

The Cryosphere Discuss., referee comment RC2
<https://doi.org/10.5194/tc-2022-61-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on tc-2022-61

Anonymous Referee #2

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-RC2>, 2022

General comments

This paper explores the use of deep learning algorithms to map debris-free glaciers in Gaofen-6 PMS (pan/multispectral) imagery, which has 2-meter spatial resolution and lacks short-wave infrared (SWIR) bands. Previously employed glacier-mapping methods frequently rely on SWIR bands in instruments such as ASTER, Landsat, and Sentinel-2, because the reflectance of snow and ice is very low in that part of the spectrum. High-resolution (meter-scale) instruments tend not to have SWIR bands, as the authors point out, so if one needs to map (e.g. smaller) glaciers at high resolution, this method could be of great use.

A limitation of this method is that it is designed to work on clean (debris-free) glaciers. Since debris cover is present on 44% of Earth's glaciers (<https://www.nature.com/articles/s41561-020-0615-0>), this method is currently limited to regional studies where glaciers are largely clean. However, the authors point out that in the future, the method could use more diverse input data and be made to work on debris-covered glaciers. The method described in this paper does appear to be a solid building block for future expansion.

While the authors provide a generally good overview of previous work, one closely related paper that was not cited is from Xie, et al. (2020), entitled "GlacierNet: A Deep-Learning Approach for Debris-Covered Glacier Mapping". The application space is a bit different for this approach, but this paper seems important to mention when discussing prior work.

This paper makes extensive use of deep-learning jargon, and it should be better described for this journal's audience. The first time the "attention mechanism" is mentioned in the main text (line 58), for example, it should be better introduced, or pointed out that it will be described in the next section. But even after that, the method is not adequately

described. I think it would be good to briefly describe the concepts behind the jargon at first mention, for example, something like "ASPP is used to obtain multi-scale context information from the imagery." Otherwise, understanding of the gist of the article will be too reliant on the reader finding outside resources.

In summary, this paper describes a method that achieves good results in its currently limited domain, and appears to be a good building block for future extension of the method to other input data sets. I recommend publishing after significant revisions for 1) readability, and 2) adequate explanation of the concepts behind the algorithms.

Some specific comments:

- I think for most people, longitude/latitude coordinates given in decimal degrees are more useful than degrees-minutes-seconds.

- The paper would benefit from a go-through by a native (or near-native) speaker of English to correct usage of articles, plurals, etc.