

Detection of Detached Ice-fragments at Martian Polar Scarps Using a Convolutional Neural Network

Shu Su , Lida Fanara, Haifeng Xiao , Ernst Hauber , and Jürgen Oberst

Abstract—Repeated high-resolution imaging has revealed current mass wasting in the form of ice block falls at steep scarps of Mars. However, both the accuracy and efficiency of ice-fragments' detection are limited when using conventional computer vision methods. Existing deep learning methods suffer from the problem of shadow interference and indistinguishability between classes. To address these issues, we proposed a deep learning-driven change detection model that focuses on regions of interest. A convolutional neural network simultaneously analyzed bitemporal images, i.e., pre- and postdetach images. An augmented attention module was integrated in order to suppress irrelevant regions such as shadows while highlighting the detached ice-fragments. A combination of dice loss and focal loss was introduced to deal with the issue of imbalanced classes and hard, misclassified samples. Our method showed a true positive rate of 84.2% and a false discovery rate of 16.9%. Regarding the shape of the detections, the pixel-based evaluation showed a balanced accuracy of 85% and an F1 score of 73.2% for the detached ice-fragments. This last score reflected the difficulty in delineating the exact boundaries of some events both by a human and the machine. Compared with five state-of-the-art change detection methods, our method can achieve a higher F1 score and surpass other methods in excluding the interference of the changed shadows. Assessing the detections of the detached ice-fragments with the help of previously detected corresponding shadow changes demonstrated the capability and robustness of our proposed model. Furthermore, the good performance and quick processing speed of our developed model allow us to efficiently study large-scale areas, which is an important step in estimating the ongoing mass wasting and studying the evolution of the martian polar scarps.

Index Terms—Change detection (CD), convolutional neural network (CNN), ice-fragment, image segmentation, Mars, scarp.

I. INTRODUCTION

METER-SCALE ice block falls have first been discovered by the High-Resolution Imaging Science Experiment (HiRISE) camera onboard NASA's Mars Reconnaissance Orbiter, which can take images at fine scales down to 0.25 m/pixel

Manuscript received 17 October 2022; revised 13 December 2022; accepted 13 January 2023. Date of publication 23 January 2023; date of current version 8 February 2023. The work of Shu Su was supported by China Scholarship Council. (Corresponding author: Shu Su.)

Shu Su, Haifeng Xiao, and Jürgen Oberst are with the Institute of Geodesy and Geoinformation Science, Technical University of Berlin, 10553 Berlin, Germany (e-mail: shu.su@campus.tu-berlin.de; haifeng.xiao@campus.tu-berlin.de; juergen.oberst@tu-berlin.de).

Lida Fanara and Ernst Hauber are with the Institute of Planetary Research, German Aerospace Center (DLR), 12489 Berlin, Germany (e-mail: lida.fanara@dlr.de; ernst.hauber@dlr.de).

Digital Object Identifier 10.1109/JSTARS.2023.3238968

Our code is available at <https://github.com/sue0892/image-segmentation-for-change-detection-pytorch>.

[1]. HiRISE operates on a nearly sun-synchronous orbit and takes images at the same local time of the day [2]. Therefore, HiRISE imagery is very well suited for scientific investigations requiring change detection techniques. Active mass wasting, such as gully activity including erosion and deposition of material [3] and ice block falls [4], [5], [6] can be investigated through change detection. Equatorward-facing steep scarps at the periphery of the martian North Polar Layered Deposits (NPLD) are composed of several-kilometer thick stacks of dusty water ice layers that record martian climate history over millions of years [7]. Due to thermoelastic stresses [8], [9], they experience fracturing that leads to ice block falls [1], [4]. There are two ways to estimate the area of the mass wasting volume: the vacant gap left in the source scarp and the collection of ice blocks at the foot of the scarp. The source regions of block fall events were first mapped with the help of changed shadows' detection by [10], [11]. Fig. 1(a) shows an example of the NPLD scarp in a HiRISE image. The topography shows that the slope of this scarp reaches up to 45° [see Fig. 1(b)]. Obvious fractured slab-like ice-fragments can be seen along the scarp. These ice-fragments can detach from the scarp [one example of which is indicated by the red arrow in Fig. 1(c)]. They rest as ice blocks of different sizes at the underlying Basal Unit [inside the red circle in Fig. 1(d)]. A coarse-to-fine ice blocks' detection approach by considering the illumination properties was proposed in [6], which provides a robust and effective way to identify the ice blocks larger than 0.5 m in diameter. However, some ice blocks break into pieces or even into finer, pulverized material during rolling downslope, leaving no trace in remote sensing images. Hence, the investigation of the sources of ice block falls (i.e., the detached ice-fragments in the source scarp) is a more reliable way to monitor the ongoing activity of the NPLD.

Manual searching and mapping of the detached ice-fragments is very time consuming as we are facing large amounts of data. Conventional computer vision methods may not reach the required efficiency and accuracy [12], [13]. Artificial intelligence not only can reduce the workload of humans, especially for image analysis over large-scale areas [14], but can also achieve satisfactory detection accuracy [15], [16], [17]. In this article, we use a deep learning method to perform change detection in order to extract the detached ice-fragments at the scarps of the NPLD. More specifically, the area of the ice-fragment is automatically mapped by segmentation, which has been widely used in medical image diagnostics [18], [19] and remote sensing image analysis [20], [21], [22], [23]. U-Net is a typical convolutional neural

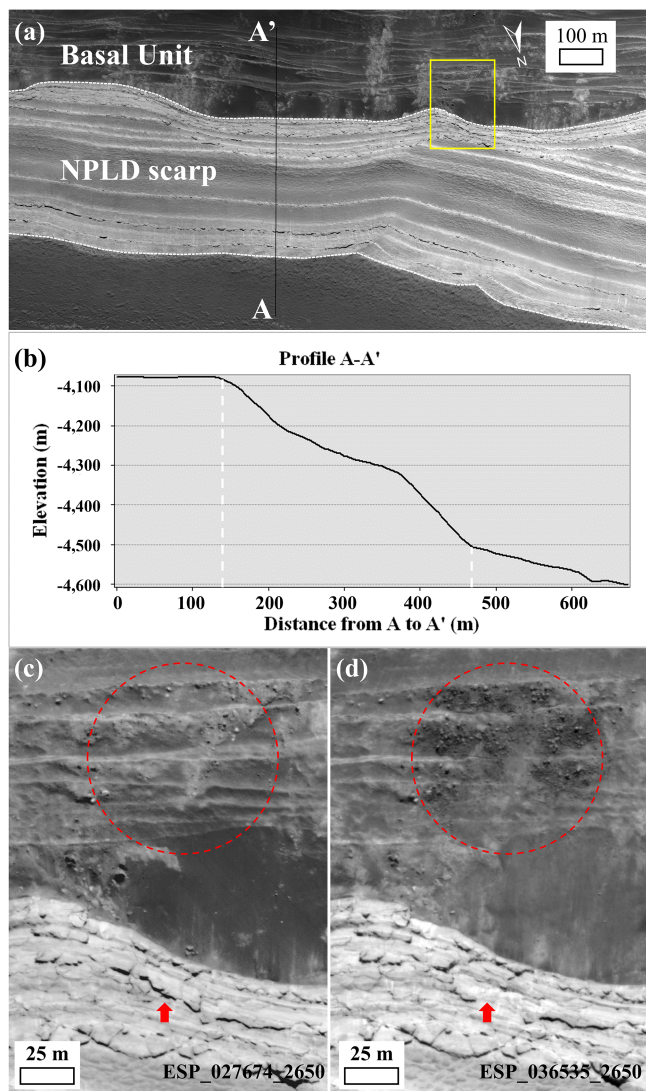


Fig. 1. (a) Red-filter HiRISE image ESP_027674_2650, at 85.0°N, 151.5°E, shows the steep NPLD scarp and its underlying sand-rich Basal Unit. The boundaries of the NPLD scarp are depicted by the white dashed lines. The scarp slope dips toward the top. The yellow box indicates the location of (c) and (d). (b) Elevation profile from A to A', in which the white dashed lines indicate the location of the scarp boundaries in (a). The topography is derived from a HiRISE DTM by [11]. (c) is the predetach image in Mars Year 31. The red arrow points to a fractured ice-fragment at the NPLD steep scarp. (d) is the postdetach image in Mars Year 32. Inside the red circle are the associated ice blocks resting at the foot of the scarp.

network architecture for image segmentation [24]. It is a U-shaped architecture consisting of a specific encoder–decoder scheme: the encoder path reduces the spatial dimensions of feature maps but increases the channels, and the decoder path reconstructs the spatial dimensions and reduces the channels. Since then, improved U-Net models have been developed, such as U-Net++ [25] and U-Net 3+ [26]. Furthermore, modified U-Net models, which substitute the basic forward convolutional units with other convolutional blocks, such as VGG-16 weight layers [27], residual blocks [28], or recurrent convolutional layers [29], have substantially increased the performance of segmentation.

The deep residual network (ResNet) introduces a shortcut connection to handle the degradation problem that the performance of the model decreases while increasing the depth of the network [30], [31]. There are many variants of the ResNet architecture, such as ResNet-34, ResNet-50, and ResNet-101, which have the same concept but with different number of layers. Starting from the 50-layer ResNet, a 3-layer ‘bottleneck’ block was introduced by [30], which contains a stack of 3 layers: 1×1 , 3×3 , and 1×1 convolution layers. ResNet combined with other networks has been widely used in object detection and classification tasks [32], [33], [34]. For example, ResU-Net, which combines ResNet and U-Net, has been used in many medical scenarios, such as identifying organs [35], [36] and detecting medical devices [37], [38]. In addition, in remote sensing, tasks like landslide mapping [39], [40], building extraction [41], [42], and land cover segmentation [43], [44] can also be handled by the ResU-Net.

In the application of change detection, unlike ordinary single image detection, we need to simultaneously feed two or more images covering the same geographical area into the model. The Siamese network is an architecture that contains two or more identical subnetworks to generate and compare feature vectors for each input [45]. It can be applied to different cases, such as detecting duplicates, finding anomalies as well as face recognition [46], [47]. The idea of using same weights while extracting features from bitemporal images is also suitable for change detection.

The attention mechanism has been well used in neural networks to improve the performance of the encoder–decoder scheme [48]. It permits the network to devote more focus on regions of interest so that the most relevant vectors will be attributed the highest weights. Oktay et al. [49] suggested to integrate an attention gate into the U-Net model (attention U-Net), which can improve the prediction performance and preserve computational efficiency as well. Li et al. [50] developed a pyramid attention network, which combines an attention mechanism and a spatial pyramid, to extract precise dense features. Ni et al. [38] proposed an augmented attention module to fuse semantic information in high-level feature maps with global context in low-level feature maps, aiming to learn discriminative features and emphasize key semantic features. He et al. [51] considered using the low-resolution semantic images as prior to guide the attention module to focus on the target of interest. Recent work has shown that adding the attention mechanism to the change detection process can improve the recognition ability of the model [52], [53]. The challenge in our study area is that shadows are much easier to identify than ice-fragments. Adding an attention module into the decoder part can greatly help the model learn to suppress irrelevant regions while highlighting specific objects.

The loss function is a way to measure how well a designed model is in predicting by quantifying the error between the prediction and the ground truth. Various loss functions can be used to handle image segmentation problems, such as cross entropy loss [54], dice loss [55], and focal loss [56]. Cross entropy loss function is a distribution-based loss function aiming to minimize the dissimilarity between the predictive distribution

and the true distribution. It is popular in segmentation tasks due to its stability [57]. However, when facing with serious class imbalance issue, i.e., the number of foreground pixels much smaller than the background pixels, the prediction is heavily biased towards the background. Dice loss, inspired from dice coefficient, measures the relative overlap between the prediction and the ground truth, and is not affected by imbalanced data [55]. The focal loss is an improved version of cross entropy loss to combat the difficulty in detecting hard, misclassified objects [56]. The loss functions behave differently when responding to specific segmentation tasks. When detecting objects that are small in proportion and difficult to identify, a combined loss function may effectively utilize their respective merits.

Our study area only consists of ice. The existing segmentation methods mainly target diverse categories that are distinguishable from each other. The main difficulty of our task is that the ice-fragments are hard to classify even by visual detection, because they are very similar to the background. Furthermore, extracting the detached parts while excluding the changed shadows requires a customized deep learning model. The contributions of our work are as follows.

- 1) Under the Siamese network architecture, the bitemporal images are handled with identical subnetworks in tandem so as to generate respective features of the bitemporal images, and the features of their difference image are also rich with multilevels.
- 2) The augmented attention module takes the features of the difference image into consideration, so that the network puts more weight on the changed area, which is consistent with the essence of change detection.
- 3) A combination of dice loss and focal loss alleviates the issue of imbalanced classes as well as hard, misclassified samples.

The rest of this article is organized as follows. Section II describes the details of our proposed deep learning-driven change detection model as well as the loss function. In Section III, we show the experimental work and our results. Section IV is the discussion of the benefits and limitations of the technique. Finally, Section V concludes this article.

II. METHODOLOGY

A. Deep Learning-Driven Change Detection Model

The overall procedure of our deep learning model for change detection is illustrated in Fig. 2. T1 and T2 are a pair of coregistered HiRISE images showing pre- and postdetach event. They are separately fed into two identical convolutional neural networks that have the same architecture, parameters, and weights. The features extracted from T1 and T2 can be expressed as

$$\begin{aligned} f_{T_1} &= H_f(T_1, w) \\ f_{T_2} &= H_f(T_2, w) \end{aligned} \quad (1)$$

where feature maps $f_{T_1}, f_{T_2} \in R^{H \times W \times C}$, H , W , and C represent the height, width, and channel dimension of the feature map. $H_f(\cdot)$ represents the residual function followed by an activation function. The weight w is mirrored to update during training.

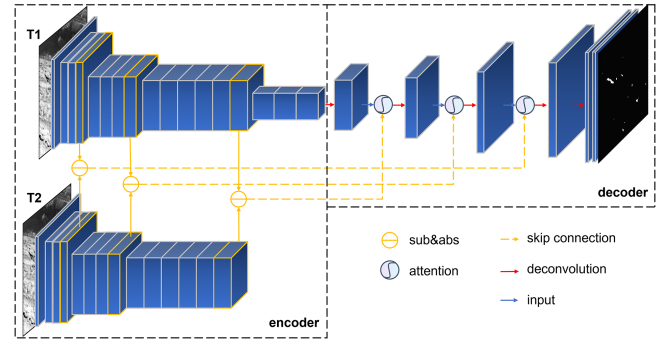


Fig. 2. Illustration of our deep learning-driven change detection model. T1 and T2 are a pair of coregistered two-period images. The black and white binary image (upper right) is the final segmentation result.

We use residual blocks from ResNet-50 as the backbone to extract features from the low level to the high level [30], [31]. Residual networks help to overcome the degradation problem while increasing the depth of the network [30], [31].

In our model, encoding takes five stages to obtain multilevel features, which is exactly applied to T1. And for T2, only the first four stages are required for extracting multilevel features. The reason for this is based on the way T1 and T2 are connected. The first stage is a convolution layer followed by a 3×3 max pooling. Max pooling picks the maximum value from each 3×3 patch to reduce the size of the feature map in order to have fewer parameters in the model while keeping essential features. Stages 2–5 apply 3, 4, 6, and 3 times “bottleneck” block in sequence.

At each stage, we subtract features of T2 from T1

$$d_{T_1-T_2} = |f_{T_1} - f_{T_2}|. \quad (2)$$

The absolute difference image $d_{T_1-T_2}$ will be skip connected into the attention module. The difference image helps guiding the network to focus on the changed area.

After the fifth stage, up-sampling is applied on T1, which goes to the decoder part. We use deconvolution, a mathematical operation that reverses the effect of convolution, to reconstruct the spatial dimension of the image, which is mapping a low dimension to a high dimension while maintaining the connectivity patterns between them. After each deconvolution layer, the absolute difference image $d_{T_1-T_2}$ is skip connected into an augmented attention module, which was introduced by [38] for segmentation on surgical instruments. The attention vector can be computed as follows:

$$V_A = H_2(H_1(G(f_{T_1})) + H_1(G(d_{T_1-T_2}))) \quad (3)$$

where $H_1(\cdot)$ represents a 1×1 convolution with batch normalization followed by the rectified linear unit activation function. $H_2(\cdot)$ represents an 1×1 convolution followed by a softmax activation function. $G(\cdot)$ is the global average pooling, directly applied on f_{T_1} and $d_{T_1-T_2}$ to squeeze the global information into $1 \times 1 \times \text{channels}$ vectors. The equation is as follows:

$$G(x_k) = \frac{\sum_{i=1}^H \sum_{j=1}^W x_k(i, j)}{H \times W} \quad (4)$$

where $1 \leq k \leq C$, C is the number of channels.

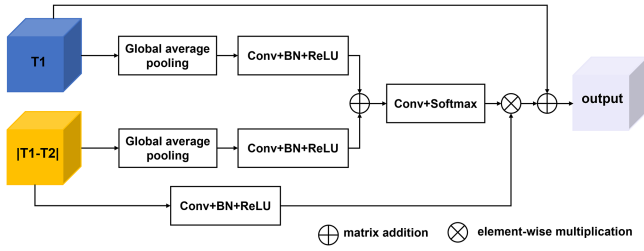


Fig. 3. Architecture of the augmented attention module, which combines T1 and the difference image $|T1-T2|$.

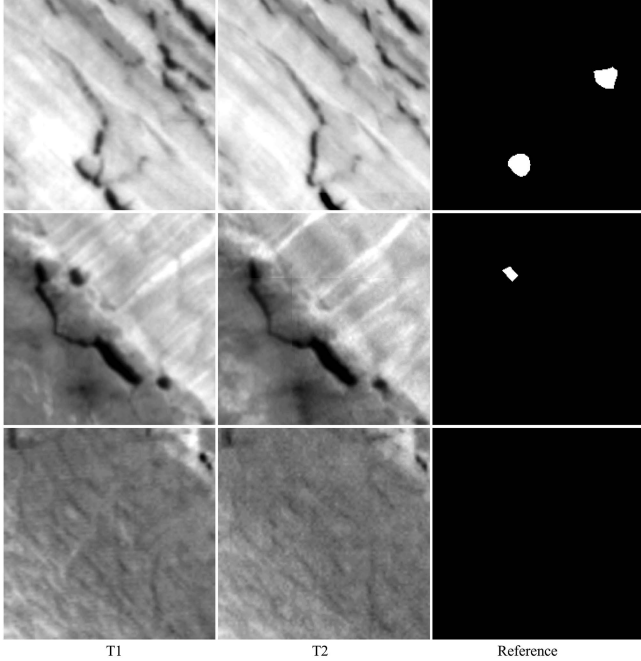


Fig. 4. T1 is the predetach image and T2 is the postdetach image. In the manually labelled images, class 0 is black and class 1 is white.

The final attentive feature map is generated as follows:

$$F(T_1, |T_1 - T_2|) = f_{T_1} + H_1(d_{T_1-T_2}) \otimes V_A. \quad (5)$$

Ni et al. [38] mentioned that the attention module can be flexibly embedded in different networks due to the advantage of using very few parameters. In our attention architecture, we capture the deep features from T1 and emphasize the target features from the difference image [see Fig. 3].

After five deconvolution steps, two convolution layers map the channels to the desired number of classes. A softmax layer transforms the outputs into a normalized probability distribution.

B. Loss Function

We annotate two classes for our segmentation task. Class 1 represents the detached ice-fragments, while Class 0 represents the background including unchanged areas and the changed shadows. However, the foreground class 1 occupies significantly smaller area than the background class 0 [see Fig. 4]. For the whole training data, the foreground class 1 has about 3.4×10^6

pixels, while the background class 0 has about 5.1×10^8 pixels. Therefore, these classes are typically imbalanced. The dice loss, based on the dice coefficient, can handle class imbalanced problems [55]. The dice loss function is formulated as

$$L_{\text{dice}} = -\log(\text{dice score}) = -\log\left(\frac{2|X \cap Y|}{|X| + |Y|}\right) \quad (6)$$

where X and Y refer to pairs of corresponding pixel values of the prediction and the reference, respectively. The value of dice score is between 0 and 1. The negative natural logarithm of the dice score extends the value range from 0 to positive infinity. If the prediction result is significantly different from the reference, the dice score is small and the dice loss value L_{dice} will be infinitely great.

Ice-fragments are harder to identify compared to the surrounding shadows. The focal loss function focuses on learning hard, misclassified samples by down-weighting the loss for easy-classified samples, which is formulated as

$$L_{\text{focal}} = \begin{cases} -\alpha(1-p)^\gamma \log(p), & \text{if } y = 1 \\ -(1-\alpha)(p)^\gamma \log(1-p), & \text{if } y = 0 \end{cases} \quad (7)$$

where p is the model's estimated probability for class 1; α and γ are two hyperparameters that can be tweaked for better performance. Here, we set $\alpha = 0.25$, $\gamma = 2$ based on our experimental experience.

Our final loss function is a combination of the dice loss and the focal loss to alleviate the problems we are facing. The hybrid loss function is formulated as

$$L = \lambda L_{\text{focal}} + (1 - \lambda) L_{\text{dice}} \quad (8)$$

where λ is a weight to balance the contribution of the dice loss and the focal loss. Based on multiple experiments, λ is set to 0.1. The experiment analysis is discussed in Section III-D.

III. EXPERIMENTS AND RESULTS

A. Dataset and Implementation

Our change detection dataset contains 10 510 pairs of coregistered HiRISE image tiles of 256×256 pixels. The training data account for 80%, the validation data account for 10% and the test data account for the rest 10%. The preprocessing of the HiRISE images including ortho-rectification and coregistration has been described in [11]. The tile width and length are set to 256 pixels, typically 64 m, because the vast majority of ice-fragments (the longer length < 100 pixels) are smaller than that. Images have 50% overlap both horizontally and vertically to ensure that the complete ice-fragments can be preserved. Fig. 4 shows three data examples: the first two rows include the detached ice-fragments (class 1) and the third row has only class 0.

Online data augmentation including perspective transformation, slight optical distortion, and randomly changing brightness and contrast have been used to increase the complexity and diversity of our training data. However, operations such as rotating or flipping the images cannot be used for the training data, as these operations change the fixed positional relationship between the ice-fragments and their corresponding shadows in the image. Online augmentation creates different datasets at each

epoch without saving them on the disk, which is more efficient than offline data augmentation [58].

All training data were fed into our deep learning-driven change detection model. We used Adam, an adaptive learning rate optimization algorithm, to optimize the model [59]. The learning rate was initially set to 0.0001 and decayed every epoch by a factor of 0.95. The 10% validation data were used to check the accuracy and calculate the validation loss after each training epoch. We have done 100 epochs of training during experiments and found that the average validation loss did no longer decrease after around 40 epochs. So, typically, our model was trained for a total of 50 epochs. The model that achieved the minimum average loss on the validation data was saved as the best model. We then applied the best model to the test data for the final evaluation of our proposed model.

B. Evaluation Metrics

Two kinds of quantitative assessment will be discussed in the following: pixel-based and event-based metrics. The F1 score and the balanced accuracy are pixel-based metrics, which are widely used for evaluating the performance of segmentation models [60], [61]. They count the number of pixels of each class. The F1 score is a measure of the accuracy of a model, which combines precision and recall into a formula

$$\text{F1 score} = \frac{2}{\text{precision}^{-1} + \text{recall}^{-1}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \quad (9)$$

The balanced accuracy allows for a better judgment of the model's performance with imbalanced classes and is based on recall and specificity

$$\text{Balanced Accuracy} = \frac{\text{recall} + \text{specificity}}{2} = \frac{\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}}{2} \quad (10)$$

where true positives (TP) are the predicted class 1 pixels associated with the reference class 1 pixels, true negatives (TN) are the predicted class 0 pixels associated with the reference class 0 pixels, false positives (FP) are the predicted class 1 pixels not associated with the reference class 1 pixels, and false negatives (FN) are pixels where the reference class 1 has no associated predictions.

The boundaries of the detached ice-fragments are not always visually clear, e.g., the one shown in the second row of Fig. 4 compared to the two in the first row where the boundaries of ice-fragments are clear. So, the manual mapping as well as the model's predictions of the ice-fragments' shape are only approximations. Therefore, we also calculated the event-based true positive rate (TPR) and false discovery rate (FDR) of the detections. They are more suitable of demonstrating the performance of our model than the pixel-based evaluation because although the exact shape is not always visible, the occurrence of an event always is. The true positives (TP) are the number of change detections. The metrics are defined as

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

$$\text{FDR} = \frac{\text{FP}}{\text{TP} + \text{FP}} \quad (12)$$

where the number of correctly detected ice-fragments are TP, the number of falsely detected ones are FP, and the ice-fragments which are not detected are FN. Instead of manually counting event numbers, we use an algorithm to calculate event numbers automatically. Each detection can be considered as a connected component. When the component of the prediction map overlaps with the corresponding component of the reference map, it is counted as one TP. On the contrary, if the detection component has no overlap with the reference map, it is counted as one FP. The rest only presenting on the reference map are counted as FNs, indicating that the model fails to detect these changes. In some cases, only parts of a large ice-fragment are detected, which means that there are multiple detections of the prediction map overlapping with the same reference component. For these cases, these multiple detections are counted as only one TP.

C. Results on Test Data

The test data were used for the final evaluation of the trained model. Three sets of results from our proposed method are visualized in Fig. 5. Our deep learning model is able to detect and delineate the detached ice-fragments, and is resistant to the complex environment and does not pick on other changes such as the shadows of the ice-fragments or the bright linear ice exposures appearing and disappearing as pointed out by white arrows in Fig. 5(c). An example such as the area indicated by the blue arrow in Fig. 5(b) has a clear boundary and is easily extracted. However, in some cases marked by pink arrows in Fig. 5, even our manual identifications are only approximate areas. Therefore, we call our detections the approximate areas of the detached ice-fragments. The examples also show inconsistencies between the predictions and the reference. When a large area of ice-fragment is released from the scarp, the change detection of the detached ice-fragment may be incomplete due to the blurry boundaries. Fig. 5(a) shows a large detached ice-fragment, while the model can only detect two parts of it. The model cannot predict the middle part that has no corresponding shadow as a change. This indicates that the shadows play an important role in helping the machine locate their corresponding ice-fragments. In Fig. 5(c), a small portion of ice (pointed by yellow arrows) was shed, however, our model did not detect it. We speculate that the machine confused it with image distortion.

The quantitative evaluation results on the test data are organized in Table I. The pixel-based evaluation shows that the F1 score for class 1 is 73.2% and the balanced accuracy is 85%. The event-based evaluation shows a true positive rate of 84.2% and a false discovery rate of 16.9%.

D. Experiment Analysis of the Weight λ of the Hybrid Loss Function

The hybrid loss function is a combination of the dice loss and the focal loss to alleviate the problems of imbalanced classes and hard, misclassified samples. Therefore, the weight λ is essential for balancing the dice loss and the focal loss. We varied λ from

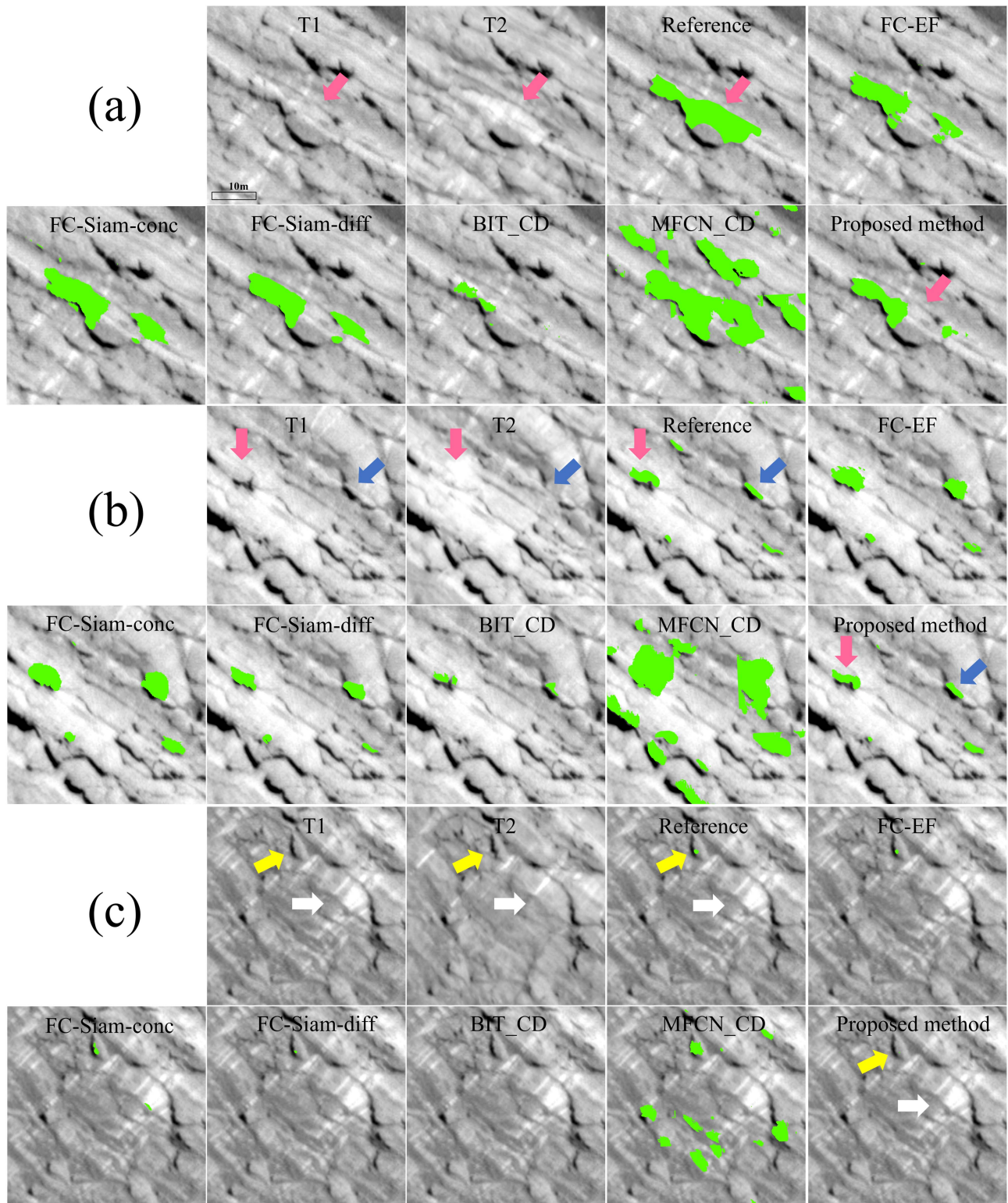


Fig. 5. Visualization results of different change detection methods applied on the test dataset. T1 is the predetach image, T2 is the postdetach image. The reference areas are our manual identifications. The underlying images for the methods are T1, which helps for visualization and localization. The detections are in green. The pink arrows point to hard-to-identify ice-fragments' boundaries, while the blue arrows point to clearly identifiable boundary. The white arrows in (c) point to changes of the bright linear ice exposures. The yellow arrows in (c) point to a small portion of the shed ice.

TABLE I
QUANTITATIVE EVALUATION RESULTS ON THE TEST DATA

Methods	Pixel-based		Event-based	
	F1 score	Balanced Accuracy	TPR	FDR
FC-EF	57.3%	87.8%	82.5%	32.9%
FC-Siam-conc	48.2%	90.7%	96.5%	41.2%
FC-Siam-diff	59.0%	87.9%	84.2%	21.0%
BIT_CD	42.7%	63.8%	45.6%	9.7%
MFCN_CD	12.1%	95.8%	98.2%	89.3%
Proposed method	73.2%	85.0%	84.2%	16.9%

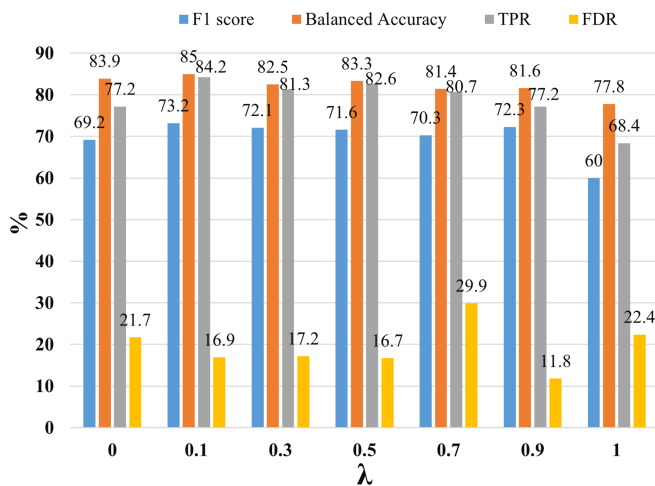


Fig. 6. Effect of different weights λ .

0 to 1 to choose the best performance weight. Note that when λ is 0, loss function is the dice loss; when λ is 1, loss function is the focal loss. The effect of different λ is shown in Fig. 6. When λ is 0.1, our model achieves the highest F1 score for class 1, balanced accuracy and TPR.

E. Ablation Studies

1) *Ablation on Attention Module*: To test if attention module helps improve the model performance, we removed the augmented attention module from our model. The test results show that it gets a TPR of 66.7% and an FDR of 9.8% for the event-based evaluation. For the pixel-based evaluation, it gets an F1 score of 65.2% for class 1 and a balanced accuracy of 76.4%. Without the attention module, the model's ability of picking on the detached ice-fragments is greatly reduced (from 84.2% to 66.7% in TPR). Moreover, both the F1 score and the balanced accuracy drop a lot when not adding the attention module. We thus demonstrate that the attention module improves the network's ability in detecting regions of interest.

2) *Ablation on Loss Function*: To verify the effectiveness of our proposed hybrid loss function, we replace it with three most common segmentation based loss functions: weighted binary cross-entropy (WBCE) loss, dice loss, and focal loss. The comparison results are listed in Table II. Our hybrid loss function, which combines the dice loss and the focal loss achieves the highest F1 score and the lowest FDR. WBCE gets a balanced

TABLE II
QUANTITATIVE EVALUATION RESULTS WITH DIFFERENT LOSS FUNCTIONS

Methods	Pixel-based		Event-based	
	F1 score	Balanced Accuracy	TPR	FDR
WBCE	39.6%	94.4%	100%	56.1%
Dice loss	69.2%	83.9%	77.2%	21.7%
Focal loss	60%	77.8%	68.4%	22.4%
Proposed method	73.2%	85.0%	84.2%	16.9%

accuracy of 94.4% and a TPR of 100%, which shows the ability of detecting all event-based changes. However, it contains large areas of FP that decrease the F1 score (39.6%), and lead to a very high FDR (56.1%), too. The significant deficiency of using WBCE is its inability of excluding the changed shadows [see Fig. 7]. The detections include the whole area of the changed shadows of the detached ice-fragments. The combination of dice loss and focus loss can not only exclude the areas of the changed shadows [see Fig. 7], but also achieve a higher accuracy compared to the case where only one of them is used [see Table II]. This demonstrates the advantage of using our proposed hybrid loss function when facing the issue of imbalanced classes and the hard-to-classify ice-fragments.

F. Comparison to State-of-the-Art Methods

A comparison to five state-of-the-art change detection methods is displayed in Fig. 5. FC-EF, FC-Siam-conc, and FC-Siam-diff were proposed in [45]. These three architecture are all based on the U-Net model. FC-EF takes the concatenation of the bitemporal images as a single input, and then pass the input through the fully convolutional network. FC-Siam-conc and FC-Siam-diff both pass the bitemporal images separately into the Siamese network. But at the decoding part, FC-Siam-conc concatenates both features from the encoding part, while FC-Siam-diff concatenates the absolute difference of the features. BIT_CD was proposed in [21], which introduces a transformer-based model to replace the last convolutional stage of the ResNet architecture. MFCN_CD was proposed in [22] that uses multiscale fully convolutional neural network to learn features of different scales.

Excluding the interference of the changed shadows is one of the important factors to measure the effectiveness of the model. Only BIT_CD and our proposed method can effectively exclude the changed shadows. However, BIT_CD can only detect parts of the changes and has a lot of miss detections. MFCN_CD totally fails in detecting the correct detached ice-fragments. The detections of FC-EF, FC-Siam-conc, and FC-Siam-diff all include parts of the changed shadows. FC-Siam-conc detects the change of the bright linear ice exposure [see Fig. 5(c)].

The quantitative evaluations in Table I show that our proposed method has the highest F1 score. Even though MFCN_CD gets the highest balanced accuracy and TPR, but its FDR is very high (89.3%). According to the visualization results, MFCN_CD not only pick the detached ice-fragments, but also detect large number of unchanged areas. FC-EF, FC-Siam-conc, and FC-Siam-diff show the ability to detect the detached ice-fragments

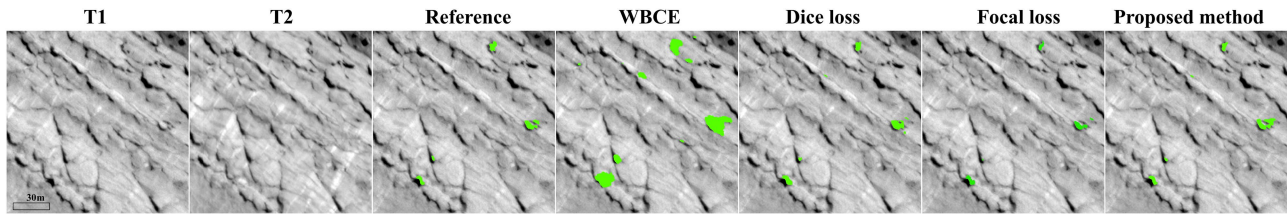


Fig. 7. Visualization results of using different loss functions. T1 is the predetach image, T2 is the postdetach image. The reference areas are our manual identifications. The underlying images for the methods are T1, which helps for visualization and localization. The detections are in green.

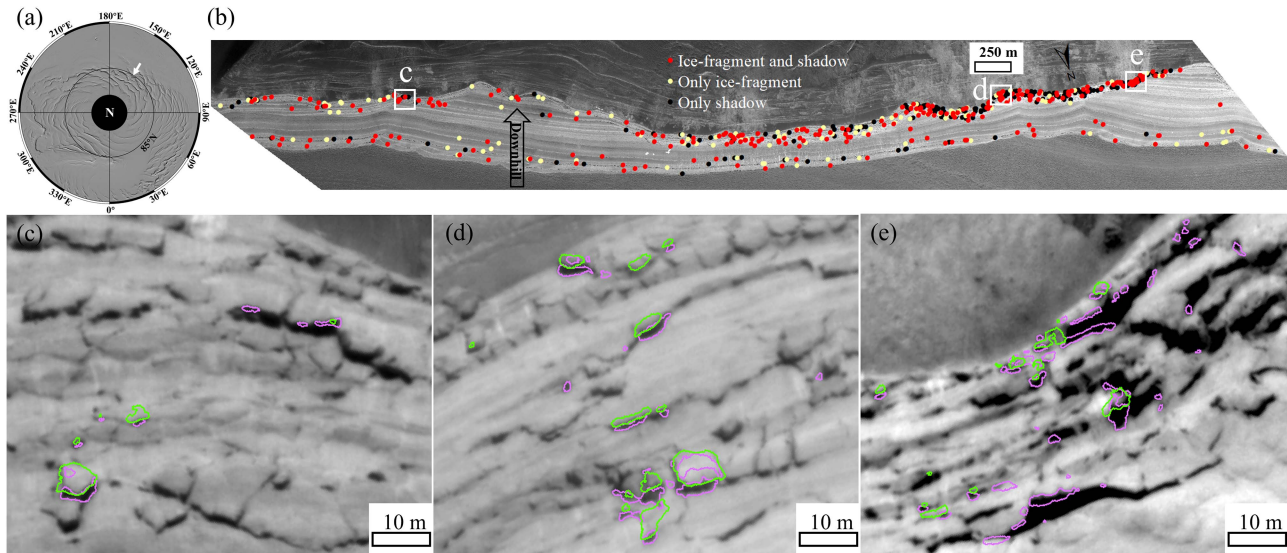


Fig. 8. (a) Context image of the north polar cap is shown by the Mars MGS MOLA Global Shaded Relief (Credit: USGS/NASA). The missing coverage poleward of 87°N is due to the orbital limit of the spacecraft within which no MOLA nadir-pointing profiles have been acquired. The location of the scarp is indicated by the white arrow. (b) Association analysis on the detections of the detached ice-fragments by the proposed deep learning-driven model with the detections of the corresponding shadows of the detached ice-fragments by Su et al. [11]. The beige dots indicate where only the ice-fragments were detected. The red dots indicate where the detached ice-fragments and their corresponding shadows were both detected. The black dots indicate where only the shadows were detected. The background image is the Red-filter HiRISE image PSP_009648_2650 of Mars Year 29. The arrow points to the direction in which the slope descends. (c)–(e) Detailed mapping of the ice-fragments (green) and the shadows (purple), and the inset locations are marked by the white boxes in (b). The background image is the same as (b). The detections (both purple and green areas) in the upper right of (c) are all false positives. In (e), a number of purple areas do not match the full shadow boundaries, which are false positives too.

and also have relatively high balanced accuracy and TPR, but their detections include shadow areas. Our proposed method achieves the basically satisfactory balanced accuracy and TPR, and has a very low FDR except BIT_CD.

G. Application and Association Analysis

We applied our trained model to detect the detached ice-fragments from an NPLD scarp on a pair of full-scene HiRISE images that were taken one Mars year apart: PSP_009648_2650_RED in Mars Year 29 and ESP_018905_2650_RED in Mars Year 30. The location of the scarp is indicated in Fig. 8(a). The size of a single image can reach several Gigabytes, which is a big challenge for manual identification. However, it only took ~ 8 min by using a NVIDIA GeForce GTX 1070 with Max-Q Design to process, orders of magnitude faster than manual work. Su et al. [11] used an automated change detection method to detect the corresponding shadows of the detached ice-fragments in the same scarp region with the same images. The change detection method proposed

by Su et al. [11] can achieve an average true positive rate of $\sim 97.6\%$ in detecting the shadows of the detached ice-fragments. As the detected ice-fragments in this study and the detected shadows in the article by Su et al. [11] are relevant, we use the results of shadow detection to evaluate the detection of the detached ice-fragments. To reduce interference from noise and radiance difference of bitemporal images, both detections of the detached ice-fragments and the shadows, smaller than 5 pixels, were excluded.

In Fig. 8(b), the red dots indicate where the detached ice-fragments and their corresponding shadows were both detected (in total 260), the beige dots indicate where only the ice-fragments were detected (in total 117), and the black dots indicate where only the shadows were detected (in total 186). Fig. 8(c), (d), and (e) shows the detailed mapping of ice-fragments (green) and shadows (purple) from three different parts of the scarp. The ice-fragments and their corresponding shadows distinct from each other, but are closely related. Note the background image is the predetach image. Some purple areas inside of the green areas are not false positives, but the

shadows appearing in the postdetach image. Because when the ice-fragments detach, their surrounding parts may cast new shadows. However, there is a probability of false negatives or false positives in the areas where only shadows or ice-fragments are detected. For example, in the upper right of Fig. 8(c), detections are false positives. This is because of the differing imaging and illumination conditions that caused discrepancies in the same shadow of different periods. The detected parts here are the discrepancies. In Fig. 8(e), a number of purple areas do not match the full shadow boundaries. They are false detections due to the severe geometric deformation of the image here. Considering the randomly selected validation areas in [11], validation may have avoided those areas with image distortion, thus lowering the false discovery rate. Therefore, the overall false discovery rate of shadow detection in [11] may be higher than 9.4%. Among the 186 black dots, the majority are probably the false positives of shadows. We find much less false positives of ice-fragments than shadows, indicating that our deep learning-based change detection method is more robust against image distortion and shadow deformation than the method in [11]. Also, the majority of the 117 beige dots could be the true positives of ice-fragments. Additionally, areas in Fig. 8(d) where only ice-fragments or shadows are detected indicate their corresponding shadows or detached ice-fragments may not have been detected correctly.

IV. DISCUSSION

The results on both test data and full-scene HiRISE application show that our deep learning-driven change detection model is able to automatically detect the detached ice-fragments at the NPLD scarps. To demonstrate that the model can be region- and time-independent, we have chosen HiRISE images from two different tens-of-kilometers-long scarps. Moreover, the time interval between the bitemporal datasets is not limited to one Mars year. Therefore, our model is flexible for short- or long-term change detection.

A feature that is recognizable on the image is the cast shadows of the fractured ice-fragments due to the low-sun conditions in the polar regions of Mars [11]. An experienced human operator is able to distinguish the ice-fragments and their corresponding shadows based on prior knowledge, e.g., the position of the sun and the slope direction of the scarp. However, small size images (256×256 pixels) are the only input to the machine training algorithm. In order to help the machine in recognizing the mutual positional relationship between the ice-fragments and their corresponding shadows, we kept the training data with a constant positional relationship, i.e., the shadows are always below the ice-fragments in the image [e.g., Figs. 4, 5, 7, and 8]. The shadow then helps the machine to locate its corresponding ice-fragment. However, this would be a restriction when training our convolutional neural network. Our augmentation on the training data excludes operations such as random image rotation as well as vertical and horizontal flipping because they would cause the positional relationship to be turned. In another case, we may also not apply the trained model directly to the images, which have an opposite positional relationship between ice-fragments and shadows, i.e., the ice-fragments are below the shadows from top to bottom of the image.

To the best of our knowledge, this is the first attempt to use a deep learning-based method to detect the detached ice-fragments at the martian scarps by comparing multitemporal images. Application of deep convolutional neural networks on a similar task such as landslide recognition has been studied [39], [40], [62]. However, in terms of specific application scenarios and data, it is difficult to make a direct comparison. An association analysis as described in Section III-G can help in such an evaluation. There are several reasons to be confident in the capability and robustness of our deep learning-based method in detecting the detached ice-fragments at the steep scarp areas. First, most of the ice-fragments detected by deep learning correspond well to the shadows detected by Su et al. [11], which achieves an average true positive rate of $\sim 97.6\%$. Second, the number of false positives of ice-fragments is much lower than that of shadows especially at areas with image distortion.

A visual assessment of our model's performance reveals that false detections or undetected changes can be caused by the following. 1) DTM generation is challenging at the steep scarp area, which will sometimes cause ortho-rectification to fail, and thus the bi-temporal images are not well aligned. 2) Indistinct intensity difference between the ice-fragments and the surrounding shadows will create difficulty for the machine to differentiate. See one falsely-detected ice-fragment at upper right of the proposed method image in Fig. 5(a). 3) When a large area of ice-fragment detaches from the scarp, our method may detect only parts of the ice-fragment because there are no shadows to help the model locate their corresponding ice-fragments, or the boundaries of the ice-fragments are blurry. Like the example shown in Fig. 5(a), our method cannot predict the middle part that has no corresponding shadow as a change. It is not very common to have a very large area of ice-fragment detaching from the scarp at one time. If it happens, our model has the ability to detect at least parts of these large ice-fragments with clear boundaries. The undetected parts reduce the pixel-based accuracy of our method, but do not affect the event-based evaluation as long as the method can detect parts of the detached ice-fragments. 4) If a small portion of ice is shed from an ice-fragment [e.g., the false negative in the proposed method image of Fig. 5(c)], it is difficult for the machine to distinguish whether it is the missing portion or the ice-fragment's deformation caused by the image distortion. The first factor could cause wrong detections when the images are misaligned. Better change detection results can be obtained with careful preprocessing of high-quality HiRISE imagery. Errors introduced by factors 2–4 are the shortcomings of our deep learning model. These false and miss detections can be accepted as they occupy a small percentage compared to the area of all detections.

When further probing into the calculation of ice-fragments' volume, we need to consider the uncertainties mentioned previously. Another general problem for accurate volume calculation is the blurry boundaries of the ice-fragments. As we have mentioned before, it is hard to map the accurate boundaries of some ice-fragments. In some cases, the break lines, where the ice-fragments break from their main body (i.e., the steep scarp), are difficult to be identified in the images. Especially in images taken multiple Mars years apart, the break line may be obscured by geological processes. This is a real limitation even

when visually mapping, let alone automatic machine detection. However, it is worthy to note that the shape of all detached ice-fragments can be roughly determined by comparing the pre- and postdetachment images as well as considering the size of the cast shadows of the fractured ice-fragments. Therefore, the size of the detections will not be extremely larger than the actual size.

Although the accurate boundaries of the detached ice-fragments are not always visible, the occurrence of these detachment events are. So, we believe that the event-based evaluation is more convincing. Evaluation based on pixels has restrictions for these blurry shape of ice-fragments. In order to address the pixel-based evaluation problem, in our future work we will divide assessment into two categories: ice-fragments with/without clear boundaries. Then, the final pixel-based evaluation will combine these two assessment results together. This may help to assess the accuracy of our predictions more precisely.

V. CONCLUSION AND OUTLOOK

In this article, a deep learning-driven change detection model is proposed to automatically detect detached ice-fragments at the steep scarps of the NPLD on Mars. We use a U-Net convolutional neural network architecture, which integrates both the ResNet-50 to extract features and an augmented attention module to highlight the target, i.e., the detached ice-fragment. The bitemporal images are fed into the Siamese network to mine their respective information. A hybrid loss function based on dice loss and focal loss is introduced to deal with the issue of imbalanced classes as well as hard, misclassified samples. Test results show that our change detection model is capable of localizing and mapping the changed areas, achieving an F1 score of 73.2% for the detached ice-fragments' class, a balanced accuracy of 85%, a true positive rate of 84.2%, and a false discovery rate of 16.9%. Compared to five state-of-the-art change detection methods, our model is more robust in extracting the approximate areas of the changed ice-fragments while excluding other changes on the images, and is more resistant to the complex topography of the NPLD scarps and even slight image distortions. An association analysis of our detection of the detached ice-fragments with a previous detection of the corresponding shadows demonstrates the capability and robustness of our deep learning-based model.

Fast processing speed and automation demonstrate the potential to apply this method across the whole NPLD area, and even to terrestrial mass wasting. The shape of the detached ice-fragments is an important parameter for estimating the flux and volume of ongoing mass wasting and studying the dynamic evolution of the NPLD scarps. From another perspective, the avalanches have been investigated visually for those fractured NPLD scarps which display ice block fall deposits at their base [7], [63]. We will extend our deep learning method to automatically detect active avalanches, to help reduce human work and complete an automated monitoring pipeline of this area [5], [64]. Ice block falls and avalanches are the main mass wasting activity of the NPLD scarps. Monitoring and investigation of long-term mass wasting over the whole NPLD scarps will provide insights into ice behavior, supporting modeling studies of martian viscous flow velocity [65], thermoelastic stress [9], and climate change.

ACKNOWLEDGMENT

The authors would like to thank HiRISE data, the open-source NASA Ames Stereo Pipeline, and the USGS Integrated Software for Imagers and Spectrometers and also like to thank eight anonymous reviewers for providing valuable and constructive reviews, which help to improve this contribution. We acknowledge support by the Open Access Publication Fund of TU Berlin.

REFERENCES

- [1] K. E. Herkenhoff, S. Byrne, P. S. Russell, K. E. Fishbaugh, and A. S. McEwen, "Meter-scale morphology of the north polar region of Mars," *Science*, vol. 317, no. 5845, pp. 1711–1715, 2007.
- [2] M. J. Brennan and M. D. D. Johnston, "Mars reconnaissance orbiter landing site reconnaissance capability," in *Proc. IEEE Aerosp. Conf.*, 2017, pp. 1–13.
- [3] J. Raack et al., "Present-day gully activity in Sisyphi Cavi, Mars—Flow-like features and block movements," *Icarus*, vol. 350, 2020, Art. no. 113899.
- [4] L. Fanara, K. Gwinner, E. Hauber, and J. Oberst, "Present-day erosion rate of north polar scarps on Mars due to active mass wasting," *Icarus*, vol. 342, 2020, Art. no. 113434.
- [5] L. Fanara, K. Gwinner, E. Hauber, and J. Oberst, "Automated detection of block falls in the north polar region of Mars," *Planet. Space Sci.*, vol. 180, 2020, Art. no. 104733.
- [6] Z. Song, S. Liu, and Y. Jin, "An automatic approach for ice-block falls detection in the martian north polar region," *Proc. SPIE*, vol. 122, 2022, Art. no. 12267.
- [7] C. M. Dundas et al., "Active Mars: A dynamic world," *J. Geophys. Res. Planets*, vol. 126, 2021, Art. no. 8.
- [8] S. Byrne, "The polar deposits of Mars," *Ann. Rev. Earth Planet. Sci.*, vol. 37, pp. 535–560, 2009.
- [9] S. Byrne et al., "Mars polar cliffs: Stressed out and falling apart," *Eur. Planet. Sci. Conv.*, vol. 11, 2017, Art. no. 936.
- [10] S. Su, L. Fanara, X. Zhang, E. Hauber, K. Gwinner, and J. Oberst, "Detecting the sources of ice block falls at the martian north polar scarps by analysis of multi-temporal HiRISE imagery," *Int. Arch. Photogrammetry Sens. Spatial Inf. Sci.*, vol. XLIII-B3-2021, pp. 673–678, 2021.
- [11] S. Su, L. Fanara, X. Zhang, E. Hauber, K. Gwinner, and J. Oberst, "Searching for the sources of ice block falls at the martian north polar scarps," *Icarus*, vol. 390, 2022, Art. no. 115321.
- [12] A. Singh, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 1989.
- [13] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, "Learning tensor low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 679–691, Jan. 2023.
- [14] Q. He, X. Sun, W. Diao, Z. Yan, D. Yin, and K. Fu, "Transformer-induced graph reasoning for multimodal semantic segmentation in remote sensing," *ISPRS J. Photogrammetry Remote Sens.*, vol. 193, pp. 90–103, 2022.
- [15] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.
- [16] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.
- [17] Q. He, X. Sun, Z. Yan, B. Li, and K. Fu, "Multi-object tracking in satellite videos with graph-based multitask modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 5619513.
- [18] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: A review," *Artif. Intell. Rev.*, vol. 54, no. 1, pp. 137–178, 2021.
- [19] R. Yang and Y. Yu, "Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis," *Front. Oncol.*, vol. 11, 2021, Art. no. 573.
- [20] T. Bao, C. Fu, T. Fang, and H. Huo, "PPCNET: A combined patch-level and pixel-level end-to-end deep network for high-resolution remote sensing image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1797–1801, Oct. 2020.
- [21] H. Chen, Z. Qi, and Z. Shi, "Remote sensing image change detection with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2021, Art. no. 5607514.

- [22] X. Li, M. He, H. Li, and H. Shen, "A combined loss-based multiscale fully convolutional network for high-resolution remote sensing image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jul. 2021, Art. no. 8017505.
- [23] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Syst. Appl.*, vol. 69, 2021, Art. no. 114417.
- [24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, 2015, pp. 234–241.
- [25] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, 2018, pp. 3–11.
- [26] H. Huang et al., "UNet 3+: A full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 1055–1059.
- [27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, Apr. 2015.
- [28] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual Unet," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
- [29] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imag.*, vol. 6, no. 1, 2019, Art. no. 014006.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [31] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.
- [32] V. T. Bickel, C. Lanaras, A. Manconi, S. Loew, and U. Mall, "Automated detection of lunar rockfalls using a convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3501–3511, Jun. 2018.
- [33] V. T. Bickel, S. J. Conway, P. A. Tesson, A. Manconi, S. Loew, and U. Mall, "Deep learning-driven detection and mapping of rockfalls on Mars," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2831–2841, May 2020.
- [34] G. Cheng, G. Wang, and J. Han, "ISNet: Towards improving separability for remote sensing image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2022, Art. no. 5623811.
- [35] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted Res-Unet for high-quality retina vessel segmentation," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ.*, 2018, pp. 327–331.
- [36] J. Zhang, Z. Jiang, J. Dong, Y. Hou, and B. Liu, "Attention gate ResU-Net for automatic MRI brain tumor segmentation," *IEEE Access*, vol. 8, pp. 58533–58545, 2020.
- [37] Z. L. Ni, G. B. Bian, X. L. Xie, Z. G. Hou, X. H. Zhou, and Y. J. Zhou, "RASNet: Segmentation for tracking surgical instruments in surgical videos using refined attention segmentation network," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2019, pp. 5735–5738.
- [38] Z. L. Ni et al., "RAUNet: Residual attention U-Net for semantic segmentation of cataract surgical instruments," in *Proc. Int. Conf. Neural Inf. Process.*, 2019, pp. 139–149.
- [39] W. Qi, M. Wei, W. Yang, C. Xu, and C. Ma, "Automatic mapping of landslides by the ResU-Net," *Remote Sens.*, vol. 12, no. 15, 2020, Art. no. 2487.
- [40] N. Prakash, A. Manconi, and S. Loew, "Mapping landslides on EO data: Performance of deep learning models vs. traditional machine learning models," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 346.
- [41] S. Chen, Q. Zuo, and Z. Wang, "Semantic segmentation of high resolution remote sensing images based on improved ResU-Net," in *Proc. Int. Conf. Pioneering Comput. Scientists, Engineers Educators*, 2021, pp. 303–313.
- [42] M. Dixit, K. Chaurasia, and V. K. Mishra, "Dilated-ResUNet: A novel deep learning architecture for building extraction from medium resolution multi-spectral satellite imagery," *Expert Syst. Appl.*, vol. 184, 2021, Art. no. 115530.
- [43] V. Pollatos, L. Kouvaras, and E. Charou, "Land cover semantic segmentation using ResUNet," in *Proc. AI Natural Sci. Technol. Workshop SETN*, Oct. 2020.
- [44] S. Yoo, J. Lee, M. G. Farkoushi, E. Lee, and H. G. Sohn, "Automatic generation of land use maps using aerial orthoimages and building floor data with a conv-depth block (CDB) ResU-Net architecture," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 107, 2022, Art. no. 102678.
- [45] R. C. Daudt, B. L. Saux, and A. Boulch, "Fully convolutional Siamese networks for change detection," in *Proc. 25th IEEE Int. Conf. Image Process.*, 2018, pp. 4063–4067.
- [46] L. Song, D. Gong, Z. Li, C. Liu, and W. Liu, "Occlusion robust face recognition based on mask learning with pairwise differential Siamese network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 773–782.
- [47] D. Chicco, "Siamese neural networks: An overview," *Artif. Neural Netw.*, vol. 2190, pp. 73–94, 2021.
- [48] D. He, Q. S. Shi, X. Liu, Y. Zhong, and L. Zhang, "Generating 2m fine-scale urban tree cover product over 34 metropolises in China based on deep context-aware sub-pixel mapping network," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 106, 2022, Art. no. 102667.
- [49] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, Amsterdam, The Netherlands, Apr. 2018.
- [50] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," 2018, *arXiv:1805.10180*.
- [51] D. He, Q. Shi, X. Liu, Y. Zhong, and X. Zhang, "Deep subpixel mapping based on semantic information modulated network for urban land use mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10628–10646, Dec. 2021.
- [52] Z. Li, C. Yan, Y. Sun, and Q. Xin, "A densely attentive refinement network for change detection based on very-high-resolution bitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4409818.
- [53] X. Peng, R. Zhong, Z. Li, and Q. Li, "Optical remote sensing image change detection based on attention mechanism and image difference," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7296–7307, Sep. 2020.
- [54] J. S. Bridle, "Probabilistic interpretation of feedforward classification network outputs with relationships to statistical pattern recognition," *Neurocomputing*, vol. 3, pp. 227–236, 1990.
- [55] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3-D Vis.*, 2016, pp. 565–571.
- [56] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
- [57] S. Jodon, "A survey of loss functions for semantic segmentation," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol.*, 2020, pp. 1–7.
- [58] Z. Tang, Y. Gao, L. Karlinsky, P. Sattigeri, R. Feris, and D. Metaxas, "OnlineAugment: Online data augmentation with less domain knowledge," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 313–329.
- [59] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, Dec. 2014.
- [60] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li, "Dice loss for data-imbalanced NLP tasks," *CoRR*, pp. 1–13, Nov. 2019.
- [61] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models assessment over imbalanced data sets," *J. Inf. Eng. Appl.*, vol. 3, 2013, Art. no. 10.
- [62] W. Shi, M. Zhang, H. Ke, X. Fang, Z. Zhan, and S. Chen, "Landslide recognition by deep convolutional neural network and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 4654–4672, Jun. 2020.
- [63] P. Becerra, C. Hery, A. Valantinas, S. Byrne, N. Thomas, and S. Conway, "Avalanches of the martian north polar cap," *Europlanet Sci. Congr.*, vol. 14, 2020, Art. no. EPSC2020-1093.
- [64] O. Martynchuk, L. Fanara, K. Gwinner, E. Hauber, and J. Oberst, "Automated NPLD scarp mapping for the block falls monitoring pipeline," *Europlanet Sci. Congr.*, vol. 15, Jun. 2021, Art. no. EPSC2021-712.
- [65] M. M. Sori, S. Byrne, C. W. Hamilton, and M. E. Landis, "Viscous flow rates of icy topography on the north polar layered deposits of Mars," *Geophys. Res. Lett.*, vol. 43, no. 2, pp. 541–554, 2016.



Shu Su was born in 1992. She received the M.S. degree in surveying and mapping engineering from Tongji University, Shanghai, China, in 2018. She is currently working toward the Ph.D. degree in planetary geodesy with the Chair of Planetary Geodesy, Institute of Geodesy and Geoinformation Science, Technical University of Berlin, Berlin, Germany.

Her research interests include digital image processing, deep neural networks for image segmentation, change detection of planetary surface, and large-scale remote sensing image analysis.



Lida Fanara received the Ph.D. degree in planetary geodesy from Technical University of Berlin, Berlin, Germany, in 2022.

She is currently a Surveying Engineer. She is a young researcher, currently with the Planetary Geodesy Department, German Aerospace Center (DLR), Berlin, Germany. She is interested in the detection of planetary surface features and change detection with machine learning. Her research interest includes image analysis for planetary research.



Ernst Hauber is currently a Geologist and studies surface processes on terrestrial planets. He is with the Institute of Planetary Research, German Aerospace Center (DLR), Berlin, Germany. His research interests include endogenic processes on Mars and Mercury, permafrost and glacial processes on Mars and Earth, and analogue fieldwork in polar regions on Earth.



Haifeng Xiao received the M.S. degree in surveying and mapping engineering from Tongji University, Shanghai, China, in 2017. He is currently working toward the Ph.D. degree in planetary geodesy with the Chair of Planetary Geodesy, Institute of Geodesy and Geoinformation Science, Technical University of Berlin, Berlin, Germany.

His research interests include planetary laser altimetry, martian seasonal polar caps, body tidal deformation of Moon and Mercury, planetary rotation, interior structures, and evolution.



Jürgen Oberst received the Ph.D. degree in planetary geodesy from The University of Texas at Austin, Austin, TX, USA, 1989.

He is currently a faculty member and Chair of Planetary Geodesy with the Institute of Geodesy and Geoinformation Sciences, Technical University Berlin, Berlin, Germany. In his research, he works on various aspects of planetary geodesy and geophysics, using data from Laser altimeters and cameras. His particular interest is on planet Mars. He has been a long-standing Co-Investigator in the team of the

High-Resolution Stereo Camera on Mars Express.