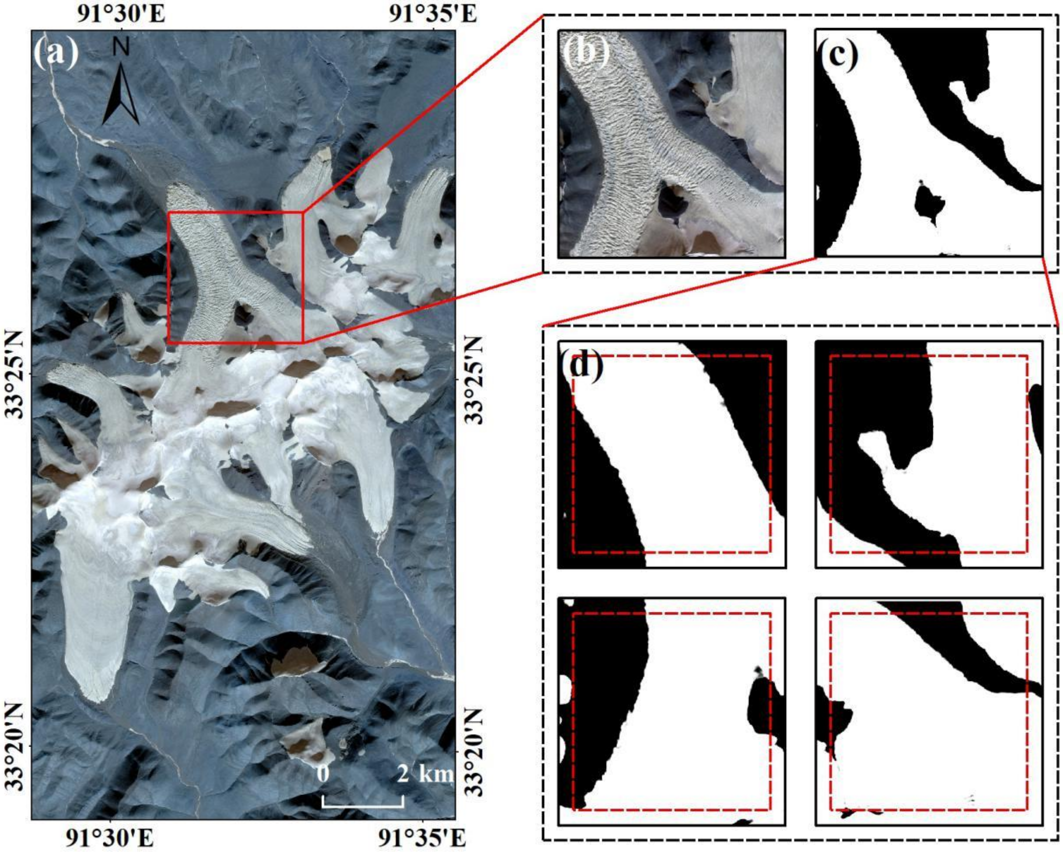


Figure 6. The RGB Gaofen-6 samples and ground truth are displayed in the first and second rows.

2.3 Complete glacier extraction

A sample usually displays only a portion of the glacier, so an image larger than the sample size needs to be input to extract the
165 complete glacier. Usually, the images to be classified were generally split into a series of images with the same size as the samples and fed into the network for prediction, and then the predicted results are merged into one final result image in the cropping order in the prediction process. However, larger classification errors and unsmooth merging after clipping could happen due to the insufficient pixel features in the edge areas of each patch. Hence, we adopt the strategy to make the two patches clipped overlap each other to preserve the features of edge pixels and make the merging of edges smoother.

170 Figure 7 illustrates the process of prediction, in which, the pre-processed original image over the area in the red checkbox (Fig. 7a and Fig. 7b) must be spilt into 1024×1024 overlapping patches before being input into the network for prediction, and then the classification results of each patch are obtained (Fig. 7d). Subsequently, 90% of the central part of each patch (red checkbox in Fig. 7d) was merged to obtain the classification result (Fig. 7c).



175 **Figure 7.** Post processing, (a) and (b) the RGB images obtained on December 20, 2020, (c) the final result of merged, (d) the predict results.

2.4 Test-time augmentation

In this paper, the TTA strategy was used in the result extraction, which can be considered as a post-processing technique because it is executed during the testing phase (Wang et al., 2021b). Therefore, it does not affect the network learning parameters, but tries to obtain multiple enhanced copies by performing data enhancement operations such as horizontal and vertical transformations on each image during the test, then combining the results of multiple enhanced copies for prediction (Fig. 8). The voting formula is as follows:

$$p = \frac{\sum_{v=1}^{n} S}{n} \tag{4}$$

where p is the probability that the pixel belongs to glacier. n is the number of each test image and its copies. S is the probability of each pixel belonging to the glacier in the image and its copies. Figure 8 shows the result of TTA.

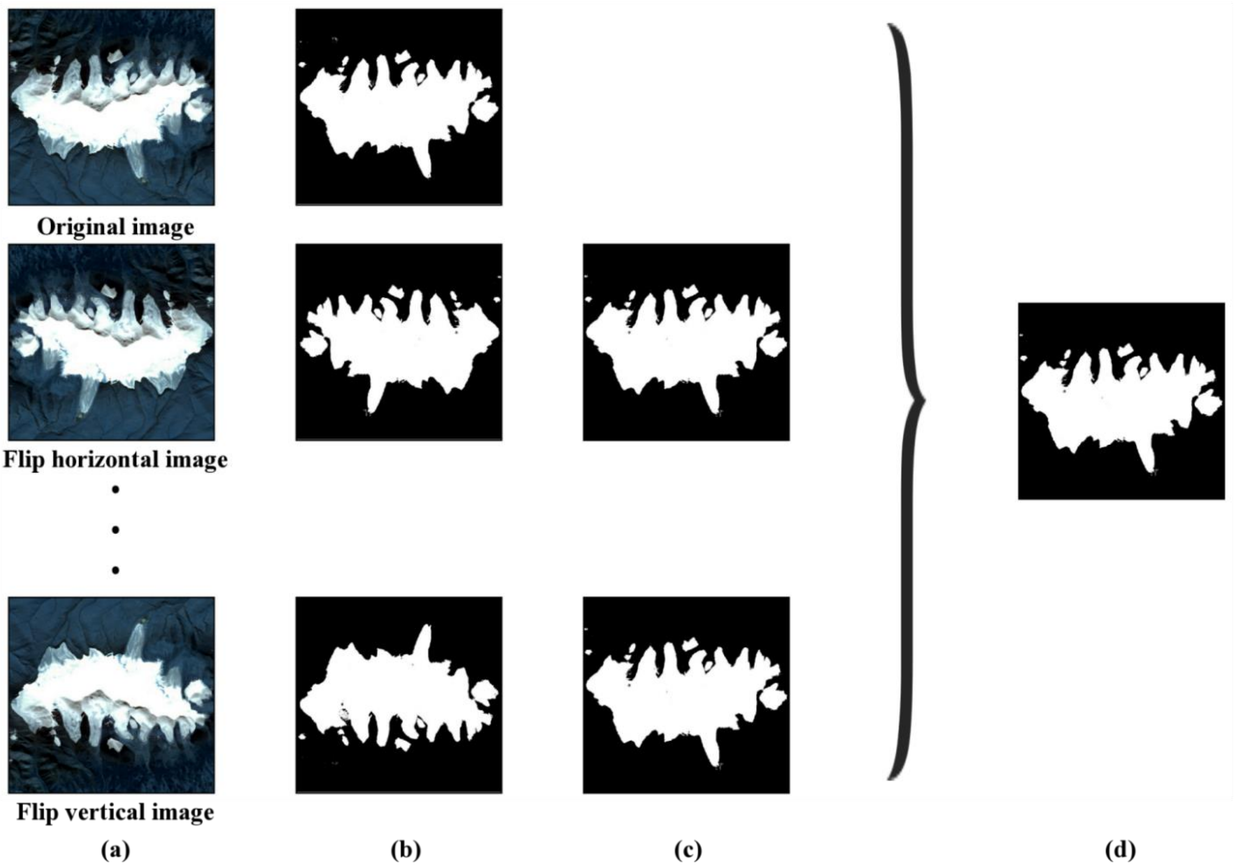


Figure 8. Test-time augmentation result. (a)-(d) the RGB image obtained on September 29, 2020, its copies, predict results, reductive copies results, the result of the vote, respectively.

2.5 Evaluation metrics

To quantitatively describe the ability to extract glaciers from high-resolution images using the method proposed by this study, the results obtained from manual visual interpretation were chosen as ground truth to be compared with the classification results. The overall accuracy (OA) and Kappa coefficient (Kappa) of the confusion matrix were used for accuracy evaluation (Olofsson et al., 2014), where *OA* is the ratio of correctly classified pixels to all pixels in the entire image, and can be calculated as follows:

$$OA = \frac{P_1}{n} \tag{5}$$

where p_1 and n are the number of correctly classified pixels and total pixels, respectively.

The *Kappa* is a statistic that measures the agreement between prediction and ground truth, and can be calculated as follows:

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \tag{6}$$

$$p_e = \frac{a_1 \times b_1 + a_2 \times b_2}{n^2} \tag{7}$$

where p_0 is the *OA*. a_1 and a_2 are the number of true samples of glaciers and other features, respectively. b_1 and b_2 are the number of predicted samples of each category, respectively.

3 Experimental result

This section will test the algorithms in this paper, analyze the experimental results and address the following two questions: (1) examining the performance of the Attention DeepLab V3+ with TTA on the test set; (2) evaluating the ability of our method to extract large scale glaciers.

3.1 Experimental setup

For the experimental platform, we used a central processing unit with processor Intel Core i9-10920 (3.50 GHz) which configured with 64GB of memory, a graphics card with Nvidia GeForce RTX 2080 Ti 11GB of video memory, Windows 10 64-bit operating system, and python programming implementation. In terms of software environment, we chose the pytorch as the deep learning framework, CUDA version 11.1 as the graphics processing unit (GPU) computing platform and cuDNN8.0 as deep learning GPU acceleration library.

In this paper, the ratio of training set, validation set and test set was 8: 1: 1. The training set was used to optimize the network parameters (weights and bias), the validation set to prevent overfitting and optimize the hyperparameters of the network (learning rate), and the test set to evaluate the effectiveness of the model trained on the training set, where the images in the test set were not processed by data augmentation.

Gaofen-6 PMS image, with spatial resolution of 2 m, allows a large number of pixels occupied by glacier. Therefore, four scenes of Gaofen-6 PMS image covering glaciers in parts regions of Tanggula Mountains and Kunlun Mountains (Image 1: 91°32′7″ E, 33°24′3″ N; Image 2: 91°48′53″ E, 32°57′27″ N: Image 3: 91°57′3″ E, 35°49′53″ N; Image 4: 92°4′47″ E, 33°6′29″ N) were selected to test the capability of extracting the large-scale glaciers by the model in this paper. The whole visual procedure to extract glacier is shown in Fig. 9.

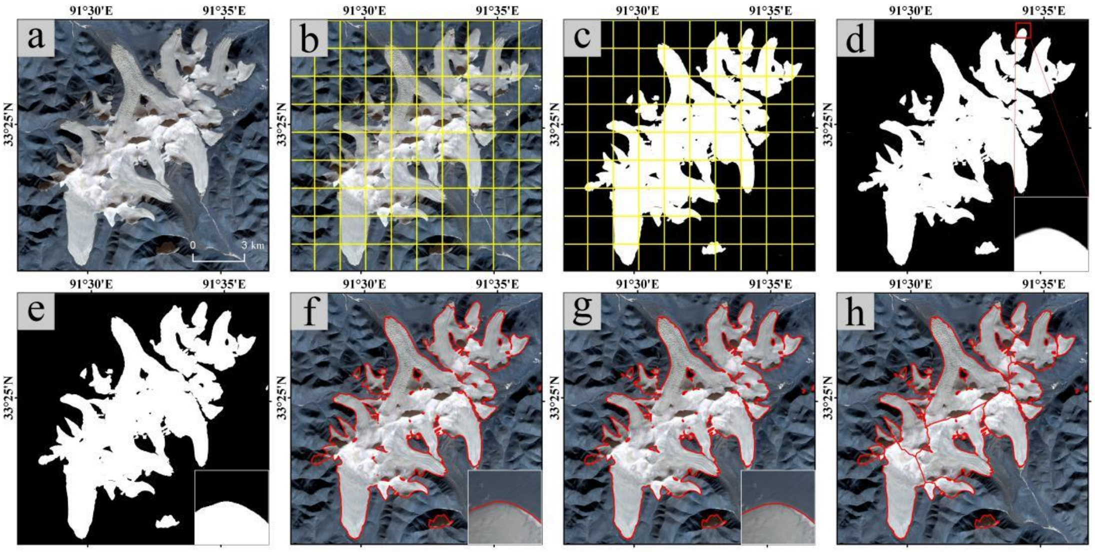


Figure 9. The procedure of glacier extraction. Firstly, the pre-processed image obtained on December 20, 2020 (a) were clipped into 1024 x 1024 overlapping patches (b), then input into the model to predicted result (c). Secondly, we merged the results and classified the pixel values greater than 0.5 as glaciers to obtain the final binary map (d and e). Thirdly, the binary map from above can be converted to a vector (f) and smoothed (g). Finally, the glacier was segmented using ASTER GDEM to obtain individual glacier vector boundaries (h).

3.2 Experiments on Gaofen-6 test set

In this section, we explored the effectiveness of our network on the test set by comparing with FCN, U-Net and DeepLab V3+, in which FCN was specifically used FCN 32s network and U-Net was added the backbone network of Resnet 18. Figure 10 shows the visualized prediction results of some test sets on different methods. The extracted results derived from FCN were almost error-free, however, had a poor performance on the test set, which may due to its direct upsampling of 16, resulting in the loss of the detailed information in glacier boundaries. The difference between the performances of U-Net and DeepLab V3+ on the test set was small, U-Net worked better in Fig. 10-4/5/6 and DeepLab in Fig. 10-1/2/3. It is obvious that the Attention DeepLab V3+ with TTA model has the best glacier extraction capability, which extracts glacier boundary with excellent continuity and fewer fine patches.

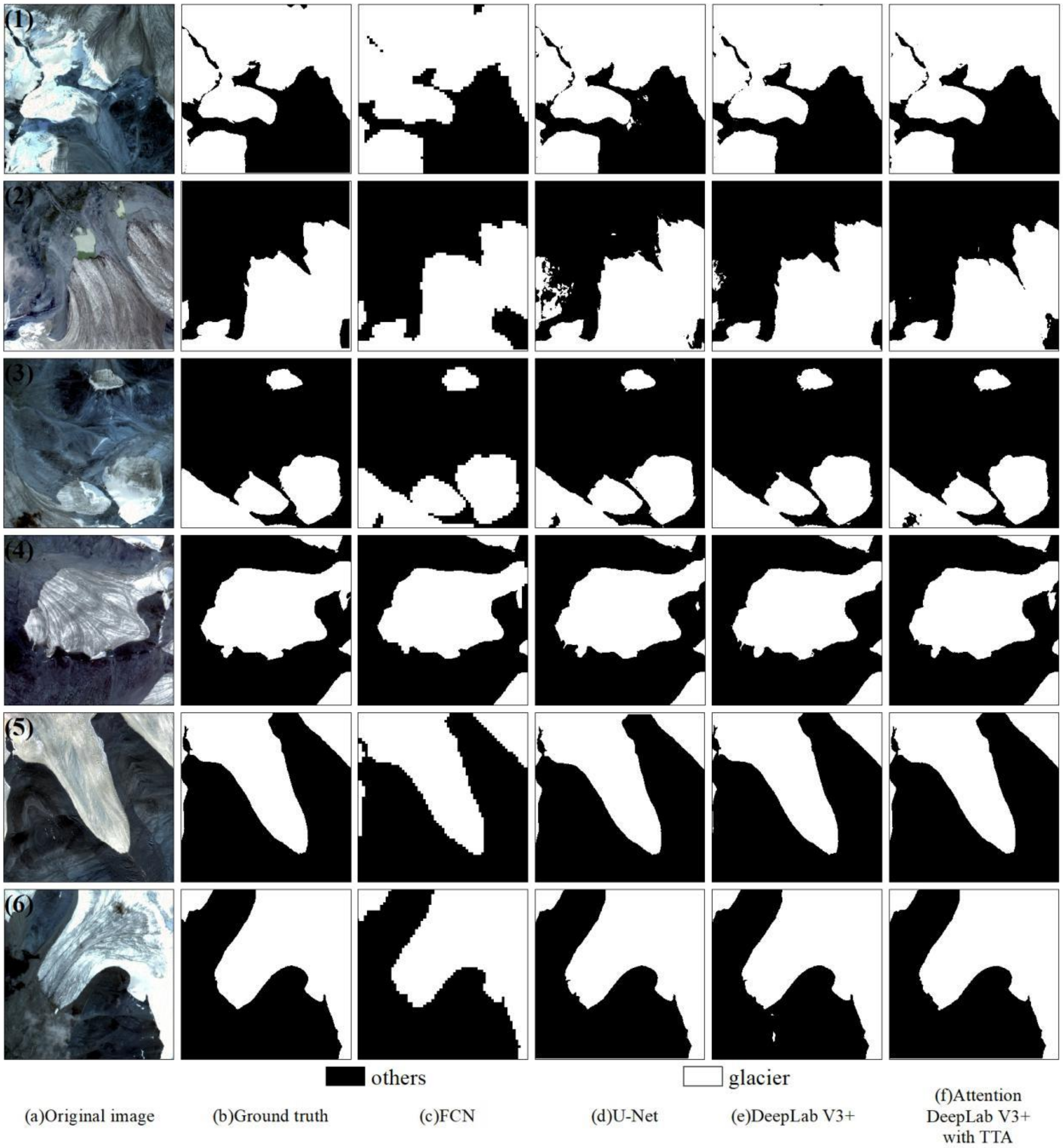
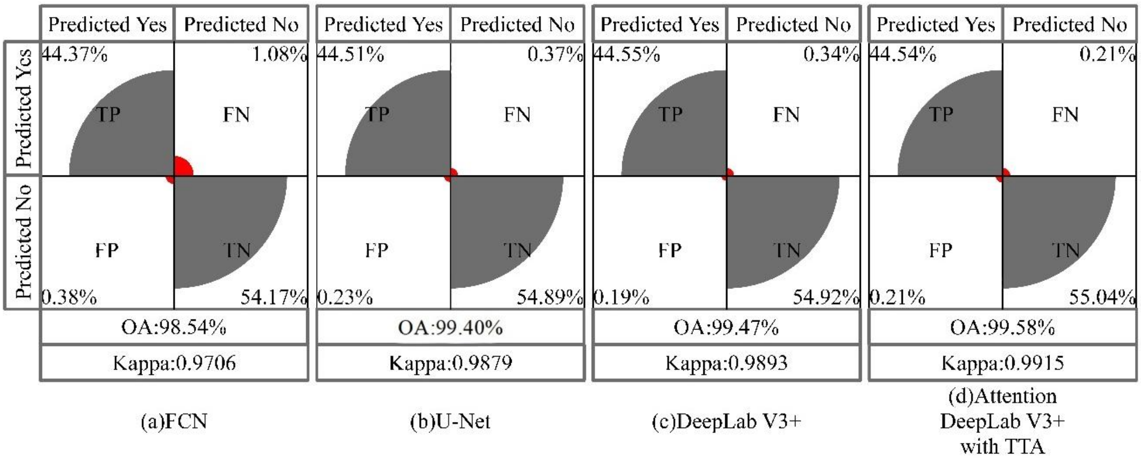


Figure 10. Comparison of test results of different networks.

Figure 11 shows the calculated confusion matrix, OA and Kappa of each model on the test sets. In terms of OA, the Attention DeepLab V3+ with TTA model achieved the highest score of 99.58%, with misidentified pixels accounted for 0.42% of the total pixels. This was followed by DeepLab V3+ with the score of 99.47% and a misidentified pixels percentage of 0.53% duo to more glacier pixels being misclassified as other features. Similar to DeepLab V3+, U-Net misclassified glaciers as other features, with the misidentified pixels percentage of 0.60% of and OA of 99.40% that lower than the Attention DeepLab V3+ with TTA model 0.18%. The comparison results of Kappa were similar to that of OA, with the highest value of Attention

DeepLab V3+ with TTA model (0.9915), followed by DeepLab V3+ model (0.9893) and U-Net (0.9879), lowest by FCN (0.9706).

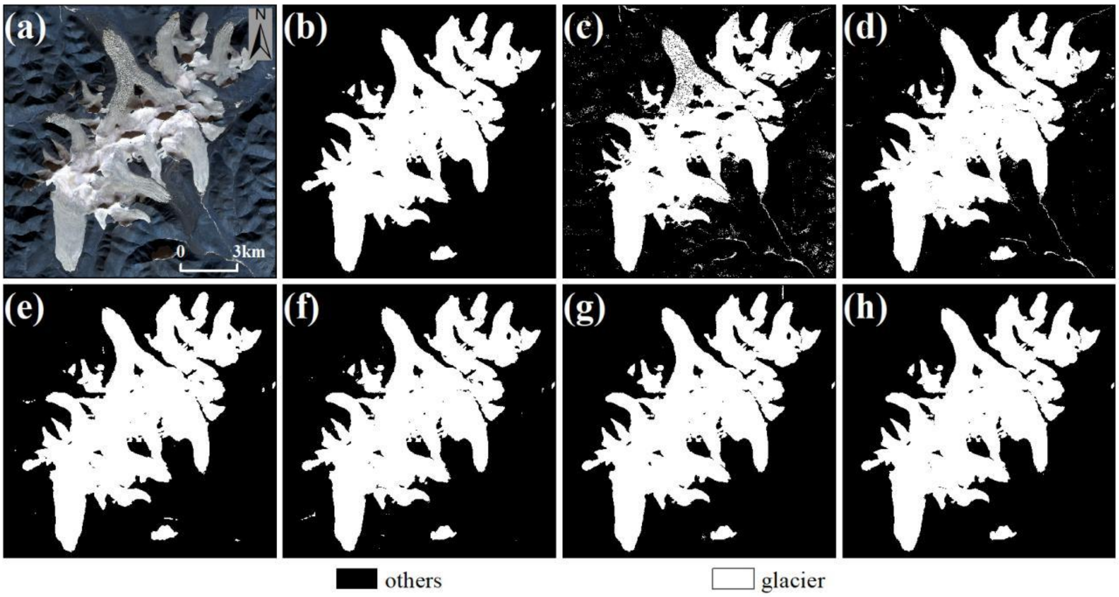


(a)FCN (b)U-Net (c)DeepLab V3+ (d)Attention DeepLab V3+ with TTA

Figure 11. Performance of extraction models by confusion matrix, OA and Kappa.

3.3 Experiments on Gaofen-6 image

In this section, we compared the ability of extracting complete glaciers based on four scene of Gaofen-6 images (section 3.1) using our method, Single-Band Threshold Method (SBTM) and Random Forest (RF). Meanwhile, FCN, U-Net and DeepLab V3+ were selected for comparation because the complexity of the large-scale remote sensing images may result in the performance on the test set being unsuitable for extracting the huge complete glaciers. The experimental results of each method are shown in Fig. 12-15, and OA and Kappa were calculated for the extraction results to evaluate the performance of different methods (Tab. 2).



255 **Figure 12.** Glacier extraction maps of the image (1) obtained on December 20, 2020. (a) the RGB image. (b) the ground truth. (c)–(h) SBTM, RF, FCN, U-Net, DeepLab V3+ and Attention DeepLab V3+ with TTA.

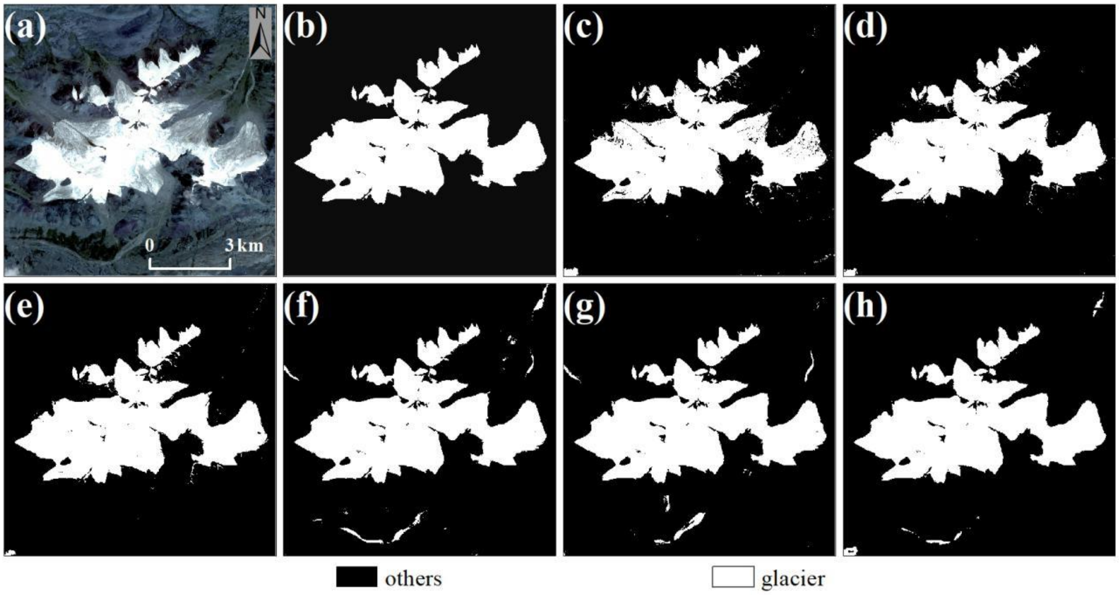


Figure 13. Glacier extraction maps of the image (2) obtained on August 19, 2020. (a) the RGB image. (b) the ground truth. (c)–(h) SBTM, RF, FCN, U-Net, DeepLab V3+ and Attention DeepLab V3+ with TTA.

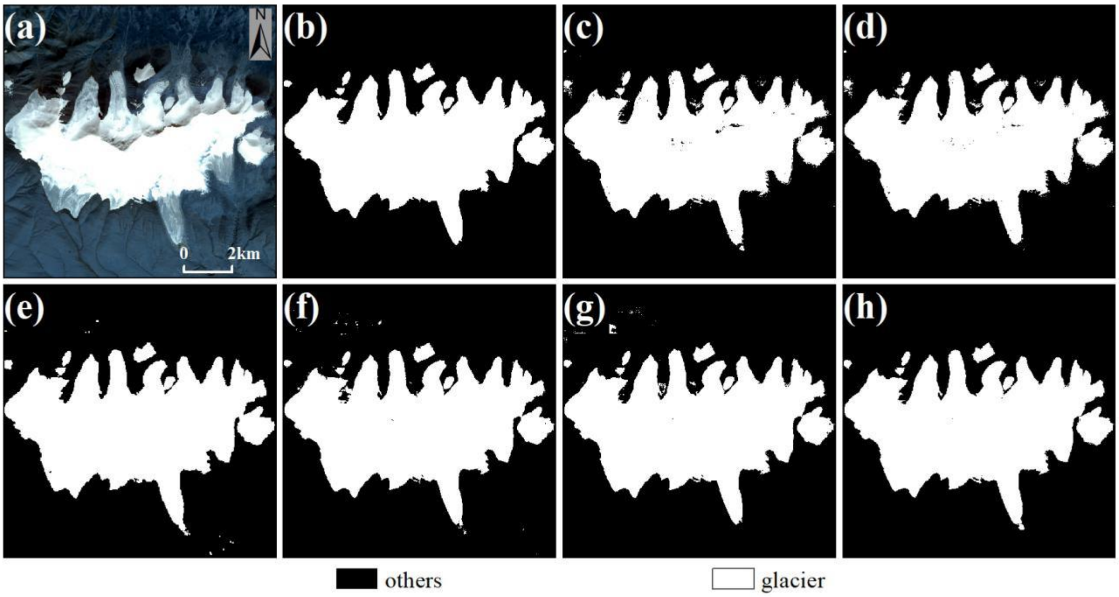


Figure 14. Glacier extraction maps of the image (3) obtained on September 29, 2020. (a) the RGB image. (b) the ground truth. (c)–(h) SBTM, RF, FCN, U-Net, DeepLab V3+ and Attention DeepLab V3+ with TTA.

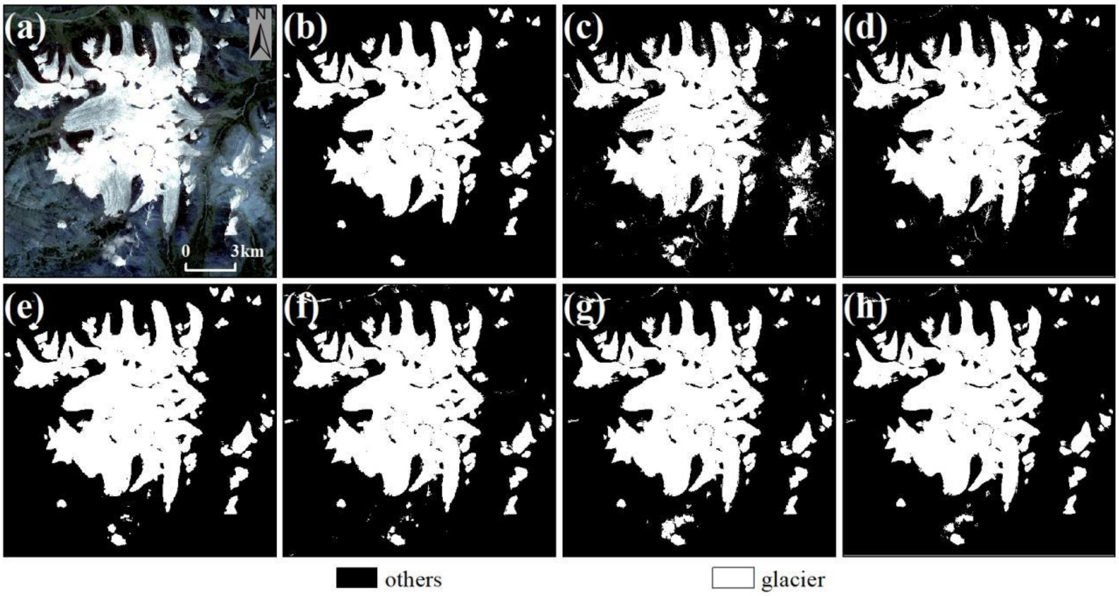


Figure 15. Glacier extraction maps of the image (4) obtained on August 19, 2020. (a) the RGB image. (b) the ground truth. (c)–(h) SBTM, RF, FCN, U-Net, DeepLab V3+ and Attention DeepLab V3+ with TTA.

The SBTM performed well when the spectral features of images are simple (Fig. 12), but had the worst effectiveness when the reflectance of some glaciers is similar to that of other features (Fig. 12 to 15). The RF had a better glacier extraction capability than SBTM due to its use of four bands. The deep learning yielded a better result than above methods, which was trace-free and smooth despite being merged of many sample-sized images. Among the deep learning methods, our method had the best performance with extraction results similar to the ground truth.

In terms of accuracy, the SBTM has the lowest accuracy with OA and Kappa of 97.3% and 0.9425, respectively, followed by the FCN and RF. It should be noted that OA and Kappa of RF were 98.97% and 0.9740, respectively, which were higher than that of FCN (98.80%, 0.9702) on the whole, but lower for image (1). It was followed by U-Net and DeepLab V3+, but they also had a good score. Our method had the best accuracy with the highest scores in all test images and the highest OA and Kappa of 99.40% and 0.9846, respectively.

Table 2. Comparison of different glacier extraction methods on Gaofen-6 image.

Gaofen-6 images	Image (1)		Image (2)		Image (3)		Image (4)		Total	
Evaluation metrics	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa
SBTM	95.51	0.8892	98.51	0.9565	99.09	0.9790	97.80	0.9454	97.73	0.9425
RF	98.99	0.9755	99.07	0.9730	99.16	0.9807	98.67	0.9669	98.97	0.9740
FCN	99.08	0.9779	98.82	0.9678	98.96	0.9761	98.35	0.9592	98.80	0.9702
U-Net	99.58	0.9897	99.10	0.9739	99.42	0.9866	98.76	0.9692	99.22	0.9799
DeepLab V3+	99.74	0.9937	99.19	0.9761	99.50	0.9885	98.77	0.9694	99.30	0.9819
Attention DeepLab V3+ with TTA	99.76	0.9942	99.31	0.9800	99.56	0.9897	98.98	0.9746	99.40	0.9846

3.4 Comparative Experiment with or without TTA

We improved the accuracy of Attention Deeplab V3+ model by employing TTA during testing, and verified the effectiveness by comparing the results with that without TTA. Table.3 shows the scores of each evaluation index, in which, the Attention DeepLab V3+ is more accurate than the other networks tested in this paper, with the higher OA and Kappa (99.54% and 0.9908) than DeepLab V3+ (99.47%, 0.9893). The addition of TTA increases the OA and Kappa of the network by 0.04% and 0.0006, respectively, and removes some discriminative errors of the pixels (Fig. 16), improving the performance of the model in extracting glaciers.

Table.3 confusion matrix of test set result.

Method	Attention DeepLab V3+		Attention DeepLab V3+ with TTA	
Pixel number	Predicted Yes	Predicted NO	Predicted Yes	Predicted NO
Actual Yes	186712209	949981	186798674	885576
Actual No	966562	230801649	880097	230866053

OA	99.54%	99.58%
Kappa	0.9908	0.9915

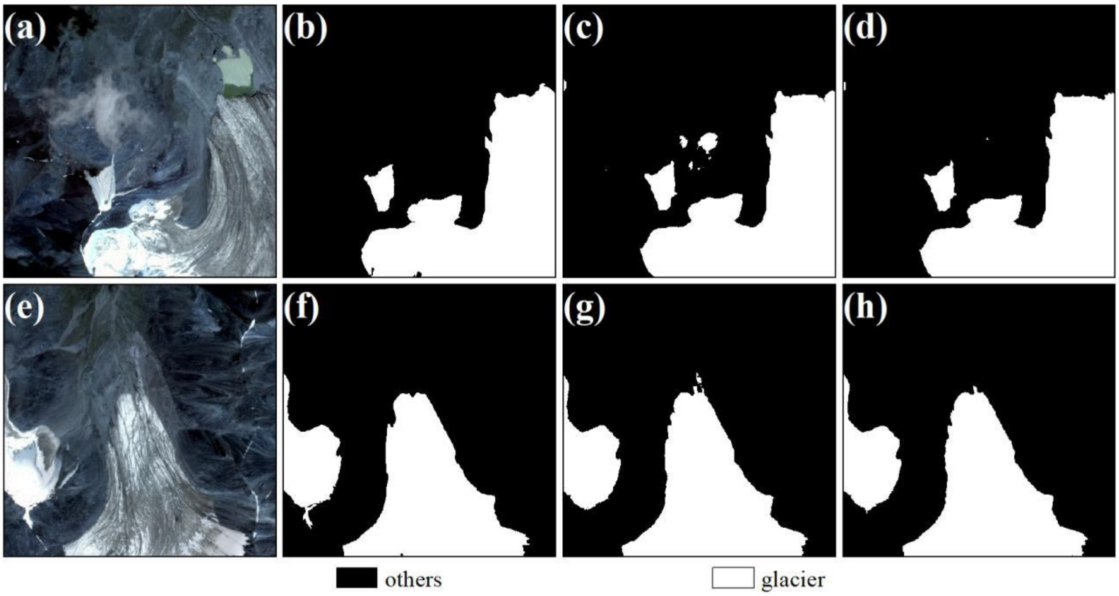


Figure 16. Examples of the results with or without TTA, (a) and (e) are RGB images, (b) and (f) are ground truth, (c) and (g) are Attention DeepLab V3+, (d) and (h) are Attention DeepLab V3+ with TTA.

4 Discussion

4.1 Advantages and limitations

Compared with previous studies on glacier extraction, the main achievements of this paper are: (1) building a glacier dataset with high spatial resolution for deep learning, (2) improving the DeepLab V3+ model by adding attention mechanism, (3) proposing an effective method to extract glaciers from high-resolution remote sensing images.

We expected that our method differs from previous glacier extraction methods that focused only on small scale, such as monitoring changes in the Antarctic ice shelf front and melting of the Greenland outlet glaciers (Baumhoer et al., 2019; Zhang et al., 2019), but can be applied to extract complete glacier over a large extent. Unlike traditional glacier extraction methods that depend only on spectrum features, the high-spatial resolution images possess rich information of texture, shape, and spatial distribution of ground objects, which contribute significantly to distinguish categories with similar spectral characteristics (Tong et al., 2020). Based on which, our method can automatically learn the features to distinguish glaciers from terminal moraine lake, thin clouds and snow (Fig. 13 and 15), showing an obvious advantage in glacier extraction in the case of complex spectrum characteristics. Our method yields the continuous glacier boundaries with fewer fine patches, which can reduce the workload for the further post-processing. Furthermore, the adoption of attention mechanism and TTA both improved the effectiveness of our model to extract glaciers in the test set (Fig. 11 and 16).

The method in this paper was not used to extract the debris-or clouds-covered glaciers, in which cases errors may occur. And, features such as frozen rivers were sometimes mistaken for glaciers. Another drawback is that despite the short prediction time of the well-training model, its production took a lot of time due to the lack of readily available glacier datasets based on high-resolution remote sense images.

4.2 The difference between inventories

The amount of glacier resources is critical to regional water resources and future sea level rise (Bolch et al., 2012) and the accurate extraction of glaciers contributes to the exact assessment of ice volume and mass balance as well as the measurement of glacier length (Immerzeel et al., 2010). Existing glacier inventories are generally based on images with low/medium spatial resolution, which may misestimate the global glacier size to some extent, therefore, we discussed the differences between GGI and TPG2017 from the test images, where the former was produced by manual delineation using Landsat TM/ETM+ images in 1999-2003 (Nuimura et al., 2015) and the latter was generated based on Landsat OLI images in 2013-2018. To better explore the differences between the inventories, glaciers were firstly divided into accumulation zone and ablation area based on the median area elevation that is deemed as the material equilibrium line altitude (ELA) which is higher than the actual ELA for some glaciers (e.g., disintegrating glaciers) (Braithwaite and Raper, 2009), but has little effect for this study. The median altitude of glaciers is different in each region, which is 5645 m in the Tanggula Mountains and 5441 m in the Kunlun Mountains (Tab. 4).

Table 4. Summary of glaciers in GGI, TPG2017 and our data.

Image	Region	Attention DeepLab V3+	GGI			TPG2017		
		Area	Area	Difference		Area	Difference	
		km^2	km^2	km^2	%	km^2	km^2	%
Image (1)	Ablation	29.12	33.43	-4.31	-14.80	32.34	-3.22	-11.06
	Accumulation	40.90	36.36	4.54	11.10	42.90	-2	4.89
Image (2)	Ablation	11.09	15.20	-4.11	-37.06	15.46	-4.37	-39.40
	Accumulation	9.65	8.00	1.65	17.10	10.22	-0.57	-5.91
Image (3)	Ablation	21.47	22.59	-1.12	-5.22	23.14	-1.67	-7.78
	Accumulation	21.24	20.54	0.7	3.30	21.38	-0.14	-0.66
Image (4)	Ablation	45.18	49.37	-4.19	-9.27	48.23	-3.05	-6.75
	Accumulation	37.60	32.76	4.84	12.87	39.03	-1.43	-3.80
Total		216.25	218.25	-2.00	-0.92	232.70	-16.45	-7.61

Our extracted glacier area differs from that of GGI by only -2.00 km^2, accounting for -0.92% of the extracted area, while the area difference with TPG2017 is -16.45 km^2, accounting for -7.61% of the extracted area (Tab. 4). In detail, our model extracts a larger glacier area in accumulation zone than GGI, which is mainly due to the omission of the shaded area in the upper glacier by GGI (Nuimura et al., 2015). However, the glacier area in the ablation zone obtained by our model is smaller than in GGI, which is rational because of the retreat of glaciers in the ablation zone under the dramatic warming (Fig. 18a-d). Compared with the results of TGP2017, the glacier areas extracted by our method are smaller in both the ablation and accumulation zones, while the difference is significantly larger in the ablation zone than in the accumulation zone, indicating that the glaciers change is mainly concentrated in the ablation area, which is consistent with the general pattern of glacier changes (Fig. 18e-h). From the perspective of data sources, GGI and TPG2017 were produced by Landsat TM/ETM+/OLI images with resolution of 30 m/15 m. Our data was obtained from Gaofen-6 PMS images with a resolution of 2 m that fewer mixed pixels in the image allow for more detail classification of glaciers and other features as well as the more accurate extraction of glacier boundaries, which leads to the smaller glacier area for the data in this paper (Fig. 17). In general, our method allows a more accurate extraction of the bare intact glaciers, although in a few cases, other features are misclassified as glaciers, but a small amount of manual modification can provide a more exact data for the glacier inventory.

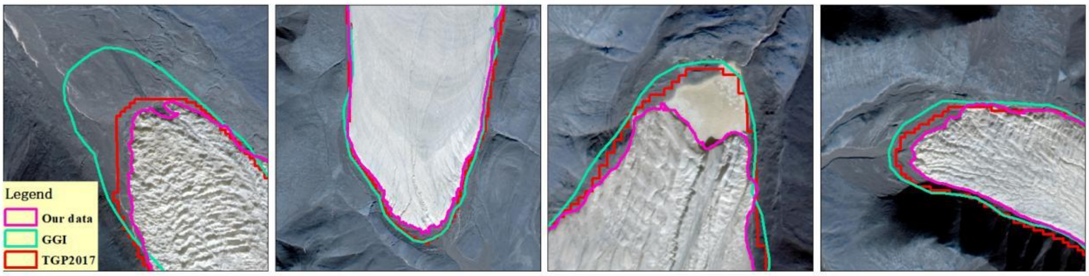


Figure 17. The comparison of different glacier inventories. Our data was obtained in 2020, the GGI was produced in 2002 and the TGP2017 was made in 2017.

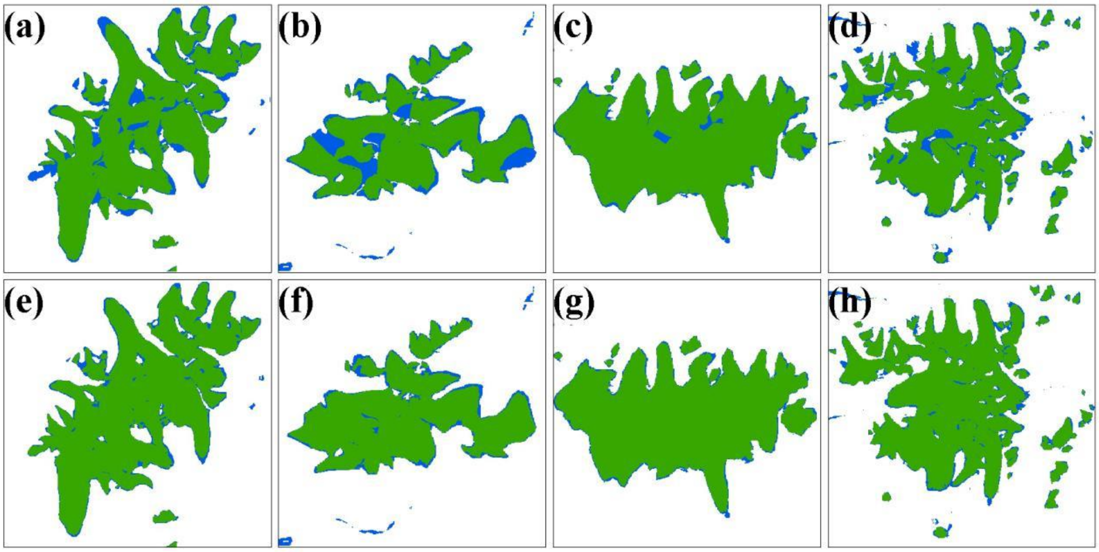


Figure 18. (a)-(d) are the comparison between our result and the GAMDAM glacier inventory, (e)-(h) are the comparison between our result and the TPG2017. The green and blue are the same and different areas in the two datasets, respectively.

5 Conclusion and prospect

In this paper, a glacier extraction method using DeepLab V3+ network with attention mechanism and TTA was proposed to accurately extract glaciers from high-resolution remote sensing images. This method can help to improve the accuracy of the automatically extracted glacier outlines and solve the problem that most high-resolution images cannot extract glacier profiles using traditional methods such as NDSI due to the lack of short-wave infrared band. By comparing with FCN, U-Net and DeepLab V3+, the glacier extracted ability of our model was demonstrated, with the highest OA and Kappa of 99.58% and 0.9915, respectively.

Four scenes of Gaofen-6 PMS image with glaciers were selected to test the ability of extracting complete glacier over large extent. And then comparison with other glacier extracted methods shows that our model has better performance with the OA and kappa of 99.40% and 0.9846, respectively, which could distinguish glacier from terminal moraine lake, thin snow and cloud. Moreover, the glacier boundary obtained from our method was continuous with less fine patches, reducing the workload for post-processing. When comparing the glaciers extracted by our method with the GGI and TPG2017, we found that our data have a more detailed representation of bare ice boundary, which can provide a more accurate data for glacier inventory after manual revision.

In the future, we will make further research and adjustment in the following aspects: (1) improving the algorithm to increase the network's ability to learn glacier features; (2) adding more samples to diversify the training samples and allow the network to learn more features to solve the existing difficulties of extracting debris-covered glaciers; (3) using other high-resolution remote sensing image and SAR image, etc., to compensate for the loss of extraction accuracy of optical images under cloud occlusion; (4) applying transfer learning to reduce the time cost of sample annotation to allow deep learning to be put to glaciers extraction more quickly, thus improving model generalization.

Code and data availability. Gaofen-6 PMS images are available at China High-resolution Earth Observation System (https://www.cheosgrid.org.cn/). The datasets including the GGI and TPG2017 used in this study are freely available. The code for deep learning is available from https://github.com/yiyou101/glacier-extraction.git, and sample datasets of glacier can be provided upon request from the corresponding readers.

Author contributions. XC designed this algorithm of extracting glacier outlines and writing the original draft; XY contributed in terms of supervision and reviewing the manuscript; HD contributed in terms of editing of the manuscript; CC, JL and WP contributed in terms of getting satellite data and processing data.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank the China High-resolution Earth Observation System for providing high resolution images used for this study.

Financial support. This research has been supported by the Project of China Geological Survey (grant no. DD20211570) and the National Natural Science Foundation of China (grant nos. 41861013 and 42071089).

References

Azam, M. F., Wagnon, P., Berthier, E., Vincent, C., Fujita, K., and Kargel, J. S.: Review of the status and mass changes of Himalayan-Karakoram glaciers, J. Glaciol., 64, 61-74, https://doi.org/10.1017/jog.2017.86, 2018.

Badrinarayanan, V., Kendall, A., and Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE T. Pattern Anal., 39, 2481-2495, https://doi.org/10.1109/TPAMI.2016.2644615, 2017.

Baumhoer, C. A., Dietz, A. J., Kneisel, C., and Kuenzer, C.: Automated Extraction of Antarctic Glacier and Ice Shelf Fronts from Sentinel-1 Imagery Using Deep Learning, Remote Sens.-Basel, 11, https://doi.org/10.3390/rs11212529, 2019.

Bishop, M. P., Olsenholler, J. A., Shroder, J. F., Barry, R. G., Raup, B. H., Bush, A. B. G., Copland, L., Dwyer, J. L., Fountain, A. G., Haeberli, W., Kääb, A., Paul, F., Hall, D. K., Kargel, J. S., Molnia, B. F., Trabant, D. C., and Wessels, R.: Global Land Ice Measurements from Space (GLIMS): Remote Sensing and GIS Investigations of the Earth's Cryosphere, Geocarto Int., 19, 57-84, https://doi.org/10.1080/10106040408542307, 2004.

Bolch, T., Kulkarni, A., Kaab, A., Huggel, C., Paul, F., Cogley, J. G., Frey, H., Kargel, J. S., Fujita, K., Scheel, M., Bajracharya, S., and Stoffel, M.: The state and fate of Himalayan glaciers, Science, 336, 310-314, https://doi.org/10.1126/science.1215828, 2012.

Braithwaite, R. J. and Raper, S. C. B.: Estimating equilibrium-line altitude (ELA) from glacier inventory data, Ann. Glaciol., 50, 127-132, https://doi.org/10.3189/172756410790595930, 2009.

Chen, L. C., Zhu, Y. K., Papandreou, G., Schroff, F., and Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, Proceedings of the European Conference on Computer Vision, Munich, Germany, 8-14 September 2018, 801-818, https://doi.org/10.1007/978-3-030-01234-2_49, 2018.

Chollet, F.: Xception: Deep Learning with Depthwise Separable Convolutions, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, 22-25 July 2017, 1251-1258, https://doi.org/10.1109/CVPR.2017.195, 2017.

Girshick, R.: Fast R-CNN, Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 13-16 December 2015, 1440-1448, https://doi.org/10.1109/ICCV.2015.169, 2015.

Grinsted, A.: An estimate of global glacier volume, Cryosphere, 7, 141-151, https://doi.org/10.5194/tc-7-141-2013, 2013.

Guo, W. Q., Liu, S. Y., Xu, J. L., Wu, L. Z., Shangguan, D. H., Yao, X. J., Wei, J. F., Bao, W. J., Yu, P. C., Liu, Q., and Jiang, Z. L.: The second Chinese glacier inventory: data, methods and results, J. Glaciol., 61, 357-372, https://doi.org/10.3189/2015JoG14J209, 2017.

He, K. M., Zhang, X. Y., Ren, S. Q., and Sun, J.: Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770-778, https://doi.org/10.1109/CVPR.2016.90, 2016.

He, Y., Yao, S., Yang, W., Yan, H. W., Zhang, L. F., Wen, Z. Q., Zhang, Y. L., and Liu, T.: An Extraction Method for Glacial Lakes Based on Landsat-8 Imagery Using an Improved U-Net Network, IEEE J-STARS, 14, 6544-6558, https://doi.org/10.1109/jstars.2021.3085397, 2021.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv [preprint], arXiv:1704.04861, 17 Apr 2017.

Huang, B., Zhao, B., and Song, Y. M.: Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery, Remote Sens. Environ., 214, 73-86, https://doi.org/10.1016/j.rse.2018.04.050, 2018.

Immerzeel, W. W., van Beek, L. P., and Bierkens, M. F.: Climate change will affect the Asian water towers, Science, 328, 1382-1385, https://doi.org/10.1126/science.1183188, 2010.

Ji, Q., Dong, J., Liu, R., Xiao, Z. L., and Yang, T. B.: Glacier Changes in Response to Climate Change in the Himalayas in 1990-2015, Scientia Geographica Sinica, 40, 486-496, https://doi.org/10.13249/j.cnki.sgs.2020.03.017, 2020.

King, M. A., Bingham, R. J., Moore, P., Whitehouse, P. L., Bentley, M. J., and Milne, G. A.: Lower satellite-gravimetry estimates of Antarctic sea-level contribution, Nature, 491, 586-589, https://doi.org/10.1038/nature11621, 2012.

Li, X., Cheng, G. D., Jin, H. J., Kang, E., Che, T., Jin, R., Wu, L. Z., Nan, Z. T., Wang, J., and Shen, Y. P.: Cryospheric change in China, Global Planet. Change, 62, 210-218, https://doi.org/10.1016/j.gloplacha.2008.02.001, 2008.

Liu, J., Yao, X. J., Liu, S. Y., Guo, W. Q., and Xu, J. L.: Glacial changes in the Gangdisê Mountains from 1970 to 2016, J. Geogr. Sci., 30, 131-144, https://doi.org/10.1007/s11442-020-1719-6, 2020.

Long, J., Shelhamer, E., and Darrell, T.: Fully Convolutional Networks for Semantic Segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, US, 8-10 June 2015, 3431-3440, https://doi.org/10.1109/CVPR.2015.7298965, 2015.

Marochov, M., Stokes, C. R., and Carbonneau, P. E.: Image classification of marine-terminating outlet glaciers in Greenland using deep learning methods, Cryosphere, 15, 5041-5059, https://doi.org/10.5194/tc-15-5041-2021, 2021.

Milletari, F., Navab, N., and Ahmadi, S.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, 2016 Fourth International Conference on 3D Vision, Stanford, US, 25-28 Oct. 2016, 565-571, https://doi.org/10.1109/3DV.2016.79, 2016.

Nie, Y., Zhang, Y. L., Liu, L. S., and Zhang, J. P.: Monitoring Glacier Change Based on Remote Sensing in the Mt. Qomolangma National Nature Preserve, 1976-2006, Acta Geographica Sinica, 65, 13-28, https://doi.org/10.11821/xb201001003, 2010.

Nuimura, T., Sakai, A., Taniguchi, K., Nagai, H., Lamsal, D., Tsutaki, S., Kozawa, A., Hoshina, Y., Takenaka, S., Omiya, S., Tsunematsu, K., Tshering, P., and Fujita, K.: The GAMDAM glacier inventory: a quality-controlled inventory of Asian glaciers, Cryosphere, 9, 849-864, https://doi.org/10.5194/tc-9-849-2015, 2015.

Oerlemans, J.: Quantifying global warming from the retreat of glaciers, Science, 264, 243-245, https://doi.org/10.1126/science.264.5156.243, 1994.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, Remote Sens. Environ., 148, 42-57, https://doi.org/10.1016/j.rse.2014.02.015, 2014.

Pfeffer, W. T., Harper, J. T., and O'Neel, S.: Kinematic Constraints on Glacier Contributions to 21st-Century Sea-Level Rise, Science, 321, 1340-1343, https://doi.org/doi:10.1126/science.1159099, 2008.

Racoviteanu, A. E., Arnaud, Y., Williams, M. W., and Manley, W. F.: Spatial patterns in glacier characteristics and area changes from 1962 to 2006 in the Kanchenjunga–Sikkim area, eastern Himalaya, Cryosphere, 9, 505-523, https://doi.org/10.5194/tc-9-505-2015, 2015.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, US, June 26th - July 1th 2016, 779-788, https://doi.org/10.1109/CVPR.2016.91, 2016.

Robson, B. A., Bolch, T., MacDonell, S., Hölbling, D., Rastner, P., and Schaffer, N.: Automated detection of rock glaciers using deep learning and object-based image analysis, Remote Sens. Environ., 250, https://doi.org/10.1016/j.rse.2020.112033, 2020.

Robson, B. A., Nuth, C., Dahl, S. O., Hölbling, D., Strozzi, T., and Nielsen, P. R.: Automated classification of debris-covered glaciers combining optical, SAR and topographic data in an object-based environment, Remote Sens. Environ., 170, 372-387, https://doi.org/10.1016/j.rse.2015.10.001, 2015.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv [preprint], arXiv:1505.04597, 18 May 2015.

Schrama, E. J. O., Wouters, B., and Rietbroek, R.: A mascon approach to assess ice sheet and glacier mass balances and their uncertainties from GRACE data, J. Geophys. Res.-Sol. Ea., 119, 6048-6066, https://doi.org/10.1002/2013jb010923, 2014.

Sun, M. P., Liu, S. Y., Yao, X. J., Guo, W. Q., and Xu, J. L.: Glacier changes in the Qilian Mountains in the past half-century: Based on the revised First and Second Chinese Glacier Inventory, J. Geogr. Sci., 28, 206-220, https://doi.org/10.1007/s11442-018-1468-y, 2018.

Telgarsky, M.: benefits of depth in neural networks, in: Proceedings of Machine Learning Research, 29th Annual Conference on Learning Theory, New York , US, 23-26 June 2016, 1517-1539, 2016.

Tong, X. Y., Xia, G. S., Lu, Q. K., Shen, H. F., Li, S. Y., You, S. C., and Zhang, L. P.: Land-cover classification with high-resolution remote sensing images using transferable deep models, Remote Sens. Environ., 237, https://doi.org/10.1016/j.rse.2019.111322, 2020.

Wang, Y. L., Du, W. B., and Wang, S. T.: Extracting glacier information from remote sensing imageries by automatic threshold method of Gaussian mixture model, National Remote Sensing Bulletin, 25, 1434-1444, https://doi.org/10.11834/jrs.20219153, 2021a.

Wang, Z. Q., Zhou, Y., Wang, F. T., Wang, S. X., and Xu, Z. Y.: SDGH-Net: Ship Detection in Optical Remote Sensing Images Based on Gaussian Heatmap Regression, Remote Sens.-Basel, 13, https://doi.org/10.3390/rs13030499, 2021b.

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S.: CBAM: Convolutional Block Attention Module, Proceedings of the European Conference on Computer Vision, Munich, Germany, 8-14 September 2018, 3-19, https://doi.org/10.1007/978-3-030-01234-2_1, 2018.

Yan, L. L. and Wang, J.: Study of Extracting Glacier lnformation from Remote Sensing, Journal of Glaciology and Geocryology, 35, 110-118, 10.7522/j.issn.1000-0240.2013.0013, 2013.

Yao, T. D., Thompson, L., Yang, W., Yu, W. S., Gao, Y., Guo, X. J., Yang, X. X., Duan, K. Q., Zhao, H. B., Xu, B. Q., Pu, J. C., Lu, A. X., Xiang, Y., Kattel, D. B., and Joswiak, D.: Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings, Nat. Clim. Change, 2, 663-667, https://doi.org/10.1038/nclimate1580, 2012.

Ye, Q. H.: Glacier coverage data on the Tibetan Plateau in 2017 (TPG2017, Version1.0), National Tibetan Plateau Data Center [dataset], https://doi.org/10.11888/Glacio.tpdc.270924, 2019.

Ye, Q. H., Zong, J. B., Tian, L. D., Cogley, J. G., Song, C. Q., and Guo, W. Q.: Glacier changes on the Tibetan Plateau derived from Landsat imagery: mid-1970s – 2000–13, J. Glaciol., 63, 273-287, https://doi.org/10.1017/jog.2016.137, 2017.

Yu, F. and Koltun, V.: Multi-scale context aggregation by dilated convolutions, arXiv [preprint], arXiv:1511.07122, 30 Apr 2016.

Zhang, E. Z., Liu, L., and Huang, L. C.: Automatically delineating the calving front of Jakobshavn Isbræ from multitemporal TerraSAR-X images: a deep learning approach, Cryosphere, 13, 1729-1741, https://doi.org/10.5194/tc-13-1729-2019, 2019.

Zhao, X. R., Wang, X., Wei, J. F., Jiang, Z. L., Zhang, Y., and Liu, S. Y.: Spatiotemporal variability of glacier changes and their controlling factors in the Kanchenjunga region, Himalaya based on multi-source remote sensing data from 1975 to 2015, Sci. Total Environ., 745, 140995, https://doi.org/10.1016/j.scitotenv.2020.140995, 2020.