The Cryosphere Discussions

The Cryosphere Discuss., referee comment RC1
https://doi.org/10.5194/tc-2022-61-RC1, 2022

Comment on tc-2022-61

Anonymous Referee #1

Referee comment on "Glacier extraction based on high-spatial-resolution remote-sensing images using a deep-learning approach with attention mechanism" by Xinde Chu et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2022-61-RC1, 2022

General Comments:

This manuscript describes the development of a deep learning methodology using neural networks and post processing of remote sensing imagery to automatically detect glacial boundaries. This work builds upon existing efforts in the field to address the labor-intensive task of surveying glacial changes over time for the purposes of further study. The method proposed uses an improved neural network design based on the Deeplabv3+ architecture to classify glacial pixels in the satellite imagery, and adds Convolutional Block Attention Modules to aid in identifying important glacier features during the automatic extraction task. The description of the method details is rigorous, which includes consideration such as the use of weighted Dice coefficient loss during neural network training, the use of image subsetting/patching to contend with computational limits/image constraints, and the use of Test Time Augmentation to produce a high confidence classification of glacial boundaries. The study evaluates this improved methodology on glaciers in the Tanggula and Kunlun Mountains using Gaogen-6 panchromatic/multispectral optical satellite imagery. The overall accuracy (99.58% out of 100%) and Kappa coefficient (0.9915 out of 1.0) of the resulting classifications is high, and while there are differences in the estimated glacial coverage between the method and existing glacial inventories, the study shows potential for further development and application.

Overall, the manuscript, its methodology, and its findings are sound. This study builds upon previous work, and it advances our understanding of automated glacial feature extraction, though it is limited in scope. There are some remarks to address and minor copy-editing which are listed below and under the specific comments. Given this, I recommend minor revisions with attention to comments, at the editor’s discretion.

Major comments:

1. One concern to raise is the choice of evaluation metrics. Existing literature in the field of automated glacial boundary extraction (Mohajerani et al. (2019, TC), Baumhoer et al. (2019, RS), Zhang et al. (2021, RSE), Cheng et al. (2021, TC), Robson et al. (2020), and He et al. (2021) use a wider variety of metrics not limited to OA and Kappa coeffeicients. More specifically, OA as a metric is subject to bias/skew depending on the test data, as small but important errors along glacial boundaries can be underrepresented given a large enough domain that is more easily classified. Most of the above also provide more robust accuracy metrics as the mean distance from the boundary in meters/pixels, and the Mean Intersection over Union. Thus, it is recommended that these two metrics be included for a more robust representation of errors along glacial boundaries, and allow for an easier comparison of this study's methodology with respect to others in the field.

2. Another concern is the scope/applicability of the technique to other glaciers, and under more difficult imaging conditions than just thin clouds/snow. While the methodology shows good results on the testing data, it is limited to 4 ideal images of 2 locations (Tanggula and Kunlun Mountains). Existing works (Cheng et al., 2021) have proven the applicability of similar methods to wider applications, but additional testing on different domains/image conditions/SAR data may be useful ensure that this neural network has not overtrained on the training data. Since this may be outside the scope of the study (as touched upon in the conclusion), this may be done at the authors' discretion, but it may be of interest to see how well the neural network model has generalized from the data it has trained on.

3. Section 4.2 provides a potentially valuable comparison between the study's glacial extent data product, and that of manually curated glacial extents. However, due to the time differences between the measurements (2020 for the study, 1999-2003 for GGI, and 2013-2018 for TPG2017), the comparison is not as useful as it could be. While this may also be out of scope for this study, it would be interesting to see a co-temporal comparison of glacial extents if such data exists. Otherwise, the differences in area/% changes make the potential errors from the methodology hard to separate from potential changes in the glaciers over time.

Specific Comments:

P2 L31: "researches of" -> "research on"

P2 L 37: "a relatively accurate results" -> "relatively accurate results"

P4 L82: "a upsampling" -> "an upsampling"

P4 L88: "internal correlation, its basic" -> "internal correlation. Its basic"

P18 L295: Zhang et al. (2019) and Cheng et al. (2021) both use single spectrum data inputs and don't utilize spectrum information, and already rely on texture/shape information.

P19 L305: "And," -> "Additionally,"

P20 L331: "2m that" -> "2m, such that", "more detail" -> "more detailed"

P20 L331- 335: Consider splitting these long sentences/rephrasing (i.e., "intact glaciers, although in a few cases…" -> "intact glaciers. However, in a few cases…"

P21 L347: "glacier extracted" -> "glacier extraction"

P21 L349: "extracting complete glacier" -> "extraction complete glaciers"

P21 L350: "And then comparison" -> "Comparison"

P21 L351: "which could distinguish glacier" -> "which could distinguish glaciers"