Review on „Image Classification of Marine-Terminating Outlet Glaciers using Deep Learning Methods” by Marochov et al.

Marochov et al. made a huge effort to revise their manuscript and to address the reviewer’s comments in every detail. The manuscript improved particularly with regard to more accurate explanations on technical details, additional test data and the implementation of an approach to extract the calving front. Additionally, the edge classification problem was addressed accurately and is clear now. Nevertheless, a few major concerns remain which are outlined below.

1. The authors added two additional test sites and more test scenes covering a wider temporal variety. This really highlights the transferability of the developed approach. Could you please explain why you decided to remove the initial test site Scoresby Sund mentioned in the first version of the manuscript? As you already have the data it would be worth to include it as an additional test set.
2. The revised manuscript includes a lot of additional information on the developed approach and describes technical aspects in every detail. On the one hand, this is an advantage of the manuscript as the approach is very transparent. On the other hand, the manuscript has become rather long and focuses more on technical details. The authors have to be careful to not only provide methodical details on their approach but also to fit within the scope of The Cryosphere by addressing a broader cryospheric community.
   To make your manuscript more suitable for a wider cryospheric community I would recommend the following:
   a. Highlight the advantages of your approach for the cryospheric community. So far, the discussion is solely technical. But you could add one section discussing the future advantages of your approach for the analysis of the cryosphere (e.g. calving front detection, change detection of class distributions, snow cover changes between different years etc.)
   b. Highlight the great performance of your classification algorithm but try not to confuse the reader with too many details about performance differences. For example, consider to shift some of the plots to the supplementary material. You could just show the data for the optimal model configuration in the manuscript and keep the rest in the supplementary materials.
   c. Consider to shorten the text and densify the information on tile sizes, patches and different model configurations. For readers with no background in machine learning all those parameters might be confusing. Probably, a table including all those different parameters and the corresponding accuracies could help for a condensed and better overview.
   d. The authors put a lot of effort into comparing different model configurations (tile and patch size) which is highlighted in several figures and graphs. In my opinion, it would be worth to merge some of the figures to shorten the manuscript. Please see the suggestions in the technical corrections below.

Technical corrections:

Figure 1: This is a really nice figure and helps to understand your classification approach much better. Well done!

Figure 2: Nice idea to combine the class examples within this figure. Looks much nicer now.

Figure 6 & 7: Consider to merge those two figures into Figure 6a & 6b. It will be easier to see the differences between the cCNN and MPL approach.

Figures 11, 12, 13, 15, 16 & 17 demonstrate the classification results. The amount of figures might be a bit too heavy compared to the length of the paper. You could consider to merge the results into less figures or shift some results to the supplementary materials.

Figure 14: You could condense the information and use one plot showing all three glaciers in the same plot with the best model configuration (RGB+NIR, Single). The remaining part could be moved to the supplementary.

Figure 15d: It is interesting, that the model confuses mélange with glacier ice so heavily. Could you explain why this is the case?

L371: "trains to learn". Please re-phrase.

L486: What does "bergy" mean? Or just a typo.

L591: Why does the joint model provide higher classification accuracies but the single model higher accuracies for the calving front extraction? This seems to be contradictory.

L644: Here you mention that the developed approach might be suitable for lake mapping but earlier it is mentioned that the approach has difficulties with classes being smaller than the tile size. Are those lakes always large enough to be captured by your model? Additionally, the class "lake" is not included in the classification or is it defined as open water enclosed by glacier ice?

Supplement:

1. Why did you use additional Helheim scenes for the joint training method even though the single model was trained on Helheim anyways? (see scenes with * in Table S1)
2. Why do you provide only a confusion matrix for the single training but not the combined training approach? It would be interesting to see the performance differences to justify the necessity of a joint and single model approach. Moreover, I would assume that the most robust (spatially transferrable) model would be achieved by including several training areas from the beginning (instead of only Helheim) over different glaciers which would make the additional joint training unnecessary.

Please don't be discouraged by the length of my review. I know that a lot of work went into the revised version and the provided comments might require some further effort. Nevertheless, my comments are mostly suggestions and not a must hopefully helping to improve your manuscript.