Anonymous Referee #1

General Comments

This paper presents the application and study of emerging machine learning techniques towards automatic calving front detection. Specifically, it utilizes a deep neural network architecture, U-Net, to automatically segment raw SAR imagery along calving fronts into digitized vectors. This study focuses on Jakobshavn from 2009-2015, and performs analysis using additional data products to cross-validate the results. The analysis correlates and validates data from the Greenland Ice Sheet CCI project, Bed Machine v3 bedrock data, and the automatically determined calving fronts from this paper. Images to describe the study, and accompanying data tables, help communicate the work done.

The paper is well written and covers a novel emerging technique (deep learning in the cryosphere). Therefore, I would like to recommend it for publication. However, I do have two concerns, though it may not be within the scope of this paper. These concerns regard the paper's wider implications/context, and may impact the rigor/novelty/impact of this study.

We highly appreciate the reviewer for the constructive comments which have significantly improved the quality of our manuscript. We have made our best effort to revise the manuscript based on the referee's comments and suggestions.

The first concern relates to existing similar work conducted by Mohajerani, Y., et. al., in Remote Sensing. Please refer to their paper here: https://www.mdpi.com/2072-4292/11/1/74/htm. While the methods are no doubt similar (deep-learning UNet), the one covered in this paper seems to be more accurate and more comprehensively analysed, though by virtue of being more focused in scope. For comparison, this paper covers Jakobshavn, TerraSAR-X, while Mohajerani covers Jakobshavn, Sverdrup, Kangerlussuaq, Helheim, Landsat 8 in Mohajerani's paper. I think it is helpful to have corroborating evidence of the validity of this methodology - especially published in The Cryosphere. Regardless, while I can still make my recommendation, I will leave others to discuss this matter.

We have added a new subsection 7.1 titled **Differences from the previous work** to discuss the differences between our work and the method of Mohajerani et al. (2019), which are summarized as follows:

- Different strategies are used to classify calving fronts. Our study classifies the surface into two types (i.e., ice mélange and non-ice mélange) to extract the calving front; Mohajerani et al. (2019) use semantic segmentation to extract the front without classifying the surrounding surfaces.
- Additional manual practices such as finding a rotation angle for each glacier are needed in the work of Mohajerani et al. (2019).
- We subdivide the images into small patches, which allows us to use images with high resolutions and various size (i.e., TerraSAR-X images). Mohajerani et al. (2019) resampled images to a fixed size (240 by 152 pixels) with low spatial resolution (49.0 to 88.1 meters).

The second concern I have relates to the generalizability of the network. While I acknowledge this is not the focus of the case study, the following are some questions I, and perhaps others, would express interest in knowing.

The generalizability of the network relies on the diversity of the training examples. With additional training examples, our method can be applied to other places using multi-sensor remote sensing datasets. Moreover, optical images with low cloud cover and Landsat 7 images with scan line errors can be used as long as the calving fronts are visually clear. See our replies to the specific comments below for more details. We did not include the results at another other domains or the results using other remote sensing datasets since they are preliminary and beyond the scope of this manuscript.

Specific Comments

Page 7 Line 1 - It was mentioned that summer imagery has higher performance than winter imagery. Though the ice melange has similar texture to glacial ice, should it not be possible for further training to be performed to close this gap? Perhaps the network needs additional capacity to handle this differentiation?

It is possible to close this gap by including more winter training examples. The accuracy of the well-trained network relies on the quality of the training examples. Delineating calving fronts in winter images with blur boundaries is challenging, and therefore the quality of winter training examples is not as good as those in summers. Including more winter training examples could make the trained network more robust and therefore mitigate the problem caused by winter data quality. However, due to the quota limitation, we only have 159 TerraSAR-X images. Therefore, we did not close this gap in the current work. Note that we did not include the discussion about the possibility to close the gap since it is beyond the scope of this manuscript.

Page 15 Line 6-8 - It is mentioned that this methodology can be applied to other domains. Do you have any analyses on how the network performs on other glacial domains, such as Sverdrup, or Helheim?

We conducted a preliminary experiment by directly applying the network generated from this work as trained by TerraSAR-X imagery from Jakobshavn to Helheim (without including any new training data). Figure R1 shows that the automatically delineated calving front at Helheim is very close to what one would get from visual inspection. Therefore, our method can be applied to other glaciers. Of course, we need to include more training examples from more glaciers to ensure reliable results on other glacier domains.
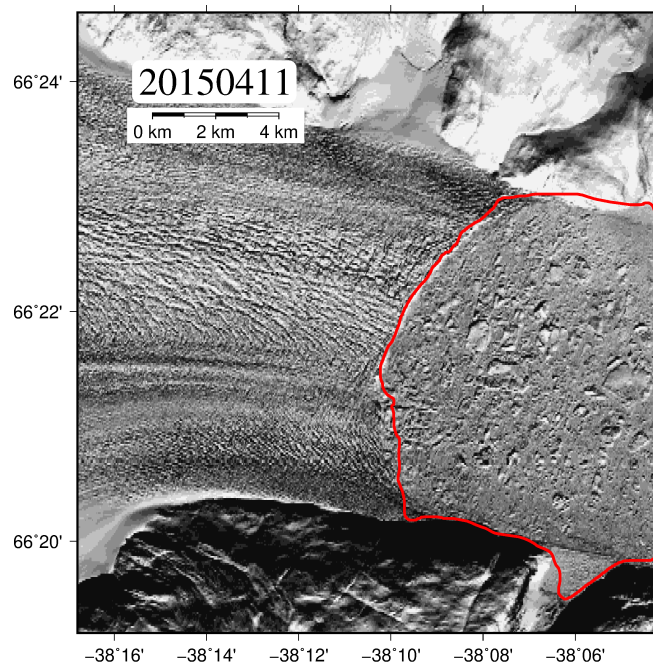
Figure R1. An example of automatically delineated calving front at Helheim. The background image is a Landsat 8 image taken on April 11th, 2015. The red line indicates the automatically delineated calving front.

Page 2 Line 29 - Does this network rely on features only visible at 3.3-3.5m? i.e., does lowering the pixel resolution adversely affect accuracy/performance? -Similarly, can the network handle lower resolution 30/60m datasets like Landsat?

This network does not rely on high-resolution images. As long as the calving front is visually clear, the network is able to handle images with different resolutions and sizes. For example, with additional training, the network can generate reasonable results using lower resolution image such as Landsat, as shown in Figure R2. Note that training dataset used to train this network does not include the image in Figure R2.
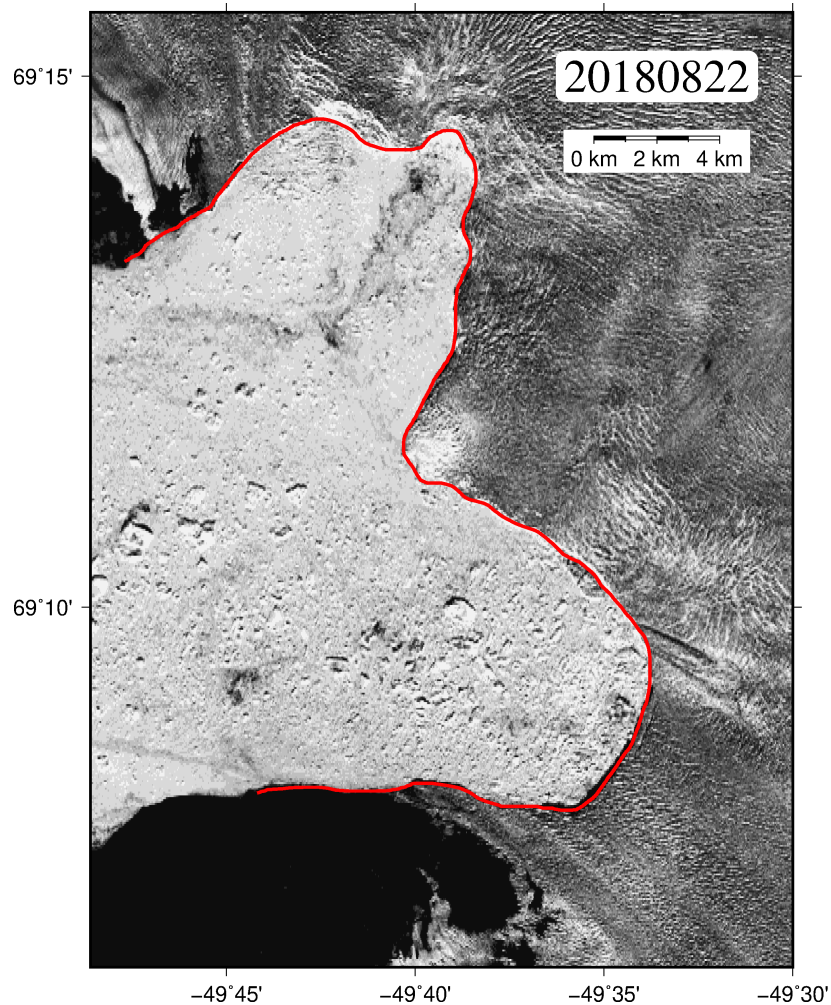
Figure R2. An example of automatically delineated calving front at Jakobshavn Isbrae using a Landsat 8 image taken on August 22nd, 2018.

Page 2 Line 34 - It is mentioned that the cloud cover issue is avoided. However, some light cloud cover does not always obscure calving front edges. Would it be feasible to train the network to handle these issues, to allow greater temporal resolution/constraints by not eliminating minor cloud covered images from the study? By extension, could the network handle Landsat 7 scan line errors, given additional training?

With additional training, it is feasible to train the network to handle the issues if cloud does not obscure calving fronts on an image. We conducted some experiments on Landsat 8 images with light cloud cover. Figure R3 shows that the results of a Landsat 8 image with light cloud cover are reasonable. This Landsat 8 image is not in the training dataset.
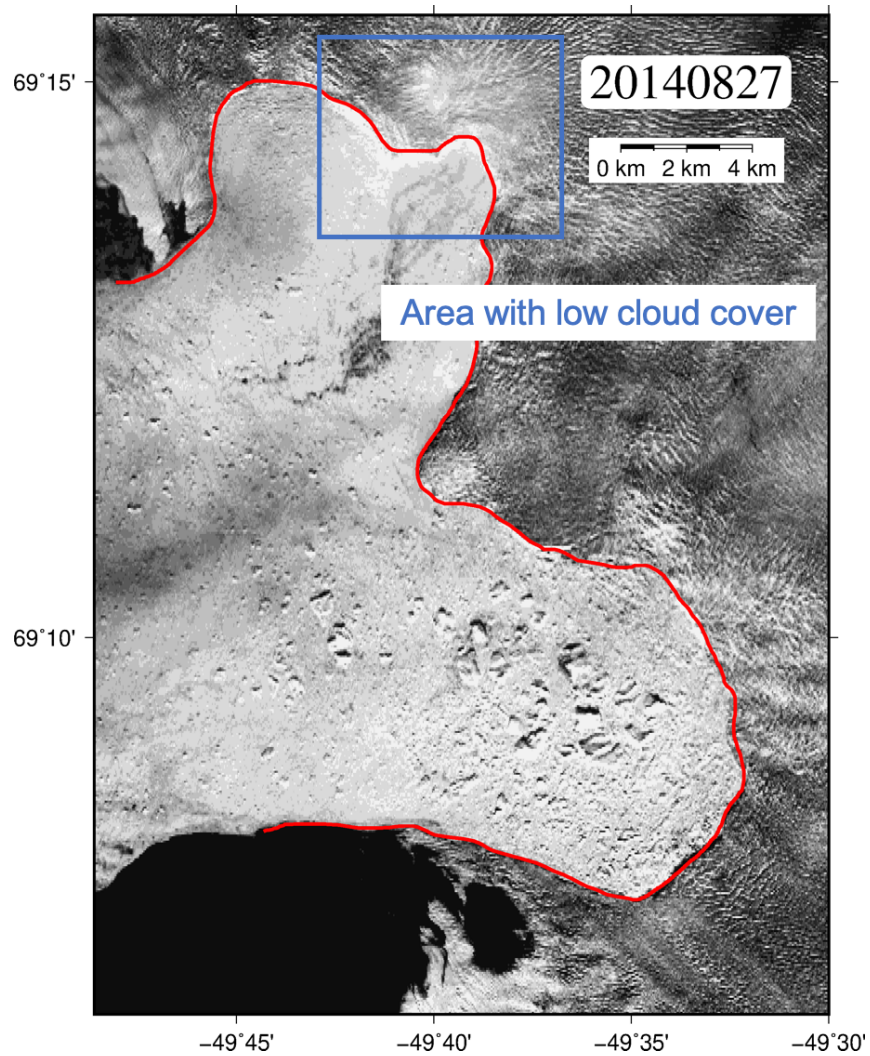
Figure R3. An example of automatically delineated calving front at Jakobshavn Isbrae using a Landsat 8 image with clouds. The image was taken on August 27th, 2014. The blue box indicates an area with low cloud cover.

The network could also handle Landsat 7 images with scan line errors (Figure R4) with additional training. Note that we have included this image in the training dataset.
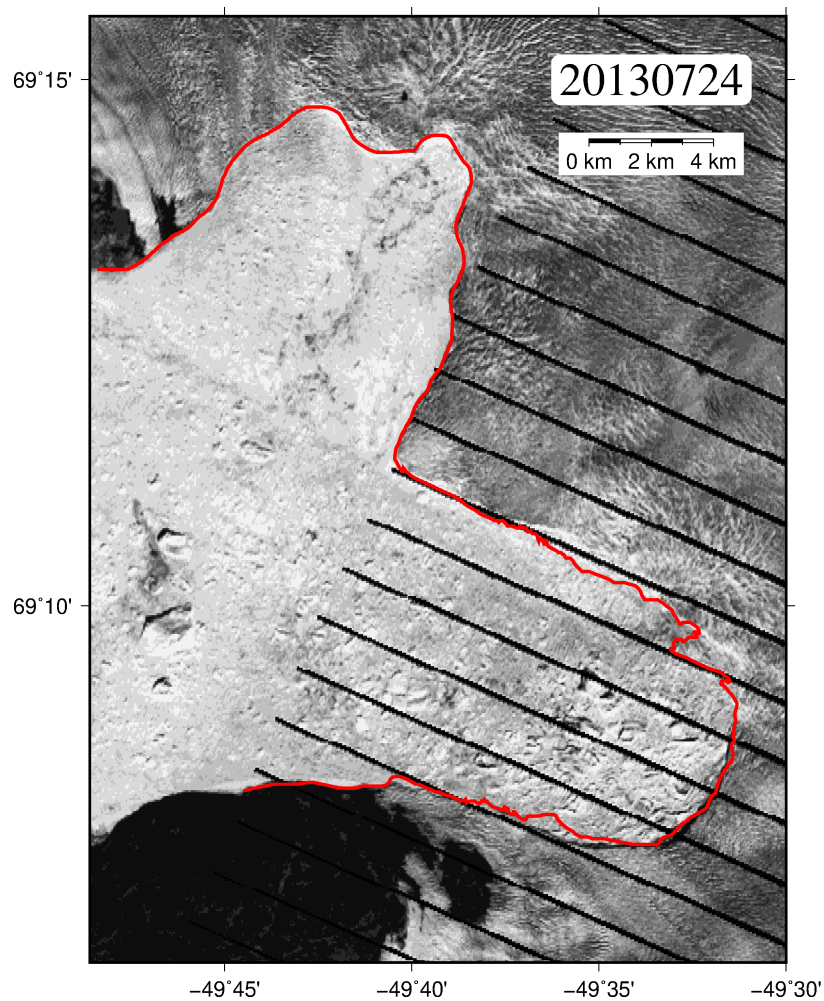

Figure R4. An example of automatically delineated calving front at Jakobshavn Isbrae using a Landsat 7 image with scan line errors. The image was taken on July 24th, 2013.

Anonymous Referee #2

The authors use a deep convolutional neural network with a U-net architecture to delineate the calving fronts of Jakobshavn Isbrae between 2009 and 2015. The network achieves reasonable results, allowing the analysis of the interannual and season behavior of the two branches of the glacier. The authors determine three distinct phases of calving front behavior, which they partially attribute to the bed elevation. There are a some issues with the manuscript regarding originality of the paper, ambiguous or incorrect technical comments, and lack of clarity in some aspects of the methods. However, it does add valuable results and showcases the uses of deep learning in SAR products. Therefore, I believe the article may be considered for publication after Major Revisions, once the following concerns have been addressed:

We highly appreciate the reviewer for the constructive comments which have significantly improved the quality of our manuscript. We have made our best effort to revise the manuscript based on the referee's comments and suggestions.

Major Comment

As the first reviewer pointed out, despite the claim in the manuscript regarding the novelty of the technique, the methodology is very similar to that of Mohajerani et al. [2019] (https://doi.org/10.3390/rs11010074). However, this study does provide a different take on this technique and the authors should point out specifically how this work improves on previous efforts. For instance, the authors here use classification of surfaces in order to obtain the calving front, while Mohajerani et al use semantic segmentation to extract the front without classifying the surrounding surfaces. Each technique has strengths in different contexts. This and other differences should be discussed.

We have added a new subsection 7.1 titled **Differences from the previous work** to discuss the differences between our work and the method of Mohajerani et al. (2019), which are summarized as follows:

- Different strategies are used to classify calving fronts. Our study classifies the surface into two types (i.e., ice mélange and non-ice mélange) to extract the calving front; Mohajerani et al. (2019) use semantic segmentation to extract the front without classifying the surrounding surfaces.
- Additional manual practices such as finding a rotation angle for each glacier are needed in the work of Mohajerani et al. (2019).
- We subdivide the images into small patches, which allows us to use images with high resolutions and various size (i.e., TerraSAR-X images). Mohajerani et al. (2019) resampled images to a fixed size (240 by 152 pixels) with low spatial resolution (49.0 to 88.1 meters).

There are some statements that are not necessarily true from a technical point of view and raise some concern, which require revision:

i) Page 6 Lines 12-15: This is not true. Even when using one architecture, the loss and/or accuracy metrics on the validation dataset can be used during training in order to avoid

overfitting, whereas the test dataset is only used after training. This is particularly important if the trained network is intended to be used in multiple areas.

> We agree and have separated our data into three parts: training, validation, and test. We revised the relevant text as: *We separate all the SAR images into a training-validation dataset (75 images) and a test dataset (84 images) (Table S1). In the training-validation dataset, we randomly choose 90% as training data and take the rest as validation data.* (Page 7 Line 13-15)

ii) Page 7 Lines 7-8: This statement is not necessarily true and could be misleading. A larger kernel provides more context, but doesn't necessarily directly increase precision. It is dependent on the scale of the desired features to be extracted, depth of network, desired level of weight sharing, and many other factors.

> Indeed, the accuracy relies on several factors such as the depth of the network and desired level of weight sharing. The primary purpose of increasing the kernel size is to get smoother calving fronts. We rephrased the relevant text as: *We utilize relatively large convolution kernel size (5 by 5) to obtain smoother calving fronts.* (Page 7 Line 7)

iii) Page 7 Line 27: It is not necessarily true that having more items in a batch reduces overfitting. This is dependent on the total number of epochs that the batches are cycled through and the rate of minimization of the loss function as a function of batch size. Large batches can indeed reduce generalizability (e.g. Keskar et al [2016] https://arxiv.org/abs/1609.04836).

> We agree that a larger batch size would not reduce overfitting but actually reduce generalizability. Typically, batch sizes are no larger than 256. A large batch size would help to increases the efficiency and improves the accuracy of the gradient estimation at each step. Here, the batch size we use is three. We revised the relevant text as: *With a given GPU memory, a smaller patch size allows more items in a batch, which increases the efficiency and improves the accuracy of the gradient estimation at each step. To strike a balance between edge effect and batch size, we choose 960×720 pixels as our patch size and the batch size is three.* (Page 8 Line 22-24)

There is no proper measure of the extent of overfitting in the study. Without a validation dataset to keep track of overfitting during training, and no regularization in the network (or lack of discussion in the manuscript), one cannot make any statements about the generalizability of the model. This is exacerbated by the fact that the authors train and test the network on only one and the same glacier.

We have added the validation dataset and halted the training when the validation error stops to decrease with patience of 5 epochs (Page 7 Line 13-15; Page 9 Line 1-2). The optimizer we use has an L2 regularization term with a factor of 0.00001 (Page 7 Line 12). These strategies help to mitigate overfitting. We chose not to include the dropout layer because we found that

adding a dropout layer caused large fluctuations for both the training loss and validation loss at the end of training.

It would be helpful to provide more detailed information on the time requirements (e.g. Page 7 Lines 30-31) and the GPU model used in the study as a point of reference.

We have provided more detailed information on the time requirements (Page 8 Line 27). We do mention the used GPU model, Quadro P5000 GPU, in the Acknowledgment section. We prefer not to mention any brand name in the main part.

There is very little discussion on the actual architecture of the U-Net model. How many layers are used, what activation functions are used, etc.?

We have added one paragraph and a graph to describe the U-Net architecture (Page 6 Line 13, Page 7 Line 1-12, Figure S1). The architecture we use has 41 layers in total, including 23 convolutional layers and 18 batch normalization layers. The activation function in the last convolutional layer is Sigmoid, and the rest activation functions are LeakyReLU.

It would be more meaningful to put the errors in context. For example Page 8 Line 28, how much of the error is purely from the delineation alone, if you had multiple investigators manually delineate the same calving front? And how do these errors and those reported in Table S3 compare with the resolution of the image in terms of the number of pixels?

We agree that including the error from delineation alone would be more meaningful. We asked another investigator to manually delineate the calving fronts from six selected images. By comparing the two sets of independent delineation results, we obtained a mean difference of 33 meters (equivalent to ~5.5 pixels). We revised the relevant text as: *To measure the manual delineation error, we have another investigator to manually delineate the above-mentioned six calving fronts again. By comparing the two sets of independent delineation results, we obtained a mean difference of 33 meters (equivalent to ~5.5 pixels) (Table S2).* (Page 9 Line 23-26)

We have added the error in terms of the number of pixels in Table S3.

Minor Comments

Page 1 Line 16: add "to" after "stabilized".

We have revised as suggested (Page 1 Line16).

Page 3 Line 13: change "speeded up" to "sped up"

We have revised as suggested (Page 1 Line15).

Table S1: please statement more clearly if 0=test and 1=train to avoid confusion.

We have revised the caption of Table S1 as suggested.

Page 4 Line 15: How are boundaries dealt with in the averaging of pixels?

The images we use to delineate the calving front manually and to apply to the network are all multi-looked images. The original TerraSAR-X images have a high spatial resolution, and their pixel size is 1.25 meters. After reducing the image size by 25 times, the boundaries in the multi-looked images remain visually clear.

Page 7 Lines 3-4: It is not very clear how the calving front is delineated front the closest temporal neighbor. Is there a set distance threshold from the calving front of the reference image?

If the boundary is not clear in an image, we will find its closest temporal neighbor with a clear edge. By observing the texture variation due to the glacier movement, we can approximately decide where the calving front is for the blur image. Figure S3 gives an example of how we dealt with this issue. The manual delineation is all based on visual observations without any quantitative analysis.

Figure S4: "(c) and (c) show the manually delineated calving fronts" should be changed to "(c) and (d) [...]".

We have revised the caption of Figure S3.

We have changed the order of the Figures in supporting information in the order they are referred to in the main manuscript:

Figure S1--> Figure S2

Figure S2--> Figure S4

Figure S3--> Figure S5

Figure S4--> Figure S3.

We have added one figure in supporting information to describe the network architecture (Figure S1).

Page 7 Line 19: Is rotation augmentation necessary if you are only working with one glacier here?

Without rotation augmentation, the trained network still can generate reasonable results. However, we prefer to keep the rotation augmentation since it could be helpful when we apply our method to other glaciers in the future.

Page 7 Line 20: Please explain what you mean by 2% linear stretch. Is this done separately in each direction (horizontal and vertical)?

We didn't do the linear stretch separately in each direction.

The linear stretching is to change the pixels' values to increase the contrast.

For all values between 2% and 98% of the pixel value range, we use the following equation to do the linear stretching

$$P_{stretched} = 255 * \frac{(P_{in} - P_{min})}{(P_{max} - P_{min})}.$$

Where $P_{stretched}$ is the pixels' value after linear stretching and P_{in} is the pixels' value before stretching. P_{min} and P_{max} are the 2nd and 98th percentile in the histogram (that is, 2% of the pixels have values lower than P_{min}, and 2% of the pixels have values larger than P_{max}).

For values lower than P_{min}, they are set as zero, and for values larger than P_{max}, they are set as 255.

We believe that "x% linear stretch" is a widely used terminology in remote sensing and therefore choose not to provide a detailed explanation in the manuscript.

Page 8 Lines 3-4: Just a suggestion: in order to avoid losing training data, you can change the weights in the loss function instead.

Thanks for your suggestion, but we prefer dropping out these one-class patches. By changing the weights in the loss function, we can indeed avoid losing training data. However, the primary purpose of dropping out one-class patches is to save computational power. The network may generate erroneous segmentation in the region that is away from the calving fronts due to dropping out one-class patches. However, we can fix this problem in post-processing by removing small isolated polygons caused by erroneous segmentation.

Page 8 Line 9: what threshold do you use to determine a "stable error"?

We have changed our strategy to ovoid overfitting. With give patience of 5 epochs, if the validation loss stops to decrease, we halt the training process (Page 9 Line 1-2).

Figure 10: the magenta and red colors are very hard to distinguish. Please consider using a more contrasting color.

We have changed the line color from magenta to green.

Section 7.2: What are the limitations of the current technique?

We have added a new subsection 7.3 titled **Limitation of current method** to discuss the limitations of the current technique, which are summarized as follows:

- The U-Net architecture requires relatively high GPU memory.
- Splitting images with overlaps increase the training time.
- The accuracy of this method relies on manual delineation and the information richness of the training dataset.

Could imagery artifacts or more varied surfaces be dealt with?

As long as the calving fronts are clear in the images, imagery artifacts or more varied surfaces will not be a problem to the network. For example, with additional training, the network could handle images with low cloud cover (Figure R1) as well as Landsat 7 images with scan line errors (Figure R2). Note that the image in Figure R1 is not in the training dataset, and the image in Figure R2 is in the training dataset.

However, imagery artifacts such as image distortion need to be corrected by pre-processing procedures other than deep learning.

We did not include the results using Landsat -7 and -8 images since they are preliminary and beyond the scope of this manuscript.
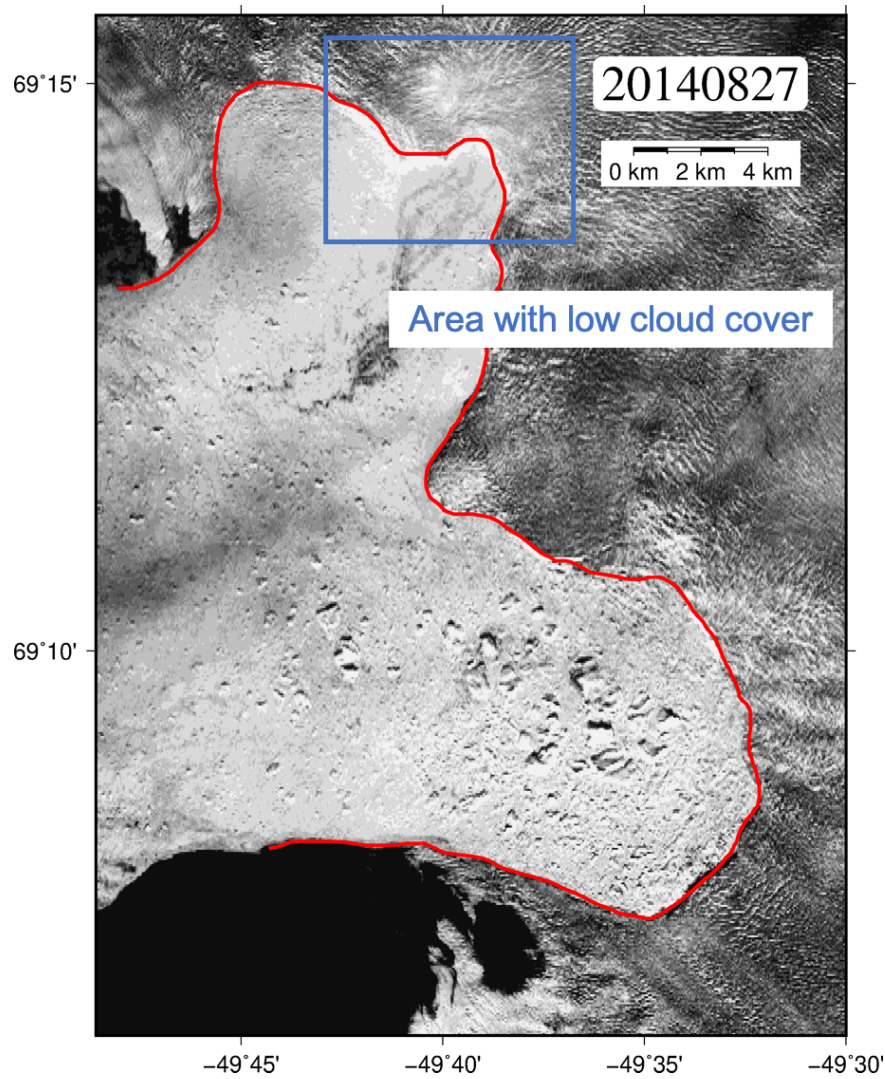
Figure R1. An example of automatically delineated calving front at Jakobshavn Isbrae using a Landsat 8 image with clouds. The image was taken on August 27th, 2014. The blue box indicates an area with low cloud cover.
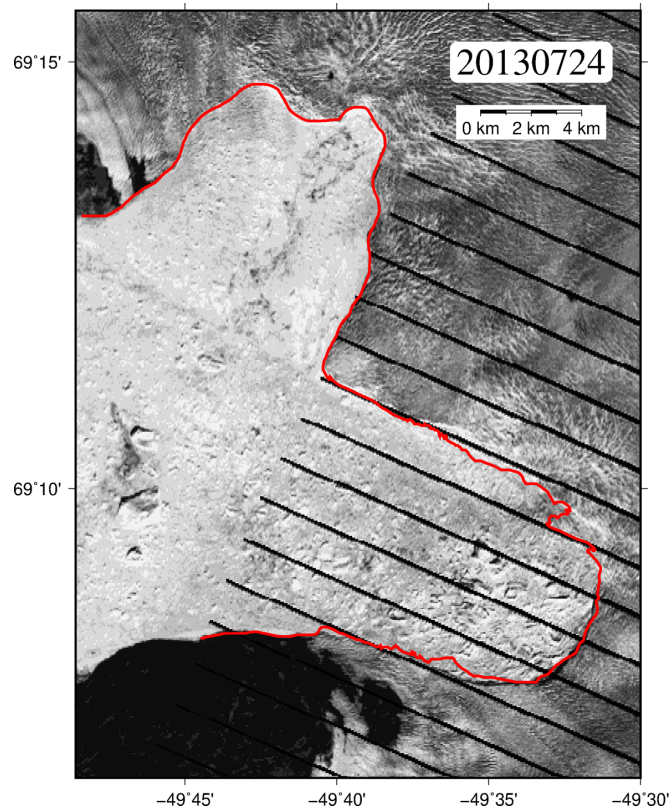
Figure R2. An example of automatically delineated calving front at Jakobshavn Isbrae using a Landsat 7 image with scan line errors. The image was taken on July 24th, 2013.

Can the trained network be applied to multiple glaciers or does it have to be retrained for every glacier?

Currently, if we want to apply the network to other glaciers, retraining is needed. We conducted a preliminary experiment by directly applying the network generated from this work as trained by TerraSAR-X imagery from Jakobshavn to Helheim (without including any new training data). Figure R3 is a superior example shows that the automatically delineated calving front at Helheim is very close to what one would get from visual inspection. Of course, we need to include more training examples from more glaciers to ensure reliable results on other glacier domains.

However, with more and more data from different glaciers included in the training dataset, the trained network has the potential to be applied to another glacier without retraining.

We did not include the results on other glacier domains since they are preliminary and beyond the scope of this manuscript.
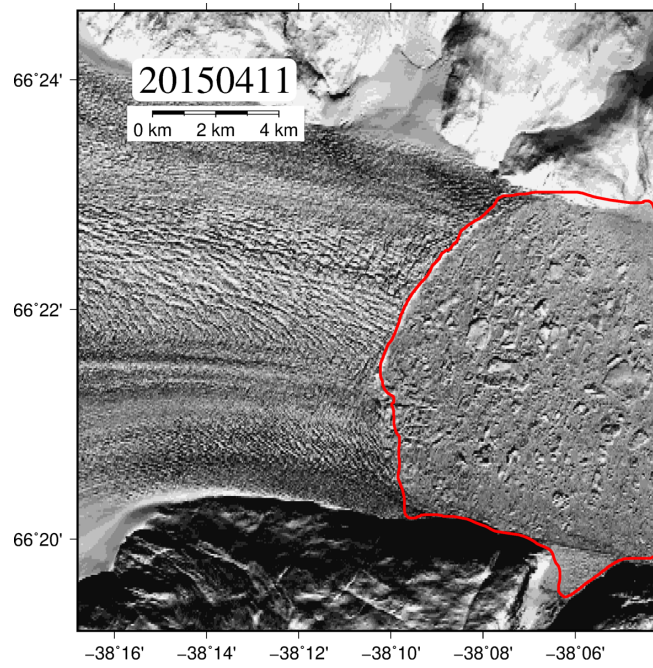
Figure R3. An example of automatically delineated calving front at Helheim. The background image is a Landsat 8 image taken on April 11th, 2015. The red line indicates the automatically delineated calving front.

Automatically delineating the calving front of Jakobshavn Isbræ from multi-temporal TerraSAR-X images: a deep learning approach

Enze Zhang[1], Lin Liu[1], Lingcao Huang[1]

[1]Earth System Science Programme, Faculty of Science, The Chinese University of Hong Kong, Hong Kong, China.

5   *Corresponding to*: Enze Zhang (zhangenze@link.cuhk.edu.hk)

Abstract. The calving fronts of many tidewater glaciers in Greenland have been undergoing strong seasonal and inter-annual fluctuations. Conventionally, calving front positions have been manually delineated from remote sensing images. But manual practices can be labor-intensive and time-consuming, particularly when processing a large number of images taken over decades and covering large areas with many glaciers, such as Greenland. Applying U-Net, a deep learning architecture, to

10   multi-temporal Synthetic Aperture Radar images taken by the TerraSAR-X satellite, we here automatically delineate the calving front positions of Jakobshavn Isbræ from 2009 to 2015. Our results are consistent with the manually delineated products generated by the Greenland Ice Sheet Climate Change Initiative project. We show that the calving fronts of Jakobshavn's two main branches retreated at mean rates of -117 ± 1 m yr^{-1} and -157 ± 1 m yr^{-1}, respectively, during the years 2009 to 2015. The inter-annual calving front variations can be roughly divided into three phases for both branches. The retreat

15   rates of the two branches tripled and doubled, respectively, from phase 1 (April 2009–January 2011) to phase 2 (January 2011–January 2013), then stabilized **to** nearly zero in phase 3 (January 2013–December 2015). We suggest that the retreat of the calving front into an overdeepened basin whose bed is retrograde may have accelerated the retreat after 2011, while the inland-uphill bed slope behind the bottom of the overdeepened basin has prevented the glacier from retreating further after 2012. Demonstrating through this successful case study on Jakobshavn Isbræ and due to the transferable nature of deep learning, our

20   methodology can be applied to many other tidewater glaciers both in Greenland and elsewhere in the world, using multi-temporal and multi-sensor remote sensing imagery.

1 Introduction

Glacier retreating is one of the processes that control the recent speedups of Greenland's tidewater glaciers (King et al., 2018). As glacier retreats, it accelerates to compensate for the loss of downstream buttress. Glacier dynamic instabilities, as suggested

25   decades ago by Meier and Post (1987), play an essential role as the glaciers retreat over depressions in the bedrock topography. For example, Joughin et al. (2008a) indicated that dynamic instabilities caused Helheim and Kangerdlugssuaq Glaciers to speed up as they retreated into an overdeepened basin whose bed is retrograde between 2001 and 2006. Examining 276 marine-terminating outlet glaciers, Bunce et al. (2018) concluded that bed geometry is an important control on the timing and magnitude of glacier retreat.

An accurate and detailed quantification of calving front variations would improve our understanding of the controlling mechanisms of glacier retreat. Moreover, observations of retreat may serve as initial indicators for other dynamic variations such as the glacier acceleration (Moon and Joughin, 2008). Calving front positions are influenced by a range of forces including ice mélange buttressing, increased runoff, and ocean-driven melt (Moon et al., 2015; Fried et al., 2018). Nevertheless, the mechanisms behind the numerous and complex controls on front positions are not yet fully understood.

Compared with manually digitizing calving fronts, automatic mapping is superior because of greater productivity and reliability and lower cost. While most of the previous studies have manually delineated the calving fronts (e.g., Howat et al., 2005; Joughin et al., 2008b), studies by Sohn et al. (1996) and Seale et al. (2011) have automatically delineated calving fronts using feature extractors. Sohn et al. (1996) designed a method to extract ice sheet margin by applying Roberts edge extractor to ERS-1 Synthetic Aperture Radar (SAR) images. Seale et al. (2011) automatically identified glacier calving fronts from daily MODIS images by combining Sobel and brightness profiling methods. With low computational complexity requiring no training and little memory resources, these feature-extracting methods are promising but require extensive prior knowledge and experience. **Deep learning method has also been applied to delineate the calving front positions. Mohajerani et al. (2019) have applied U-Net architecture to Landsat-5, -7, and -8 images over Jakobshavn, Sverdrup, Kangerlussuaq, and Helheim glaciers.**

Deep learning can solve more complex problems with little prior knowledge required and take advantage of increased data volume (LeCun et al., 2015). With the continuous accumulation in the past decades and in recent space missions, the data volume of remote sensing imagery in the polar regions has increased dramatically. Moreover, glacier systems are complex, as conditions such as weather and glacier dynamic behaviors vary from place to place and from season to season. There are therefore obvious advantages to applying deep learning techniques to automatically extract glaciological features from the available big data.

Here we aim to design a ~~novel~~ method to automatically delineate a glacier calving front from multi-temporal TerraSAR-X (TSX) images based on deep convolution neural networks (DCNNs). More specifically, we delineate the glacier calving front of Jakobshavn Isbræ (Fig. 1a) and quantify its seasonal and inter-annual variations. With this new set of observations, we investigate the possible link between calving front variations and bed elevation.

DCNNs are a class of the deep learning methods, and have made important breakthroughs in image processing. DCNNs can discover both low-level (e.g., edges, corners, and lines) and mid-level features (e.g., shapes, sizes, and locations) (Sun et al., 2014; Zhang et al., 2015). Recently, some studies have used DCNNs on high-resolution SAR images to perform classification tasks (Geng et al., 2015; Huang et al., 2017). These studies unanimously agree that DCNNs outperform traditional classification methods on SAR images.

We use TSX images due to their high temporal resolution (11 days), high spatial resolution (3.3 to 3.5 meters), and ability to penetrate cloud cover. These high-temporal-resolution images have been acquired in all seasons and allow us to investigate calving front variations with a high degree of continuity and consistency. With these high-spatial-resolution images, we can easily digitize the calving fronts (known as "ground truth" in the context of deep learning), and verify the accuracy of the

DCNN. Using SAR images can avoid the cloud cover problem associated with optical images such as the Landsat-8 image shown in Fig. 1b.

2 Jakobshavn Isbræ

Jakobshavn Isbræ, located in central-west Greenland, is one of the largest and fastest tidewater glaciers in the world. In Jakobshavn, the ice flows westward to the ocean and divides into two branches near the coast (Fig. 1a). Before summer 2004, these two branches merged and flowed into the Kangia fjord. Afterwards, as the glacier retreated, the two branches became disconnected (Bondzio, 2017). During the past few years, Jakobshavn Isbræ has undergone dramatic acceleration as the glacier has retreated and thinned (Joughin et al., 2008c; Joughin et al., 2012). Jakobshavn's calving front retreated 16 km between 2002 and 2008 (Rosenau et al., 2013). This glacier alone has contributed nearly 1 mm to the global sea level rise from 2000 to 2011 (Howat et al., 2011).

Observations have shown that the calving front variations were correlated with the glacier velocity changes in Jakobshavn Isbræ. In 1998, the glacier sped up by 18% in its frontal regions, coinciding with the initial retreat of the ice tongue (Thomas, 2004; Luckman and Murray, 2005). The glacier doubled its speed by spring 2003, when nearly the entire floating ice tongue had disintegrated (Joughin et al., 2004). After the loss of this ice tongue, the glacier's velocity fluctuated seasonally from 2004 to 2007 (Joughin et al., 2008b). The glacier slowed down when it was advancing, and ~~speeded~~**sped** up when it was retreating (Joughin et al., 2012).

The variations of Jakobshavn's calving front are also strongly influenced by the presence of ice mélange, namely a mixture of calved icebergs and sea ice (Fig. 1a). The seasonal variation of the calving front in Jakobshavn Isbræ is well correlated with the growth and recession of sea ice in the Kangia fjord (Sohn et al., 1998; Joughin et al., 2008c). Temporal variations of the ice mélange strength can also control the timing of calving events and influence the evolution of the calving front position (Amundson et al., 2010).

Our study area covers a 14×18 km section of the frontal area of Jakobshavn, and includes bedrock, ice mélange, and glacier regions. We restrict the extent of the study area to reduce the computational costs, while also ensuring the coverage of all the calving fronts within our investigation period (2009–2015, determined by the TSX images we have access to). We classify our study area into two classes: ice mélange and non-ice mélange regions (including both glacier and bedrock regions). We delineate the boundaries between these two regions and retrieve glacier calving fronts. The repetitive texture of crevasses in the glacier region clearly distinguishes it from ice mélange, where icebergs are distributed discretely. It is easy to identify the bedrock region because of the distinct bedrock texture, including cracks and land-based lakes.
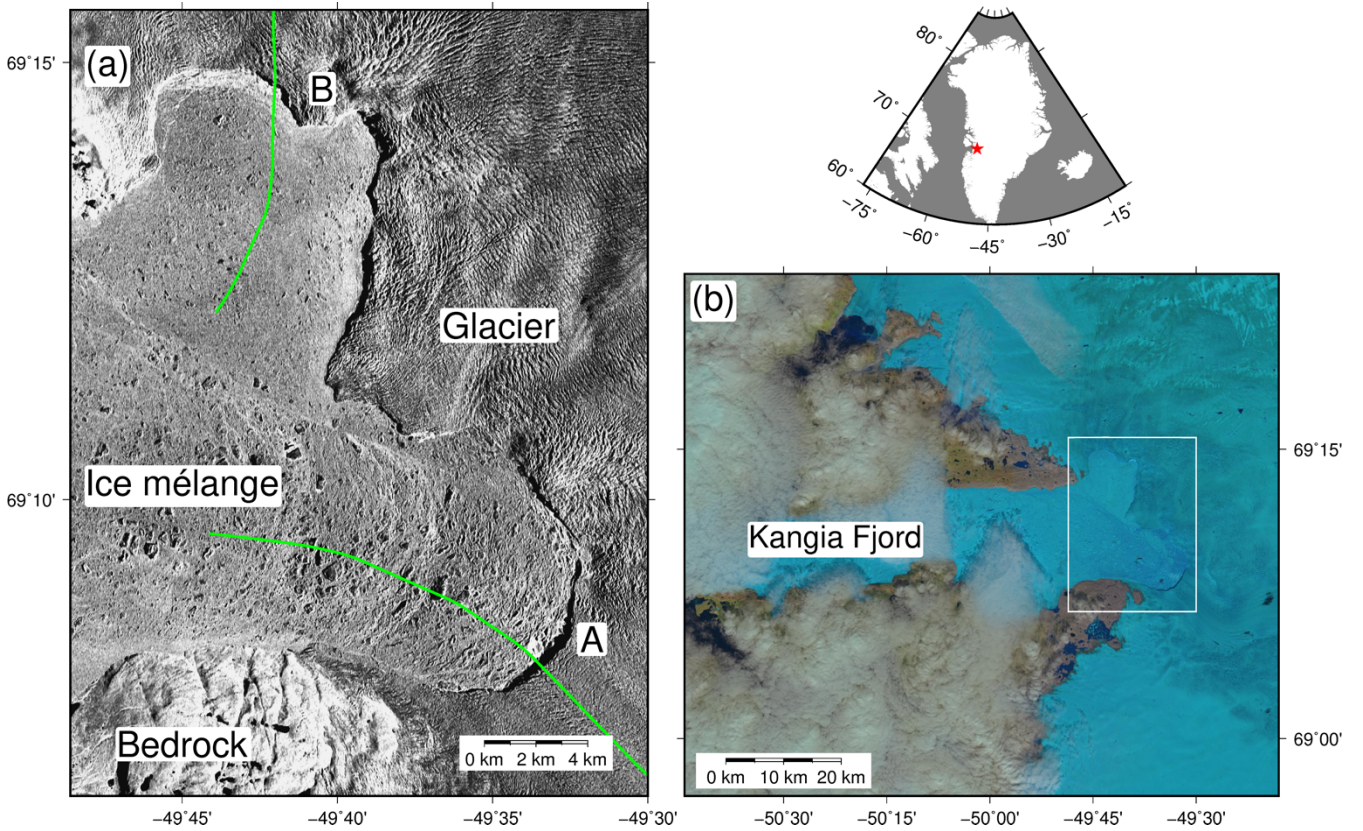
Figure 1. (a) TerraSAR-X image taken on 11th July 2015 showing the frontal area of Jakobshavn Isbræ. Its two branches are labeled as 'A' and 'B'. The green lines indicate the location of the bed elevation profiles shown in Fig. 8. **(b)** Landsat-8 image taken on 13th July 2015. The white box shows the area illustrated in Fig. 1a.

3 TerraSAR-X images and pre-processing

The German SAR satellite TerraSAR-X was launched in June 2007 and carries an X-band SAR sensor. In this study, we use TSX images taken in both ascending and descending orbits and in stripmap imaging mode. We use the enhanced ellipsoid corrected (EEC) products, which are multi-looked, projected and resampled to the WGS84 reference ellipsoid. We use 159 images in total, taken between April 16th, 2009 and December 23rd, 2015 (listed in Table S1). We apply three pre-processing procedures including despeckling, multi-looking, and re-georeferencing. Figure 2 shows an illustrative example of our pre-processing workflow, which we will describe below in detail.

Because the quality of SAR images is adversely affected by the speckle noise (Fig. 2a), we apply the median blur filter to mitigate the speckle noise (Fig. 2b) and then multi-look the filtered images to reduce their size by 25 times (Fig. 2c). The median blur filter is widely used in image processing and is particularly effective for speckle noise. With the despeckled images, we average five neighboring pixels (vertically and horizontally) by using Geospatial Data Abstraction Library (GDAL) package (www.gdal.org). Moreover, both despeckling and multi-looking can smooth images without the loss of essential

4

information for delineating the calving fronts. After despeckling and multi-looking, the pixel size of our images is six meters.

We choose the EEC products because they include topographic correction and are the standard geocoded products of TSX (Roth et al., 2004). However, even for the EEC products, we observe that the geocoding information for our study area is inaccurate. First, overlaying the EEC images on Google Earth, we note obvious offsets between these two. Second, the geocoding information is inconsistent in different orbit directions of EEC products. Therefore, we need to re-georeference the EEC products. For the images we have, we observe that the images in the same orbit direction have identical geometry. Based on this observation, we assume that the differences between the EEC products and the Google Earth images are systematic, namely that they are consistent for the EEC products in the same orbit direction. We correct the geocoding information of the EEC products using 16 ground control points on Google Earth images, including the center of lakes and cross sections of the bedrock, due to their stability and ease of identification. For all the EEC products in the same orbit direction, we apply the same thin plate spline transformation using the GDAL package.
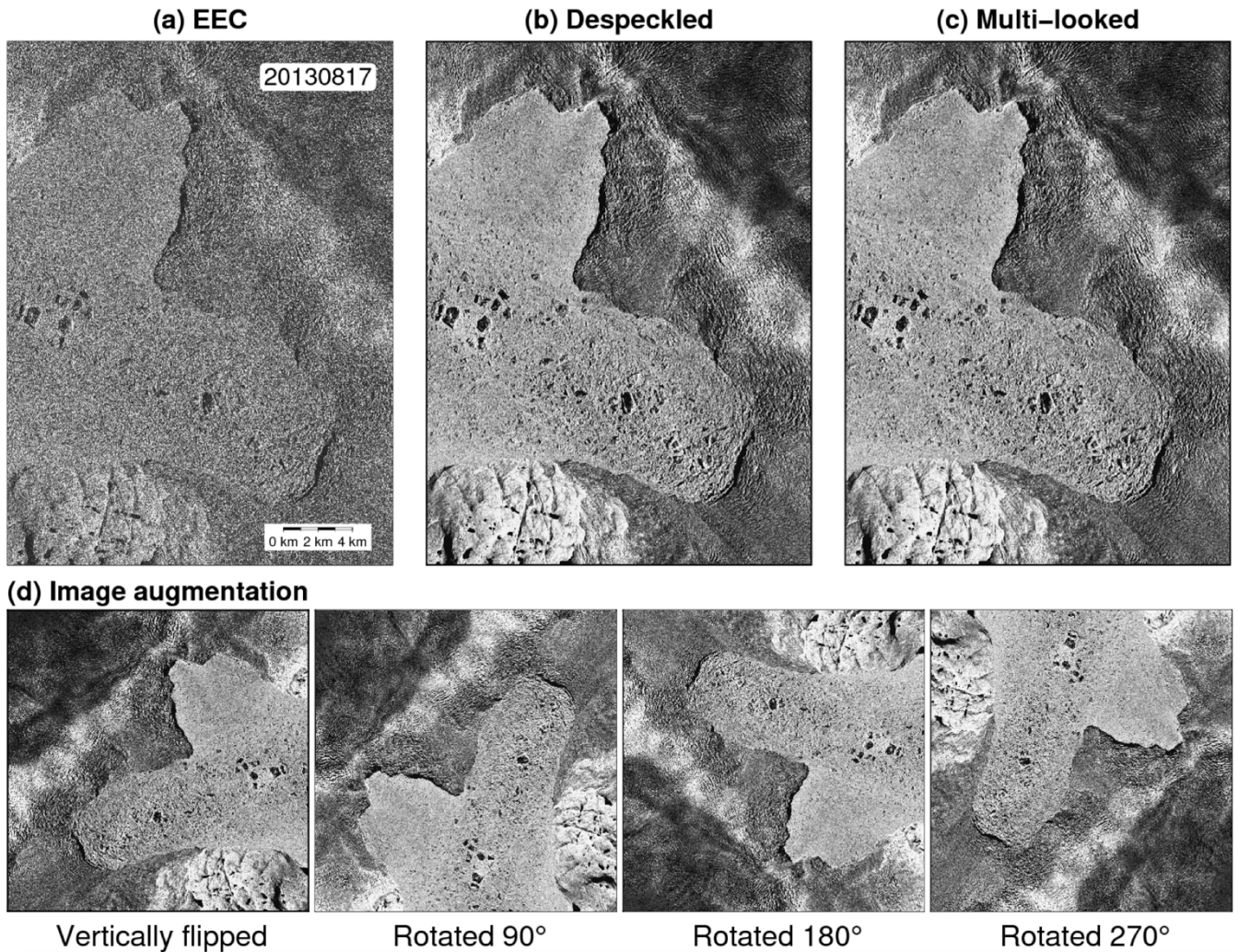
Figure 2. A set of examples of TerraSAR-X data pre-processing and preparation, including (a) the enhanced ellipsoid corrected (EEC) product, (b) despeckled image after reducing the speckle noise, (c) multi-looked image after decreasing the image size, and (d) images after vertically flipping and rotating Fig. 2c by 90º, 180º, and 270º, respectively. For ease of presentation, the images in Fig. 2d are not to scale with (a)–(c).

4 Deep learning and post-processing

DCNNs are a class of neural networks that consist of numerous convolutional layers, each of which contains learnable weights and biases. A network's architecture refers to its overall structure, including the number of units and layers the network has and how they are connected. Here, we use the U-net architecture, which has achieved outstanding performance in biomedical segmentation applications and is among the best methods in image segmentation (Ronneberger et al., 2015). This network is fast, taking less than a second on a mainstream graphics processing unit (GPU) to segment a 512×512 image. **The U-net architecture consists of a contracting path and an expansive path (Fig. S1). The contracting path consists of repeated**

6

application of two 5×5 convolution layers, each followed by a batch normalization layer and a leaky rectified linear unit (LeakyReLU) activation function, and 2×2 max pooling operation for downsampling feature maps and doubling the number of feature channels. Every step in the expansive path consists of a 4×4 up-convolution layer that upsamples the feature map and halves the number of feature channels, a concatenation with the corresponding feature map from the contracting path and two 5×5 convolution layers, each followed by a batch normalization layer and a LeakyReLU activation function. The final layer is a 3×3 convolutional layer with Sigmoid activation function to get the final segmentation patch. We utilize relatively large convolution kernel size (5 by 5) to obtain smoother calving fronts. We use LeakyReLU activation functions with a slope of 0.1 below zero, which allows for small, non-zero gradient when the unit is not active (Mass et al., 2013), making optimization potentially more robust. We use binary cross-entropy (BCE) between the ground truth images, and the network outputs to measure the training error because it avoids the problem of slow learning (the training loss decreases slowly) (Goodfellow et al., 2016). We use adaptive moment estimation max (AdaMax) (Kingma and Ba, 2014) as the optimizer with a learning rate of 0.0001 and an L2 regularization factor of 0.00001.

The proposed framework for using deep learning to delineate the calving fronts is summarized in Fig. 3. We separate all the SAR images into a training-validation dataset (75 images) and a test dataset (84 images) (Table S1). ~~This practice is different from~~In the ~~deep learning convention that separates the entire dataset into three groups:~~ training, validation, dataset, we randomly choose 90% as training data and ~~test. As~~take the rest as validation data. The validation dataset is ~~used to evaluate the performance of different architectures, and since we only use one architecture (i.e., U-net), we choose not to include a validation dataset~~for minimizing overfitting and tuning the hyperparameters of the network such as learning rate and kernel size.
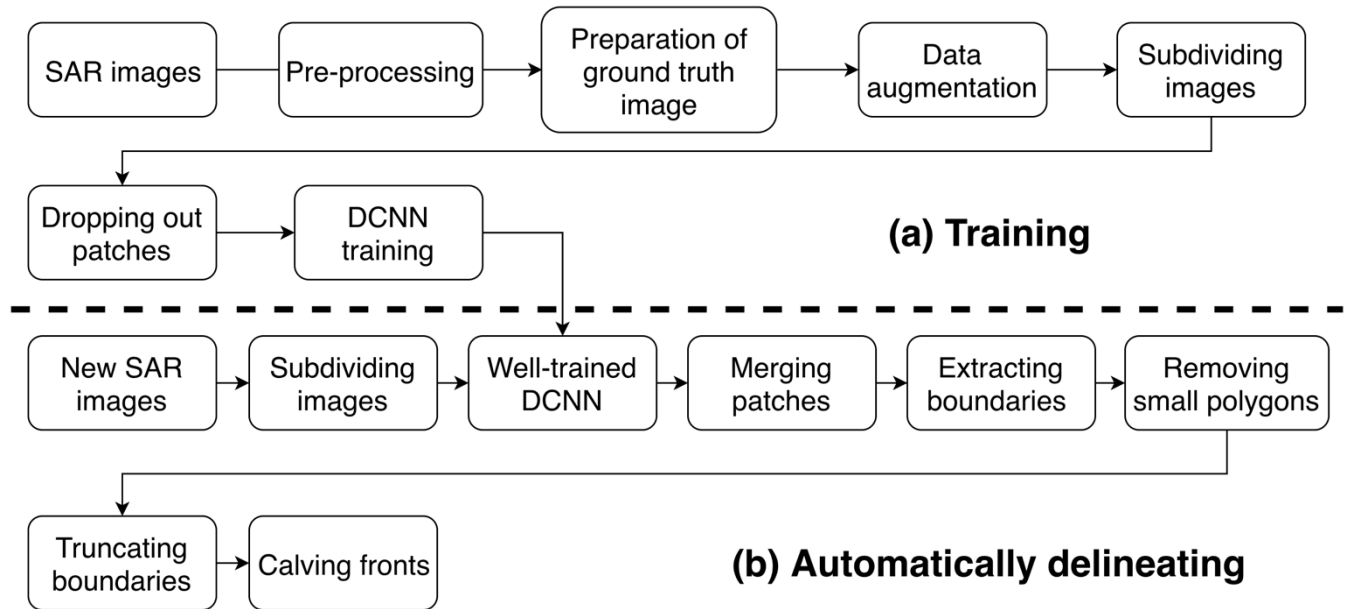

(a) Training

(b) Automatically delineating

Figure 3. Diagram of the proposed framework. Details are described in Sections 3 and 4.

Before training the network, we prepare the training dataset, including training images (SAR images) and their corresponding ground truth images. The ground truth images have two classes: the ice mélange region is set as zero, and the non-ice-mélange region (including both glacier and bedrock regions) is set as one (Fig. ~~S1~~**S2**). The ground truth images are derived by converting

5 the vector of manually delineated calving fronts to rasters using GDAL. Manual delineation is simple on most TSX images. However, it is challenging to delineate the calving fronts on few TSX images acquired in winter and spring because the boundaries are obscured due to snow ~~covering~~**cover** and sea ice bonding. For each of these obscure images, we use its temporally closest image with a clear calving front as a reference, and require that our manually-delineated fronts are smooth (Fig. ~~S4~~**S3**).

10 ~~We increase the smoothness of our network results in two ways. First, despeckling and multi-looking can smooth images without the loss of essential information for delineating the calving fronts. Second, we increase the convolution kernel size from the default 3×3 pixels to 5×5 pixels. This practice gives us more precise and smoother results. However, a larger convolution kernel increases the training time.~~

To ensure the effectiveness and accuracy of deep learning for a set of SAR images taken in all seasons spanning seven years,

15 it is important to prepare a sufficiently diverse training dataset. We include at least one image in each month into the training dataset to represent various conditions related to radar backscatter and image texture. First, radar backscatter can vary with the dielectric properties of the surface scatterers in the study area due to changes in snow ~~covering~~**cover**, wetness, and variations in geometric properties such as roughness, grain size, and internal structure (Fahnestock et al., 1993). Since our study area is in the ablation region, backscatter increases in winter because of dry snow ~~covering~~**cover** and decreases in summer due to

20 snow melting. Second, the seasonal and inter-annual variations of ice mélange condition can change the image texture. Sea ice formation in winter solidifies ice mélange, while ice mélange weakens in summer, resulting in freely floating icebergs (Amundson et al., 2010; Xie et al., 2016).

We also perform data augmentation to enrich our training dataset. We adopt the following two strategies. First, we vertically flip and rotate our training images by 90°, 180°, and 270°, respectively, to constitute many possible locations of the calving

25 front in the study area (Fig. 2d). Second, we apply 2% linear stretch to the training images to enhance the edges.

We subdivide each image (3565×1634 pixels) in the training dataset into small patches (960×720 pixels). Otherwise, the resolution would be limited by the GPU memory. We split images with overlaps, and obtain 36414 patches in total. Such a strategy allows a seamless segmentation after merging, which reduces the edge effect. A larger patch size can also better mitigate the edge effect. ~~Another~~**A** common ~~issue of DCNN~~ training ~~is overfitting, which refers to the cases when the test~~

30 ~~errors are large even though the training error is small. We adopt the common overfitting mitigation~~ strategy ~~by stacking~~**in deep learning is to train** several ~~items~~**training examples as a batch each time instead** of training ~~data together as a batch.~~**the whole dataset.** With a given GPU memory, a smaller patch size allows more items in a batch, which ~~consequently better mitigates overfitting.~~**increases the efficiency and improves the accuracy of the gradient estimation at each step.** To strike

<div align="center">8</div>

a balance between ~~overfitting and~~ edge effect **and batch size**, we choose 960×720 pixels as our patch size **and the batch size is three**.

Due to different computational time used in training and automatic delineation, the overlap areas between adjacent patches are set differently in the training and the test datasets. Taking the GPU we use as an example, training the network takes longer time (~~dozens of~~**80** hours) than automatic delineation (~~several~~**20** minutes) after the network is well trained. Therefore, we split the training images with smaller overlap (two-thirds of the patch size) to save computational power and split the test images with larger overlap (four-fifths of the patch size) to make denser samplings so that the results become more robust.

Balancing the number of training samples between classes is crucial in deep learning (Batista et al., 2005; Anantrasirichai et al., 2018). Compared with patches with two classes, patches with only one class are not equally helpful for delineating boundaries. However, one-fifth of the 36414 patches only have one class. Therefore, we randomly drop out 80% of these patches to make the network perform better on the boundary between two classes and also to save computational power.

Training the network starts with initializing all weights as zero~~, and then we use backpropagation to reduce the loss. Two properties are needed for the loss function: (1) the loss should be positive; and (2) it should tend towards zero as the network output is closer to the ground truth. We use binary cross-entropy (BCE) between the ground truth images and the network outputs to measure the training error,~~**. We stop the training when the validation error starts to increase with patience of five epochs.** ~~because it avoids the problem of slow learning (the training loss decreases slowly) (Goodfellow et al., 2016). When the error is no longer decreasing and stable, we consider the network as being well trained.~~

After the training, we first subdivide each test TSX image into small patches and use the well-trained network to segment all the patches into ice mélange and non-ice-mélange classes. Then, we merge the segmented patches (binary images with a pixel value of one or zero) into a single segmentation image by averaging the overlaps. After merging, if the pixel value is larger than 0.5, we consider the pixel to be in a non-ice mélange region. We use GDAL to convert the segmentation image into a vector, which contains a large polygon constituted by both the calving front and the image border, and small isolated polygons caused by erroneous segmentation. After removing the small polygons and truncating the large polygon to separate the calving front from the image border, we finally obtain the calving front for each image.

Using the post-processed delineation results, we can quantify the temporal calving front variations of both branches. Taking the earliest calving front (April 16th, 2009) as the reference, we calculate the enclosed area bounded by the reference and the calving front in a given TSX image. We adopt these metrics of area changes because they take both calving front position and shape into account.

5 Data validation and error estimation

Our results are validated by calving front products from the Greenland Ice Sheet Climate Change Initiative (CCI) project (http://products.esa-icesheets-cci.org). The CCI calving fronts are derived by manual delineation using ERS & Sentinel-1 SAR, and Landsat-5,7,8 optical imagery. We validate our results in the following two aspects.

First, the validation of the re-georeferencing (Section 3) is derived by directly comparing the manually delineated calving fronts obtained from this study and the CCI products. **The calving fronts from these two datasets should be on the same date, and therefore, only six calving fronts are compared.** We manually delineate the calving fronts from the TSX image after re-georeferencing and then calculate the averaged width of the enclosed area bounded by both the calving fronts from

5 these two datasets ~~on the same date.~~. The mean difference is 104 meters (**equivalent to ~17.3 pixels) (**Table S2~~,~~ **and** Fig. ~~S2~~**S4**). Several reasons could cause such a seemingly large difference. The geocoding information of the CCI products also has uncertainties. Moreover, manual delineation from both the CCI and ours are subjected to image quality and the different criteria we adopt for front delineation. **To measure the manual delineation error, we have another investigator to manually delineate the above-mentioned six calving fronts again. By comparing the two sets of independent delineation**

10 **results, we obtained a mean difference of 33 meters (equivalent to ~5.5 pixels) (Table S2).**

Second, the difference of calving front variations between ours and the CCI presents an overall validation that sums up both re-georeferencing and network-delineation uncertainties. We quantify the calving front variations of the CCI products with the same method and reference used in our results. Finally, we calculate the difference between these two variations in terms of both area and equivalent length.

15 The errors in the test dataset represent the error of the network. Unlike the BCE-measured segmentation error in training, the test error is for calving front delineation. We measure the test error by calculating the averaged width of the enclosed area bounded by the manually delineated and the network-delineated calving fronts (Fig. ~~S3~~**S5**).

6 Results

We present our results in the following order: (1) the network-delineated calving fronts from April 16th, 2009 to December

20 23rd, 2015, which are shown in a movie (Movie S1); (2) two examples of our automatically delineated calving fronts (i.e., results in test dataset) (Fig. 4); (3) retreat rates (Table 1) and time series of calving front variations (Fig. 5); (4) inter-annual calving front variation (Fig. 6 and 7).

The individual network-delineated results are influenced by ~~the~~ image quality. Usually, the boundary is more distinct in summer than in other seasons, yielding superior results (Fig. 4a). In winter and spring, the boundary is obscure due to the low

25 contrast and similar texture of the images, for example, the Branch B and the northern part of Branch A (Fig. 4b, 4c, and 4d). The backscatters of the snow-covered ice mélange and the glacier are similar. Moreover, sea ice formation in winter solidifies the ice mélange and even bonds it with the glacier. As a result, our detected edge deviates from the ground truth. Table S3 lists all of the test error with a mean of 38 meters. It also shows that our network performs better in summer than other seasons.
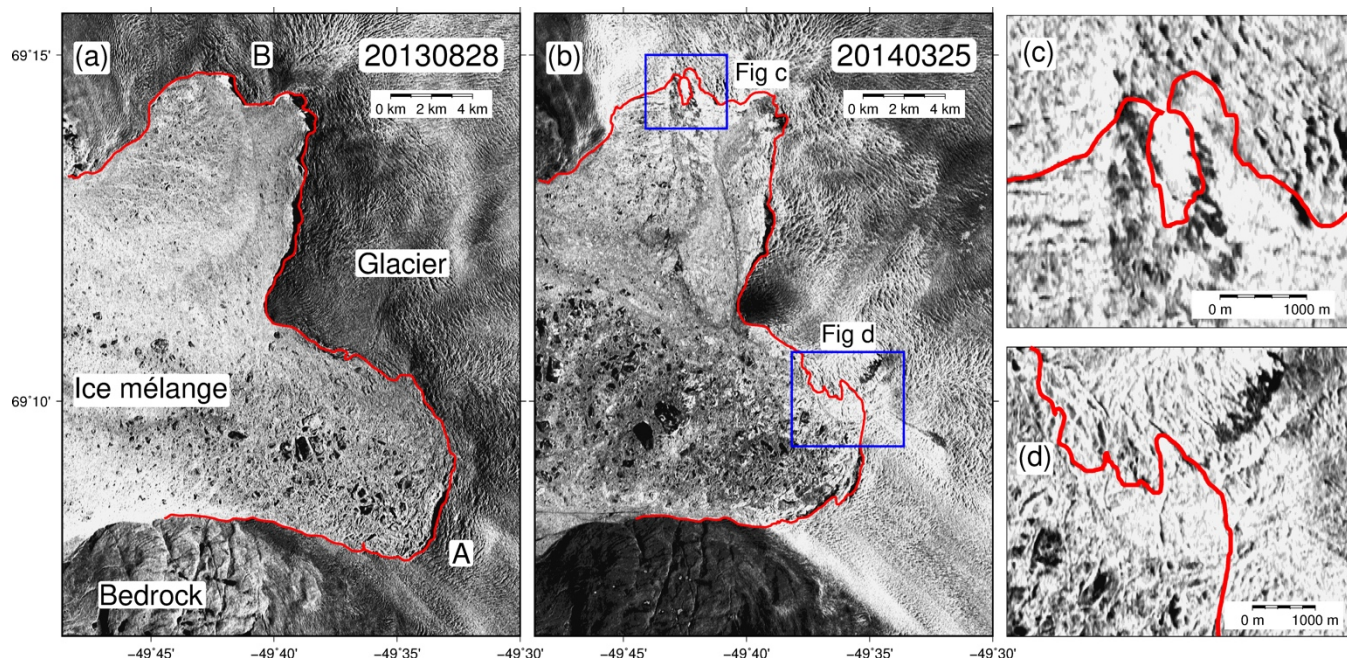
Figure 4. Examples of (a) superior and (b) inferior delineation from our deep-learning-based method. In both (a) and (b), the red line shows the calving front delineated by the network. (c) and (d) show the zoom-in figure of the obscure calving front positions within the blue boxes in (b).

Overall, our results agree well with the CCI products (Fig. 5). The area difference is 2.14×10^6 m^2, and the equivalent length difference is 73 meters. Moreover, our results have a higher temporal resolution (about two measurements every month) than the CCI products (about four measurements every year). Therefore, we can observe the seasonal and inter-annual variations more clearly. Based on our results, Branches A and B retreated from 2009 to 2015 with linear trends of -117 ± 1 m yr^{-1} and -157 ± 1 m yr^{-1}, respectively. The inter-annual variation can be roughly divided into three phases (Fig. 5 and summarized in Table 1). (1) From April 2009 to January 2011, the retreat rates were -141 m yr^{-1} and -228 m yr^{-1} along Branches A and B, respectively. (2) From January 2011 to January 2013, the glacier retreated 170% and 61% faster than in the previous phase in Branches A and B, respectively. (3) From January 2013 to December 2015, these two branches behaved differently. In Branch A, the glacier retreated and advanced seasonally, but at much slower average rates (-23 m yr^{-1}). In Branch B, the seasonal variations were minor, and the glacier retreated slowly (-46 m yr^{-1}).

Table 1. Retreat rates in area and equivalent length during different phases.

|	Period	Mean retreat rate	
		Branch A	Branch B
Area change $(10^6$ m^2 yr^{-1})	Apr 2009–Jan 2011	-3.07 ± 0.05	-4.97 ± 0.09
	Jan 2011–Jan 2013	-8.30 ± 0.04	-8.03 ± 0.07
	Jan 2013–Dec 2015	-0.50 ± 0.03	-1.01 ± 0.03
	Apr 2009–Dec 2015	-2.56 ± 0.01	-3.41 ± 0.01
Equivalent length change (m yr^{-1})	Apr 2009–Jan 2011	-141 ± 4	-228 ± 9
	Jan 2011–Jan 2013	-381 ± 3	-368 ± 5
	Jan 2013–Dec 2015	-23 ± 2	-46 ± 2
	Apr 2009–Dec 2015	-117 ± 1	-157 ± 1
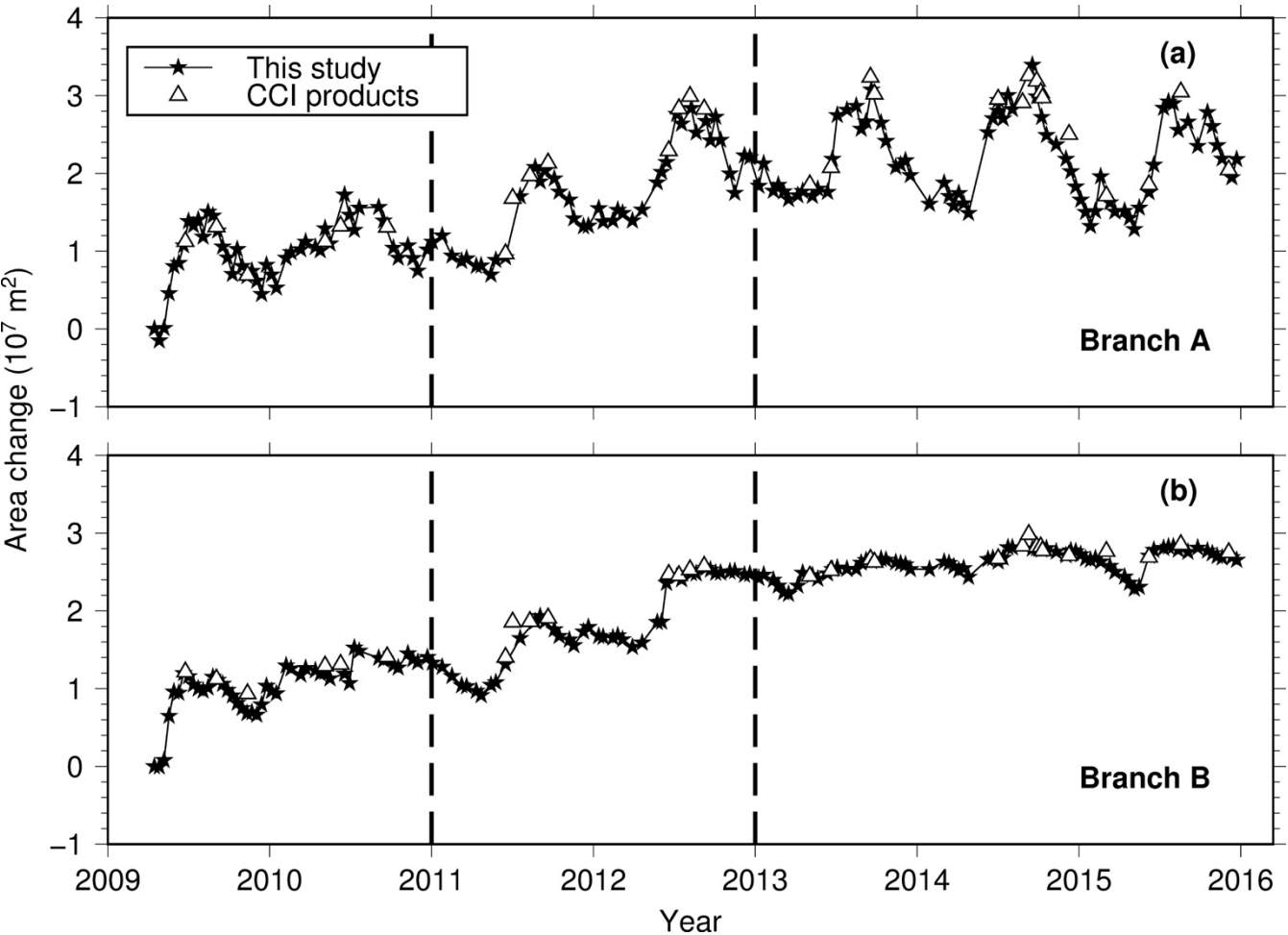
Figure 5. Time series of calving front variations (in area changes) of Branches A and B from our deep learning method (stars) and the Greenland Ice Sheet CCI project (triangles). Dashed vertical lines divide the time series into three separate phases (see text).

Further examination of the inter-annual variation indicates that the calving front exhibited different seasonal variations from year to year. First, even within a close distance of ten kilometers around the coastal area, Branches A and B behaved

12

asynchronously. For example, in 2010, Branch A began to retreat in May, while Branch B started to retreat one month later (Fig. 6a and 7). Moreover, after 2012, Branch A's front underwent strong seasonal variation while Branch B's front remained relatively stable (Fig. 6b). Second, the retreat timing of the glacier varied in different years. In Branch A, the front began to retreat around May in most years, while in 2011 and 2013 the retreat started in June. In 2010, both branches experienced a sudden retreat from mid-January to early February, and then became stable. Third, the calving front variation became regular after 2012. In Branch A, the front stopped retreating in July of each year, and its position remained unchanged up to September to October. In Branch B, the front advanced in spring and retreated in early summer, while its position remained almost unchanged in other seasons (Fig. 5 and 7).
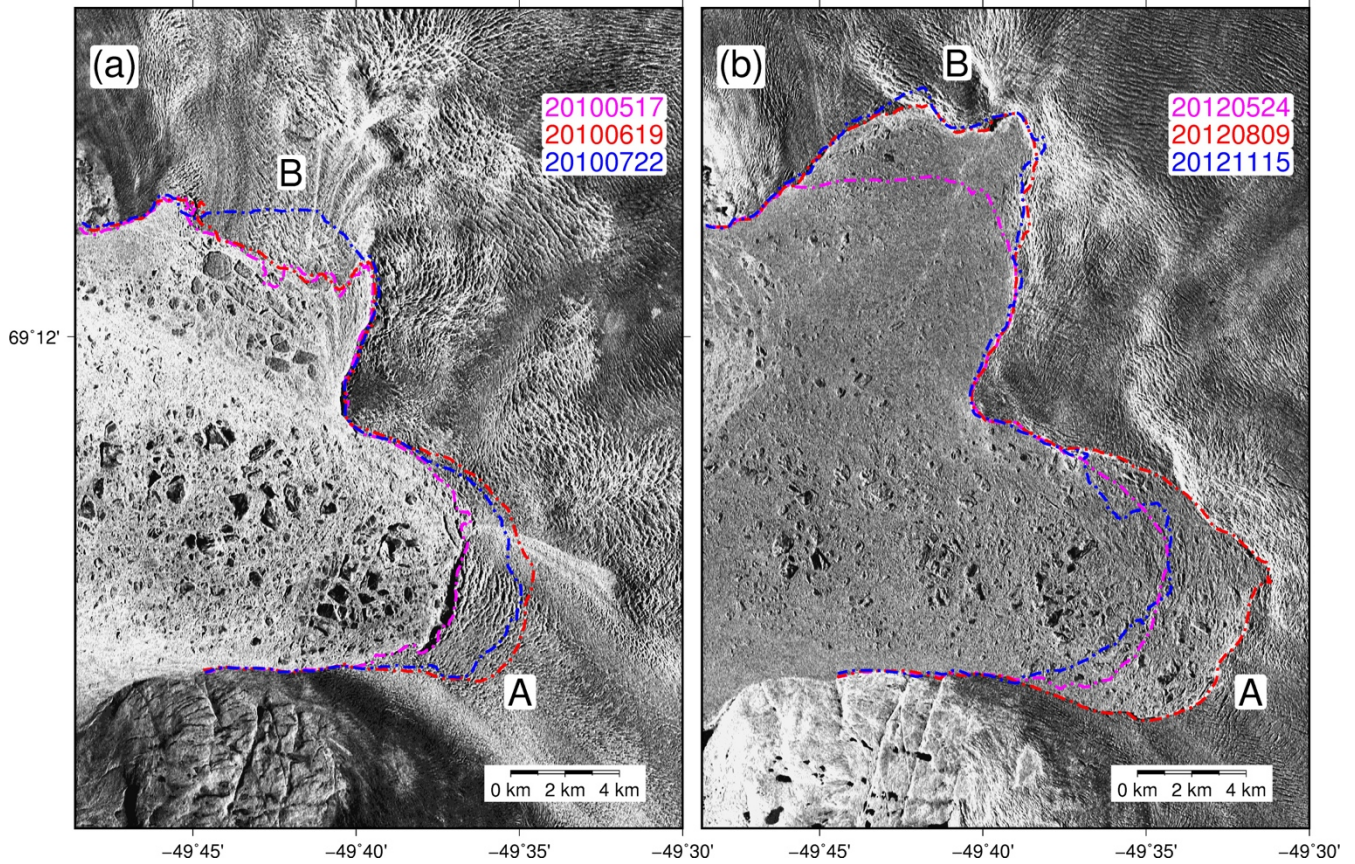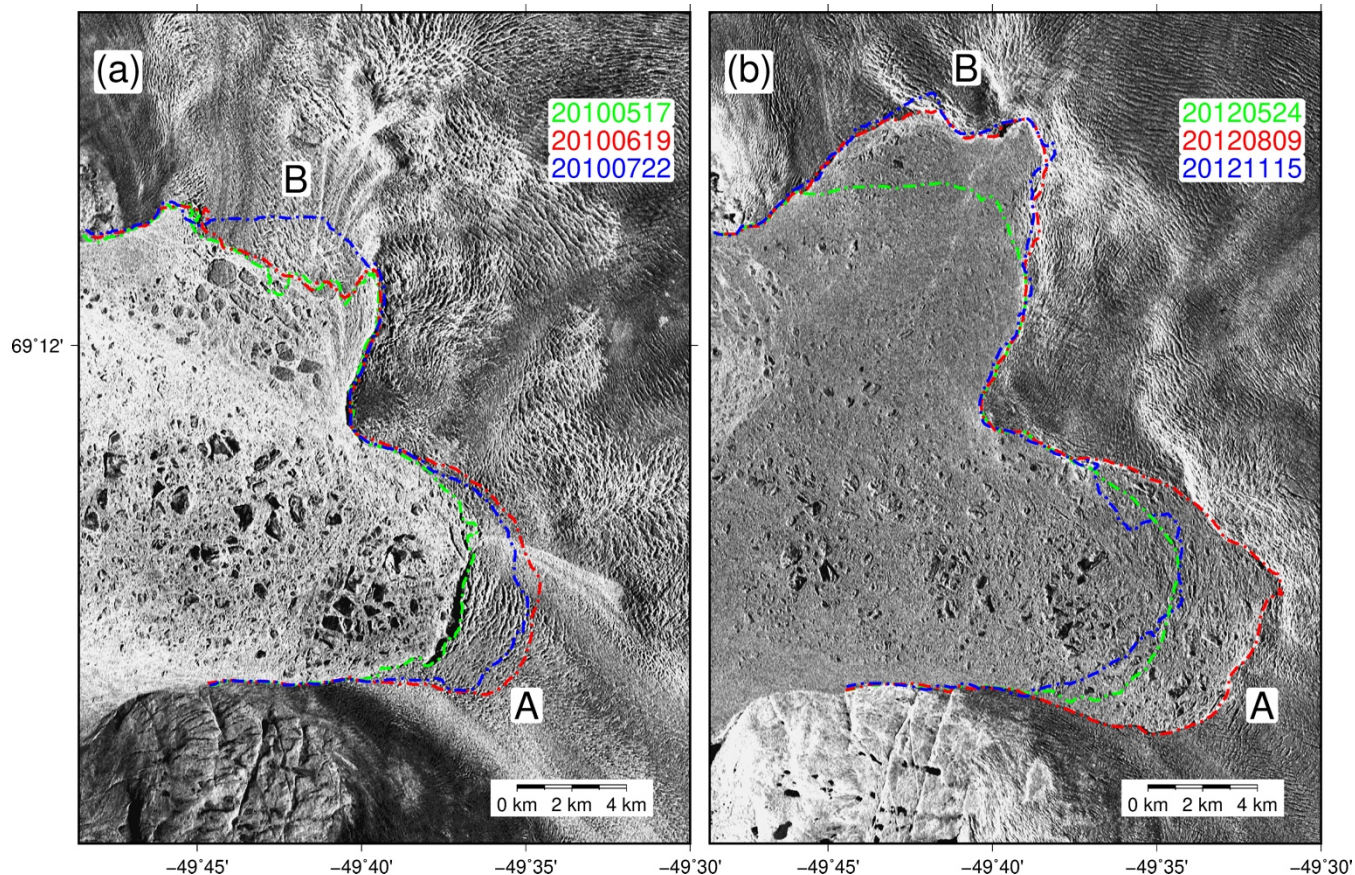
Figure 6. Two examples showing the asynchronous behaviors of Branches A and B. (a) Branch A began to retreat in May 2010, while Branch B started to retreat one month later. (b) Branch A's calving front underwent strong variation between August to November 2012, whereas Branch B's calving front was relatively stable. The magenta line in both (a) and (b) shows the calving front position just before the annual retreat.
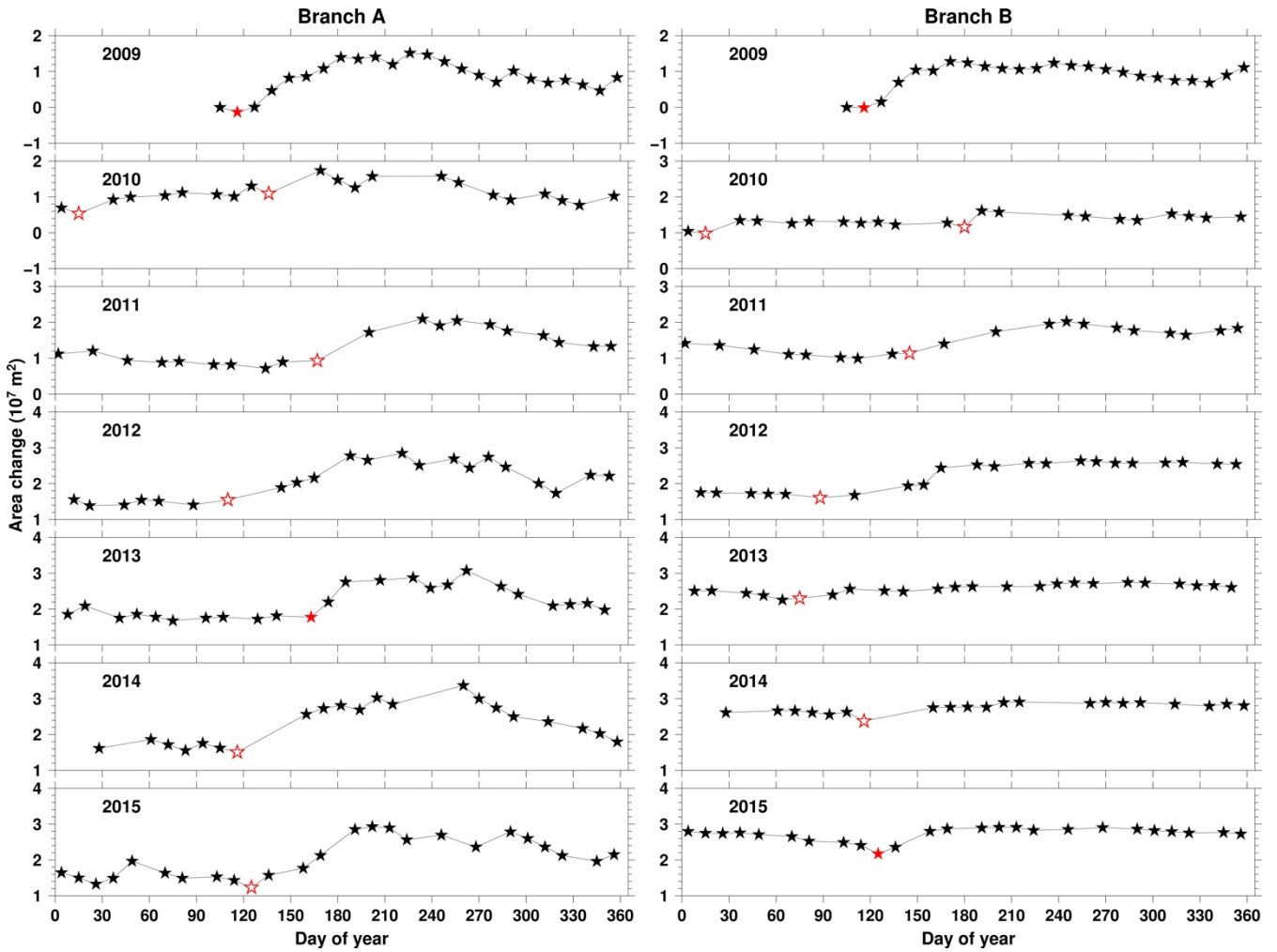
Figure 7. Similar to Fig. 5 but showing the time series of calving front changes (in area changes) of our deep-learning-based results in different years. The red filled stars mark the dates when the glacier started to retreat. Red open stars mark starting dates that cannot be reliably determined due to data gaps (e.g., Branch A in 2011) and small variations (e.g., Branch B in 2013).

5   **7 Discussion**

~~7.1~~**7.1 Differences from the previous work**

Mohajerani et al. (2019) have applied U-Net architecture to Landsat images over Jakobshavn, Sverdrup, Kangerlussuaq, and Helheim glaciers in Greenland. Despite both using the U-Net architecture, our study is different from Mohajerani et al. (2019) in datasets, result accuracy, transferability, strategies for classification, post-processing,

10   and image resampling. The usage of high-resolution TSX images allows us to generate more accurate calving fronts. Without additional manual practices, our method is more transferable, particularly when applying to large areas with many glaciers. Below we discuss the technical differences in detail.

15

First, our study classifies the surface into two types (i.e., ice mélange and non-ice mélange) to extract the calving front, while Mohajerani et al. (2019) used semantic segmentation to extract the front without classifying the surrounding surfaces. Both strategies require post-processing procedures. In our method, erroneous segmentation can cause small isolated polygons within the ice mélange or the non-ice mélange regions. Yet, we can solve this problem by removing these small polygons in the post-processing. The semantic segmentation used by Mohajerani et al. (2019) can be affected by icebergs, crevasses, etc. Nonetheless, the least-cost path search method could solve this problem (Mohajerani et al., 2019). Second, additional manual practices are needed in the work of Mohajerani et al. (2019). For instance, images of every single glacier in their work were adjusted by a certain angle to make all the glaciers flow in the same direction in the pre-processing. Third, we subdivide the images into small patches, which allows us to utilize the advantages of images with high resolution and various sizes. Mohajerani et al. (2019) resampled images to a fixed size (240 by 152 pixels) with low spatial resolution (49.0 to 88.1 meters), therefore the position accuracy is limited. They did so to save computation power. However, if applying their method to high-resolution images, they need to change the network's structure, and the size of input images would be limited by the GPU memory.

7.2 Calving front variation and bed elevation

In general, calving front variations are influenced by multiple factors, including floating or grounding conditions (McFadden et al., 2011; Murray et al., 2015; Bondzio et al., 2017; Fried et al., 2018), interaction with the ocean (Holland et al., 2008; Howat et al., 2008; Motyka et al., 2011; Vieli and Nick, 2011; Straneo et al., 2013), ice mélange and sea ice conditions (Amundson et al., 2010; Moon et al., 2015; Cassotto et al., 2015), basal lubrication (Joughin et al., 2008b; Moon et al., 2014) and bed elevation (Joughin et al., 2008a; Joughin et al., 2014; Kehrl et al., 2017; Bunce et al., 2018). Here we examine the possible link between the observed variations of the calving fronts with bed elevation.

Bed elevation has a substantial influence on the glacier retreat. In the first situation where the bed is flat, glacier retreat decreases resisting force, which accelerates the glacier. The acceleration of the glacier can also thin the ice. Thinning reduces the effective pressure at the bed, $N = P_i - P_w$, where P_i is the overburden pressure and P_w is water pressure. A decreased N reduces basal drag, causing stretching and faster flow and constituting positive feedback. In the second situation, as the glacier retreats into an overdeepened basin where the bed slopes down inland or is retrograde, the positive feedback is reinforced, and the glacier becomes more unstable, for two reasons. First, ice thickness at the calving front increases as the retreat progresses, increasing driving stress. Second, because the calving front moves into deeper water, this retreat decreases N further. In the third situation where the bed slopes inland-uphill, the glacier may stabilize, since retreating into shallower water increases N and decreases driving stress.

Previous studies also suggest that bed elevation has a substantial influence on glacier calving front variations. Examining the height above flotation of Branch A in Jakobshavn, Joughin et al. (2014) suggested that retreating into an overdeepened basin where the bed slope is retrograde may lead to an unstable calving front retreat, and a bed sloping inland-uphill may stabilize

the glacier. Other studies have also suggested that retreating into deeper water may accelerate the glacier, resulting in an unstable retreating (Howat et al., 2005; Howat et al., 2007; Nick et al., 2009; Catania et al., 2018).

In our study area, the bed elevation derived from BedMachine v3 (Morlighem et al., 2017) shows two overdeepened basins along the main channel of Branch A (Fig. 8a). During the period from 2009 to 2015, the calving front of Branch A retreated

5   into the second overdeepened basin in August 2011 for the first time, which may have ~~produce~~**produced** a faster rate of retreat. In July 2012, the glacier retreated to the bottom of the overdeepened basin and stopped retreating further (Movie S2). The inland-uphill bed slope behind the bottom of the overdeepened basin may have prevented the glacier from further retreating. In Branch B, after June 2012, the glacier retreated into a zone where the bed slopes uphill inland (Fig. 8b, Movie S3). We suggest that retreating into this zone may have led to the more regular and stable behavior of Branch B after June 2012 (Fig.
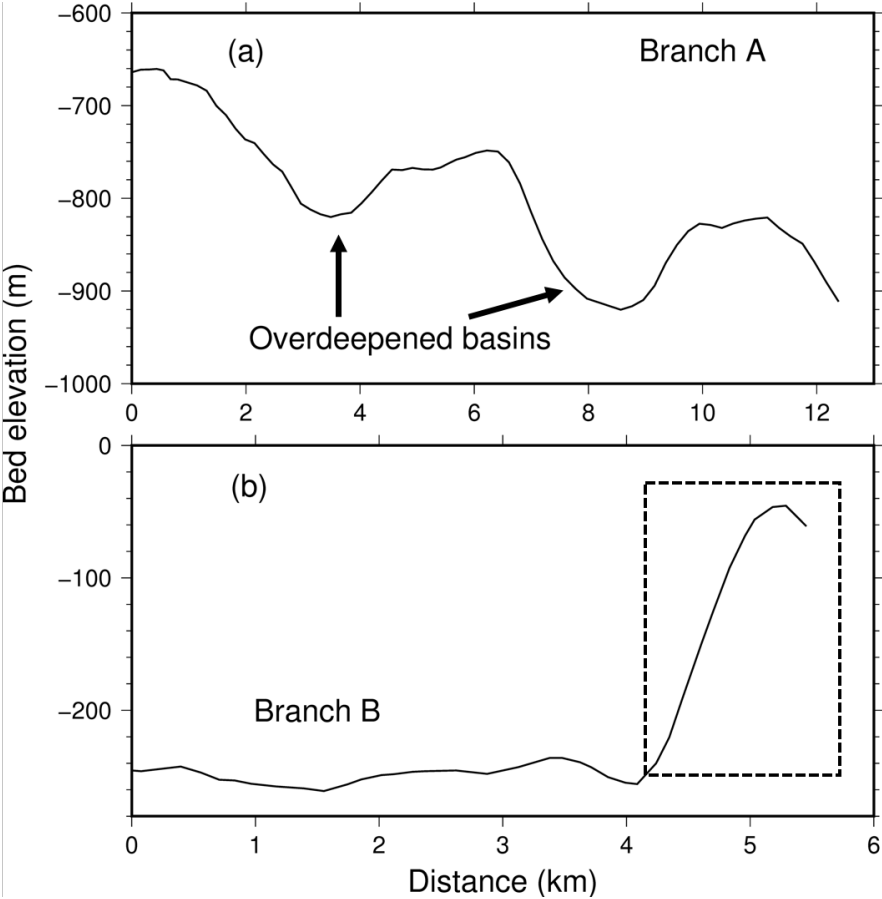
10   5b).

Figure 8. Bed elevation profiles of two branches derived from BedMachine v3 (Morlighem et al., 2017). The profile locations are shown in Fig. 1a. The dashed box shows the zone where the bed slopes uphill inland.

7.3 Limitations of current method

5 The current method is limited by high computational power requirement, and manual delineation largely control its accuracy. First, the U-Net architecture requires relatively high GPU memory for large images. In our configuration, around 15 gigabyte (GB) GPU memory is needed for training the network. Second, although splitting images with overlaps allow as to apply the network to images with different sizes, the overlaps increase the training time. These two limitations can be overcome by hardware development. With more powerful GPU in the future, we can increase the

10 calculation efficiency and lessen the training time. Third, the accuracy of this method relies on manual delineation as well as the information richness of the training dataset (Goodfellow et al., 2016). If the training examples are not representative for the actual task or if the manual delineation in these examples is of low quality or inconsistent, U-Net will either fail to train or will reproduce inconsistent results on new data. To further increase the accuracy and robustness of the network, more training examples are needed.

15 ## 7.4 Prospects for future work

In the near future, we will include more training examples to minimize network error. In this study, the well-trained network is limited to a specific dataset, namely TSX images. However, it is feasible to apply the DCNN to multi-sensor remote sensing imagery, which has been proved by previous studies (Nogueira et al. 2017; Lang et al., 2018). Moreover, as long as the calving fronts are clear in the images, our method can also use images with light cloud cover and Landsat 7 images

20 with scan line errors.

The effectiveness and transferable nature of deep learning (Anantrasirichai et al., 2018) promises that our methodology can be applied to many other glaciers, both in Greenland and elsewhere in the world. Besides Jakobshavn Isbræ, other Greenland tidewater glaciers such as Helheim and Kangerdlugssuaq also show strong calving front variations (Howat et al., 2005; Howat et al., 2007; Joughin et al., 2008a). In theory, the DCNN can be retrained whenever new data is added to the training dataset.

25 Moreover, including more data over other places can increase the generalization of the network, making it applicable to more situations (Goodfellow et al., 2016).

8 Conclusions

This study designs a method based on DCNNs to automatically delineate calving fronts of Jakobshavn Isbræ from TerraSAR-X SAR images acquired from April 2009 to December 2015. Small test error suggests that the accuracy of a well-trained

30 network can be close to the human level. Our results reveal that the two branches of Jakobshavn Isbræ behaved asynchronously. We suggest that bed elevation may have a major influence on the observed calving front variations. Our methodology can be

applied to many other tidewater glaciers both in Greenland and elsewhere in the world using multi-temporal and multi-sensor remote sensing imagery.

Code and data availability

The Greenland Ice Sheet CCI products are available from http://products.esa-icesheets-cci.org. The Bed elevation
5 (BedMachine v3) is available from http://sites.uci.edu/morlighem/dataproducts/bedmachine-greenland. The code of the whole framework (Fig. 3) will be provided by Enze Zhang upon request. The network-delineated calving fronts obtained in this study are available from https://doi.pangaea.de/10.1594/PANGAEA.897064 (Zhang et al., 2019).

Author contribution

EZ developed the code, performed the TSX images analysis and calving front quantification, and wrote the manuscript. LL
10 advised EZ and edited the manuscript. LH actively helped to develop the code.

Competing interest.

The authors declare that they have no conflict of interest.

References

20 Amundson, J. M., Fahnestock, M., Truffer, M., Brown, J., Lüthi, M. P., and Motyka, R. J.: Ice mélange dynamics and implications for terminus stability, Jakobshavn Isbræ, Greenland, J. Geophys. Res.-Earth, 115, F01 005, https://doi.org/doi:10.1029/2009JF001405, 2010.

Anantrasirichai, N., Biggs, J., Albino, F., Hill, P., and Bull, D.: Application of Machine Learning to Classification of Volcanic Deformation in Routinely Generated InSAR Data, J. Geophys. Res.-Sol. Ea., 123, 6592–6606,
25 https://doi.org/10.1029/2018JB015911, 2018.

Batista, G. E. A. P. A., Prati, R. C., and Monard, M. C.: Balancing Strategies and Class Overlapping, in: Lect. Notes Comput. Sc., pp. 24–35, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

Bondzio, J. H.: Analysis of Recent Dynamic Changes of Jakobshavn Isbrae, West Greenland, Using a Thermomechanical Model, Ph.D. thesis.

5 Bondzio, J. H., Morlighem, M., Seroussi, H., Kleiner, T., Rückamp, M., Mouginot, J., Moon, T., Larour, E. Y., and Humbert, A.: The mechanisms behind Jakobshavn Isbræ's acceleration and mass loss: A 3-D thermomechanical model study, Geophys. Res. Lett., 44, 6252–6260, https://doi.org/10.1002/2017GL073309, 2017.

Bunce, C., Carr, J. R., Nienow, P. W., Ross, N., and Killick, R.: Ice front change of marine-terminating outlet glaciers in northwest and southeast Greenland during the 21st century, J. Glaciol., 64, 523–535, https://doi.org/10.1017/jog.2018.44, 2018.

10 Cassotto, R., Fahnestock, M., Amundson, J. M., Truffer, M., and Joughin, I.: Seasonal and interannual variations in ice melange and its impact on terminus stability, Jakobshavn Isbræ, Greenland, J. Glaciol., 61, 76–88, https://doi.org/10.3189/2015JoG13J235, 2015.

Catania, G. A., Stearns, L. A., Sutherland, D. A., Fried, M. J., Bartholomaus, T. C., Morlighem, M., Shroyer, E., and Nash, J.: Geometric Controls on Tidewater Glacier Retreat in Central Western Greenland, J. Geophys. Res.-Earth, 123, 2024–2038, 15 https://doi.org/10.1029/2017JF004499, 2018.

Fahnestock, M., Bindschadler, R., Kwok, R., and Jezek, K.: Greenland Ice Sheet Surface Properties and Ice Dynamics from ERS-1 SAR Imagery, Science, 262, 1530–1534, https://doi.org/10.1126/science.262.5139.1530, 1993.

Fried, M. J., Catania, G. A., Stearns, L. A., Sutherland, D. A., Bartholomaus, T. C., Shroyer, E., and Nash, J.: Reconciling Drivers of Seasonal Terminus Advance and Retreat at 13 Central West Greenland Tidewater Glaciers, J. Geophys. Res.-Earth, 20 123, 1590–1607, https://doi.org/10.1029/2018JF004628, 2018.

Geng, J., Fan, J., Wang, H., Ma, X., Li, B., and Chen, F.: High-Resolution SAR Image Classification via Deep Convolutional Autoencoders, IEEE Geosci. Remote S., 12, 2351–2355, https://doi.org/10.1109/LGRS.2015.2478256, 2015.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, 2016.

Holland, D. M., Thomas, R. H., De Young, B., Ribergaard, M. H., and Lyberth, B.: Acceleration of Jakobshavn Isbrae triggered 25 by warm subsurface ocean waters, Nat. Geosci., 1, 659, https://doi.org/10.1038/ngeo316, 2008.

Howat, I. M., Joughin, I., Tulaczyk, S., and Gogineni, S.: Rapid retreat and acceleration of Helheim Glacier, east Greenland, Geophys. Res. Lett., 32, L22 502, https://doi.org/10.1029/2005GL024737, 2005.

Howat, I. M., Joughin, I., and Scambos, T. A.: Rapid Changes in Ice Discharge from Greenland Outlet Glaciers, Science, 315, 1559–1561, https://doi.org/10.1126/science.1138478, 2007.

30 Howat, I. M., Joughin, I., Fahnestock, M., Smith, B. E., and Scambos, T. A.: Synchronous retreat and acceleration of southeast Greenland outlet glaciers 2000–06: ice dynamics and coupling to climate, J. Glaciol., 54, 646–660, https://doi.org/10.3189/002214308786570908, 2008.

Howat, I. M., Ahn, Y., Joughin, I., van den Broeke, M. R., Lenaerts, J. T. M., and Smith, B.: Mass balance of Greenland's

three largest outlet glaciers, 2000–2010, Geophys. Res. Lett., 38, L12 501, https://doi.org/10.1029/2011GL047565, 2011.

Huang, Z., Pan, Z., and Lei, B.: Transfer Learning with Deep Convolutional Neural Network for SAR Target Classification with Limited Labeled Data, Remote Sens-Basel, 9, 907, https://doi.org/10.3390/rs9090907, 2017.

Joughin, I., Abdalati, W., and Fahnestock, M.: Large fluctuations in speed on Greenland's Jakobshavn Isbrae glacier, Nature, 432, 608, https://doi.org/doi.org/10.1038/nature03130, 2004.

Joughin, I., Howat, I., Alley, R. B., Ekstrom, G., Fahnestock, M., Moon, T., Nettles, M., Truffer, M., and Tsai, V. C.: Ice-front variation and tidewater behavior on Helheim and Kangerdlugssuaq Glaciers, Greenland, J. Geophys. Res.-Earth, 113, F01004, https://doi.org/10.1029/2007JF000837, 2008a.

Joughin, I., Das, S. B., King, M. A., Smith, B. E., Howat, I. M., and Moon, T.: Seasonal Speedup Along the Western Flank of the Greenland Ice Sheet, Science, 320, 781–783, https://doi.org/10.1126/science.1153288, 2008b.

Joughin, I., Howat, I. M., Fahnestock, M., Smith, B., Krabill, W., Alley, R. B., Stern, H., and Truffer, M.: Continued evolution of Jakobshavn Isbrae following its rapid speedup, J. Geophys. Res.-Earth, 113, F04 006, https://doi.org/10.1029/2008JF001023, 2008c.

Joughin, I., Smith, B. E., Howat, I. M., Floricioiu, D., Alley, R. B., Truffer, M., and Fahnestock, M.: Seasonal to decadal scale variations in the surface velocity of Jakobshavn Isbrae, Greenland: Observation and model-based analysis, J. Geophys. Res.-Earth, 117, F02 030, https://doi.org/10.1029/2011JF002110, 2012.

Joughin, I., Smith, B. E., Shean, D. E., and Floricioiu, D.: Brief Communication: Further summer speedup of Jakobshavn Isbrae, The Cryosphere, 8, 209–214, https://doi.org/10.5194/tc-8-209-2014, 2014.

Kehrl, L. M., Joughin, I., Shean, D. E., Floricioiu, D., and Krieger, L.: Seasonal and interannual variabilities in terminus position, glacier velocity, and surface elevation at Helheim and Kangerlussuaq Glaciers from 2008 to 2016, J. Geophys. Res.-Earth, 122, 1635–1652, https://doi.org/10.1002/2016JF004133, 2017.

King, M. D., Howat, I. M., Jeong, S., Noh, M. J., Wouters, B., Noël, B., and van den Broeke, M. R.: Seasonal to decadal variability in ice discharge from the Greenland Ice Sheet, The Cryosphere, 12, 3813–3825, https://doi.org/10.5194/tc-12-3813-2018, 2018.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, in: ICLR, http://arxiv.org/abs/1412.6980, 2015.

Lang, G., Tian, T., Xiao, S., and Hang, L.: Fusion of shallow and deep features for classification of high-resolution remote sensing images, in: MIPPR 2017: Multispectral Image Acquisition, Processing, and Analysis, vol. 10607, p. 106070O, https://doi.org/10.1117/12.2284777, 2018.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436, https://doi.org/10.1038/nature14539, 2015.

Luckman, A. and Murray, T.: Seasonal variation in velocity before retreat of Jakobshavn Isbræ, Greenland, Geophys. Res. Lett., 32, L08 501, https://doi.org/10.1029/2005GL022519, 2005.

Maas, A. L., Hannun, A. Y., and Ng, A. Y.: Rectifier nonlinearities improve neural network acoustic models, in: ICML Workshop on Deep Learning for Audio, Speech and Language Processing, 2013.

Meier, M. F. and Post, A.: Fast tidewater glaciers, J. Geophys. Res. Sol. Ea., 92, 9051–9058,

~~https://doi.org/10.1029/JB092iB09p09051, 1987.~~

McFadden, E. M., Howat, I. M., Joughin, I., Smith, B. E., and Ahn, Y.: Changes in the dynamics of marine terminating outlet glaciers in west Greenland (2000–2009), J. Geophys. Res.-Earth, 116, F02022, https://doi.org/10.1029/2010JF001757, 2011.

Meier, M. F. and Post, A.: Fast tidewater glaciers, J. Geophys. Res.-Sol. Ea., 92, 9051–9058,

5   https://doi.org/10.1029/JB092iB09p09051, 1987.

Mohajerani, Y., Wood, M., Velicogna, I., and Rignot, E.: Detection of Glacier Calving Margins with Convolutional Neural Networks: A Case Study, Remote Sens, 11, 74, https://doi.org/10.3390/rs11010074, 2019.

Moon, T. and Joughin, I.: Changes in ice front position on Greenland's outlet glaciers from 1992 to 2007, J. Geophys. Res.-Earth, 113, F02 022, https://doi.org/10.1029/2007JF000927, 2008.

10   Moon, T., Joughin, I., Smith, B., van den Broeke, M. R., van de Berg, W. J., Noël, B., and Usher, M.: Distinct patterns of seasonal Greenland glacier velocity, Geophys. Res. Lett., 41, 7209–7216, https://doi.org/10.1002/2014GL061836, 2014.

Moon, T., Joughin, I., and Smith, B.: Seasonal to multiyear variability of glacier surface velocity, terminus position, and sea ice/ice mélange in northwest Greenland, J. Geophys. Res.-Earth, 120, 818–833, https://doi.org/10.1002/2015JF003494, 2015.

Morlighem, M., Williams, C. N., Rignot, E., An, L., Arndt, J. E., Bamber, J. L., Catania, G., Chauché, N., Dowdeswell, J. A.,

15   Dorschel, B., Fenty, I., Hogan, K., Howat, I., Hubbard, A., Jakobsson, M., Jordan, T. M., Kjeldsen, K. K., Millan, R., Mayer, L., Mouginot, J., Noël, B. P. Y., O'Cofaigh, C., Palmer, S., Rysgaard, S., Seroussi, H., Siegert, M. J., Slabon, P., Straneo, F., van den Broeke, M. R., Weinrebe, W., Wood, M., and Zinglersen, K. B.: BedMachine v3: Complete Bed Topography and Ocean Bathymetry ~~Map- ping~~**Mapping** of Greenland From Multibeam Echo Sounding Combined With Mass Conservation, Geophys. Res. Lett., 44, 11,051–11,061, https://doi.org/10.1002/2017GL074954, 2017.

20   Murray, T., Selmes, N., James, T. D., Edwards, S., Martin, I., O'Farrell, T., Aspey, R., Rutt, I., Nettles, M., and Baugé, T.: Dynamics of glacier calving at the ungrounded margin of Helheim Glacier, southeast Greenland, J. Geophys. Res.-Earth, 120, 964–982, https://doi.org/10.1002/2015JF003531, 2015.

Motyka, R. J., Truffer, M., Fahnestock, M., Mortensen, J., Rysgaard, S., and Howat, I.: Submarine melting of the 1985 Jakobshavn Isbræ floating tongue and the triggering of the current retreat, J. Geophys. Res.-Earth, 116, F01 007,

25   https://doi.org/10.1029/2009JF001632, 2011.

Nick, F. M., Vieli, A., Howat, I. M., and Joughin, I.: Large-scale changes in Greenland outlet glacier dynamics triggered at the terminus, Nat. Geosci., 2, 110, https://doi.org/10.1038/ngeo394, 2009.

Nogueira, K., Penatti, O. A., and dos Santos, J. A.: Towards better exploiting convolutional neural networks for remote sensing scene classification, Pattern Recogn., 61, 539–556, https://doi.org/https://doi.org/10.1016/j.patcog.2016.07.001, 2017.

30   Rignot, E., Koppes, M., and Velicogna, I.: Rapid submarine melting of the calving faces of West Greenland glaciers, Nat. Geosci., 3, 187, https://doi.org/10.1038/ngeo765, 2010.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Lect. Notes Comput. Sc., edited by Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., pp. 234–241, Springer International Publishing, Cham, 2015.

Rosenau, R., Schwalbe, E., Maas, H.-G., Baessler, M., and Dietrich, R.: Grounding line migration and high-resolution calving dynamics of Jakobshavn Isbræ, West Greenland, J. Geophys. Res.-Earth, 118, 382–395, https://doi.org/10.1029/2012JF002515, 2013.

Roth, A., Huber, M., and Kosmann, D.: Geocoding of TerraSAR-X data, Proc. Int. 20th ISPRS Congr, pp. 840–844, 2004.

5 Seale, A., Christoffersen, P., Mugford, R. I., and O'Leary, M.: Ocean forcing of the Greenland Ice Sheet: Calving fronts and patterns of retreat identified by automatic satellite monitoring of eastern outlet glaciers, J. Geophys. Res.-Earth, 116, F03 013, 2011.

Sohn, H.-G. and Jezek, K. C.: Ice sheet margin detection using ERS-1 synthetic aperture radar, in: Int. Geosci. Remote Se., vol. 1, pp. 148–150, https://doi.org/10.1109/IGARSS.1996.516273, 1996.

10 Sohn, H.-G., Jezek, K. C., and van der Veen, C. J.: Jakobshavn Glacier, west Greenland: 30 years of spaceborne observations, Geophys. Res. Lett., 25, 1998.

Straneo, F. and Heimbach, P.: North Atlantic warming and the retreat of Greenland's outlet glaciers, Nature, 504, 36, https://doi.org/10.1038/nature12854, 2013.

Sun, Y., Wang, X., and Tang, X.: Deep learning face representation from predicting 10,000 classes, in: Proceedings of the
15 IEEE conference on computer vision and pattern recognition, pp. 1891–1898, 2014.

Thomas, R. H.: Force-perturbation analysis of recent thinning and acceleration of Jakobshavn Isbræ, Greenland, J. Glaciol., 50, 57–66, https://doi.org/10.3189/172756504781830321, 2004.

Vieli, A. and Nick, F. M.: Understanding and Modelling Rapid Dynamic Changes of Tidewater Outlet Glaciers: Issues and Implications, Surv. Geophys., 32, 437–458, https://doi.org/10.1007/s10712-011-9132-4, 2011.

20 Xie, S., Dixon, T. H., Voytenko, D., Holland, D. M., Holland, D., and Zheng, T.: Precursor motion to iceberg calving at Jakobshavn Isbræ, Greenland, observed with terrestrial radar interferometry, J. Glaciol., 62, 1134–1142, https://doi.org/10.1017/jog.2016.104, 2016.

Zhang, E., Liu, L., Huang, L.: Calving front and the comparison between its variations and bed elevation in Jakobshavn Isbræ, https://doi.pangaea.de/10.1594/PANGAEA.897066, 2019.

25 Zhang, F., Du, B., and Zhang, L.: Saliency-Guided Unsupervised Feature Learning for Scene Classification, IEEE T. Geosci Remote, 53, 2175–2184, https://doi.org/10.1109/TGRS.2014.2357078, 2015.