The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2019-14-RC2, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

The Cryosphere
Discussions
Open Access



# *Interactive comment on* "Automatically delineating the calving front of Jakobshavn Isbræ from multi-temporal TerraSAR-X images: a deep learning approach" *by* Enze Zhang et al.

**Anonymous Referee #2**

Received and published: 20 March 2019

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:





CC BY

MAJOR COMMENTS:

1. 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.

2. 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.

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.

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).

3. There is no proper measure of the extent of overfitting in the study. Without a





CC BY

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.

4. 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.

5. 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.?

6. 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?

MINOR COMMENTS:

Page 1 Line 16: add “to” after “stabilized”.

Page 3 Line 13: change “speeded up” to “sped up”

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

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

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?

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

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





CC BY

glacier here?

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

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.

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

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

Section 7.2: What are the limitations of the current technique? Could imagery artifacts or more varied surfaces be dealt with? Can the trained network be applied to multiple glaciers or does it have to be retrained for every glacier?

---

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-14, 2019.





Printer-friendly version

Discussion paper