Review on Herrmann et al 2023: Out-of-the-box calving front detection method using deep learning

Herrmann et al. use a nnU-Net to extract the calving front of marine-terminating glaciers. To train the network they used a benchmark dataset published by Gourmelon et al. 2022 including images from six different radar sensors for seven marine-terminating glaciers. They perform six different experimental set-ups to test the best label and network set-up to achieve the highest accuracy compared to manual delineated fronts. The manuscript contains a large variety of figures. The visualizations are appealing and help to understand the study design but sometimes more profound explanations would be desirable (especially in the discussion). The manuscript discusses some interesting experiments on model set-up, labels and effect of seasonality and sensor type for the accuracy of glacier front detection. This is interesting for DL-model developers and remote sensing scientists but I would expect more added value for glaciologists as well (especially when publishing in The Cryosphere). For example, this could include the out-of-the box usage of your repository for own radar images to generate time-series for an area of interest for glaciological studies (maybe this is already possible but I cannot access the GitLab link provided in the manuscript).

Some further comments:

- So far, the results & discussion section is a bit difficult to follow and very short. Think about providing sub-headers and providing a more in-depth discussion.
- Addressing the seasonal effects on accuracy is very interesting. But a more sophisticated analysis would be required. In Figure 10 it looks like that the sensor type influences the accuracy much more than the season. Therefore, it would be interesting to know the sensor type for the scenes considered for winter and summer. Especially for Columbia glacier as only TSX and S1 scenes were available and for each sensor the MDE is very different.
- Provide an analysis on the effect of different spatial resolutions on the accuracy of the front extraction. Also introduce the different spatial resolutions and polarizations of all sensors as not every reader might be familiar with that.
- The conclusion is a bit unstructured and could be much stronger. Emphasize the identified best practice approach and give a better outlook for your study and applications in glaciology.

Line specific comments:

L16: The GitLab repository is only accessible for people affiliated to the Friedrich-Alexander-University. Please publish your scripts on a platform easily accessible for everyone.

L45: "Most of these methods use optical imagery": Three out of four cited references here use SAR imagery. Please provide proper reference for your statement.

L66: What means "visual preparation". I would assume a benchmark dataset can be used right away.

L76: Please do not only state that many methods are based on the U-Net from Ronneberger but also provide references for these studies/methods and mention that they mostly used modified versions of the original U-Net (e.g. Loebel et al. 2022, Mohajerani et al. 2019, Zhang et al. 2019 etc.)

L98: In the study of Heidler et al. 2021 the digital elevation model was used during training for different experiments. A further development of the HED-UNet for a circum-Antarctic approach (Baumhoer et al. 2023) uses indeed a DEM in the post-processing.

L145ff: If the network itself decides on batch and patch size I wonder how big the effect on the accuracy is. For example, a smaller patch size would provide less spatial context and probably decrease the accuracy. Did you do any experiments on the effect of batch and patch size on the model accuracy? Furthermore, please state on what infrastructure (GPU type, RAM etc.) your model was trained and what patch and batch size was finally used.

L225: Maybe add a sentence why you propose that the fused label approach is the best even though the boundary label approach has better accuracies.

L245: Dry snow is penetrated by active microwave sensors and the ground can be the major source of the backscatter signal (depending on snowpack thickness and wavelength). Wet snow can be detected by SAR sensors. Hence, is snow cover really the reason for less accurate front predictions in winter? Additionally, I would assume that surface melt in summer and sea ice in winter make the front prediction more difficult. Did you investigate this further?

L228: Why did you decide to use the ensemble prediction results for the final evaluation instead of doing this evaluation for the most feasible method you identified prior to that (fused label approach)? Wouldn't it be more likely for users to use this 'best' method instead of the ensemble approach requiring much more computational power and effort?

L248: You can not directly compare the accuracies for both sensors. For ERS, Envisat and Palsar you only have data for Mapple glacier which seems to have a front that is better captured by the nn-UNet. The accuracy comparison for different sensors is very interesting especially regarding spatial resolution and polarization but this would require a more sophisticated comparison.

L250: I would recommend to investigate further why the results are worse for Sentinel-1. Maybe provide some plots in the appendix on these in-accurate results plotted over the Sentinel-1 image to see the sea ice conditions at that time.

Figure 10: Please properly label the X-axis for the logarithmic scale. Consider splitting the figure in two or insert a little gap between the seasonal analysis and the comparison of accuracy for different satellite sensors. For ERS, Envisat and Palsar you need to mention that this data is for Mapple only.

Figure 11c: The yellow front seems to be a circle. One part of the prediction fits very well whereas the other is completely off. How is this considered in the accuracy estimation?

L269: Rather "increase" model performance I guess?! To which sensor do you refer when talk about low resolution imagery?

Literature mentioned within this review:

Baumhoer, C. A., Dietz, A. J., Heidler, K., and Kuenzer, C.: IceLines – A new data set of Antarctic ice shelf front positions, Sci Data, 10, 138, https://doi.org/10.1038/s41597-023-02045-x, 2023.

Loebel, E., Scheinert, M., Horwath, M., Heidler, K., Christmann, J., Phan, L. D., Humbert, A., and Zhu, X. X.: Extracting glacier calving fronts by deep learning: the benefit of multi-spectral, topographic and textural input features, IEEE Transactions on Geoscience and Remote Sensing, 1–1, https://doi.org/10.1109/TGRS.2022.3208454, 2022.

Mohajerani, Y., Wood, M., Velicogna, I., and Rignot, E.: Detection of Glacier Calving Margins with Convolutional Neural Networks: A Case Study, Remote Sensing, 11, 1–13, https://doi.org/10.3390/rs11010074, 2019.

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