

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.

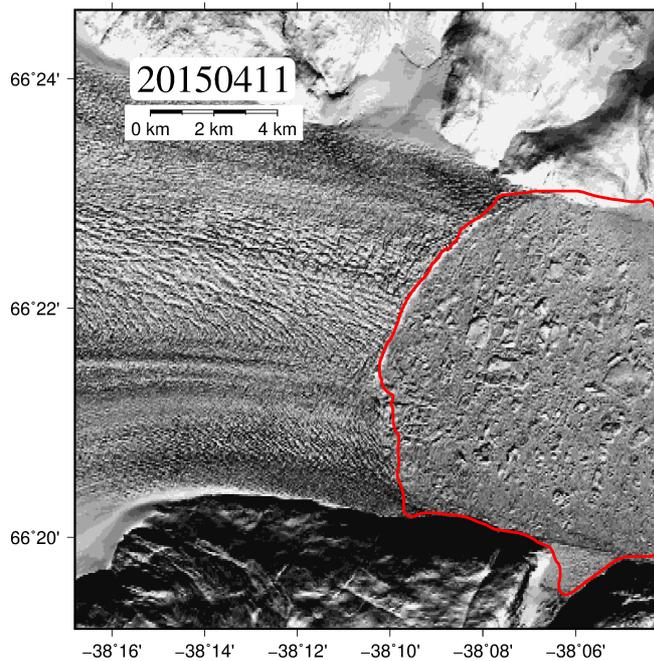


Figure R1. An example of automatically delineated calving front at Helheim. The background image is a Landsat 8 image taken on April 11th, 2015. The red line indicates the automatically delineated calving front.

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.

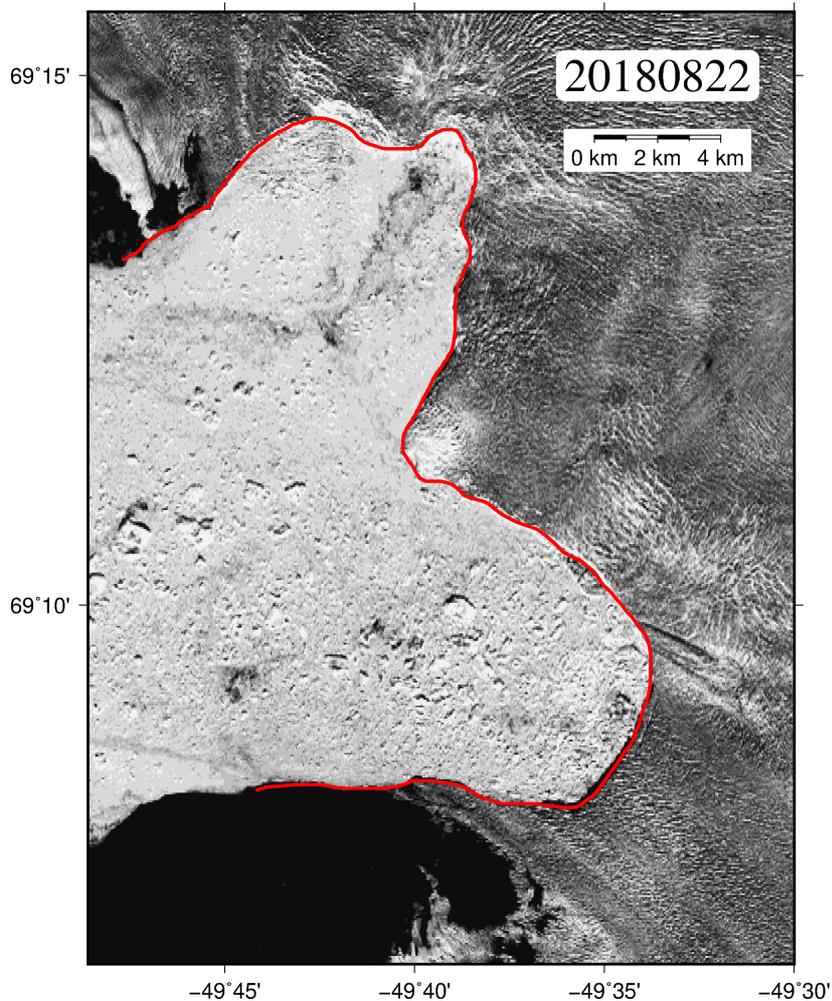


Figure R2. An example of automatically delineated calving front at Jakobshavn Isbrae using a Landsat 8 image taken on August 22nd, 2018.

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.

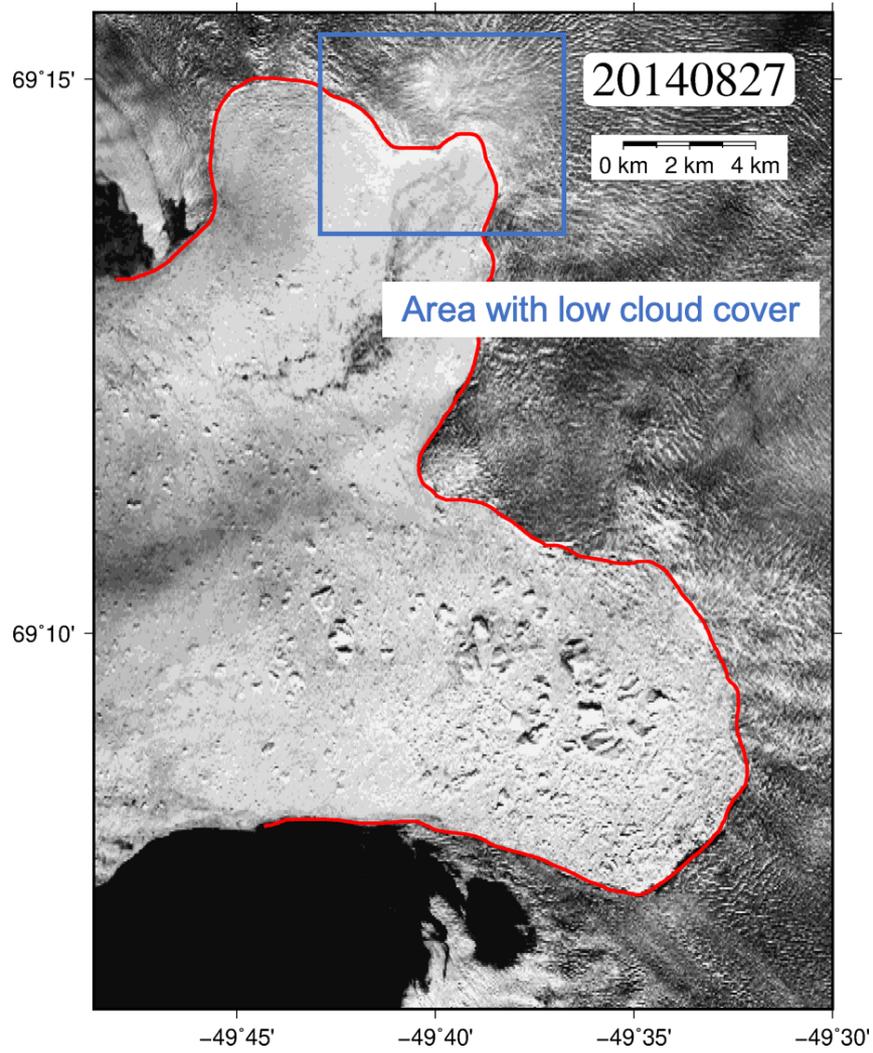


Figure R3. An example of automatically delineated calving front at Jakobshavn Isbrae using a Landsat 8 image with clouds. The image was taken on August 27th, 2014. The blue box indicates an area with low cloud cover.

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.

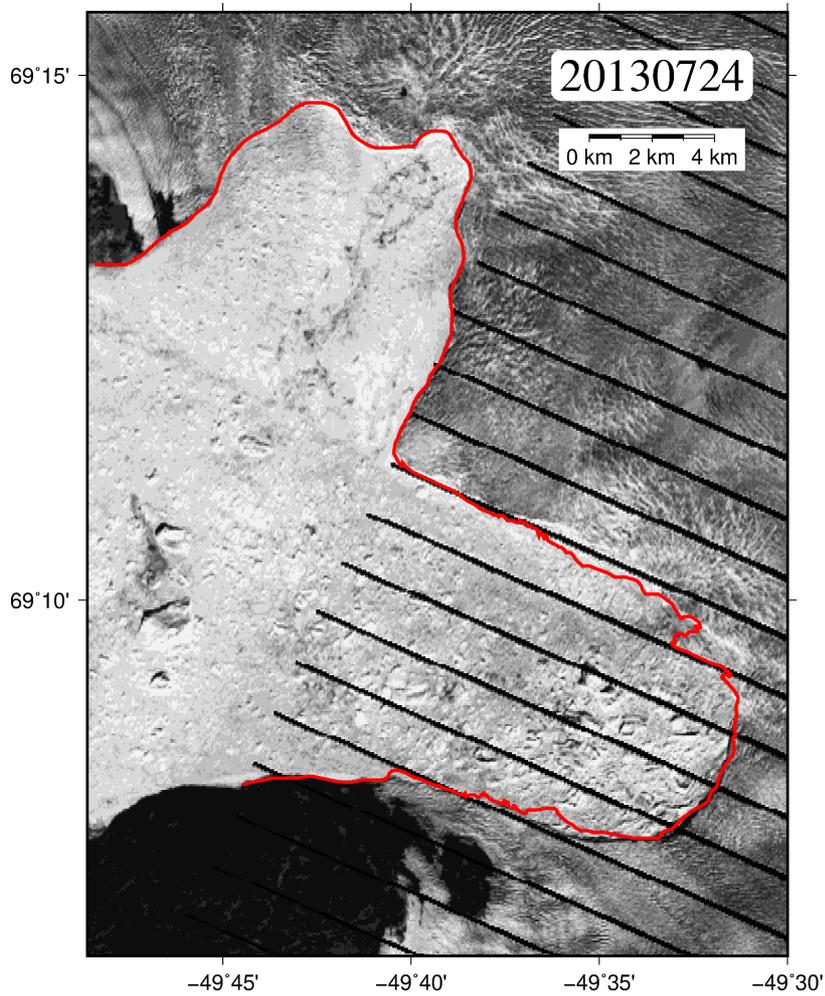


Figure R4. An example of automatically delineated calving front at Jakobshavn Isbrae using a Landsat 7 image with scan line errors. The image was taken on July 24th, 2013.