

## ***Interactive comment on “Arctic sea ice cover data from spaceborne SAR by deep learning” by Yi-Ran Wang and Xiao-Ming Li***

**Anonymous Referee #1**

Received and published: 3 January 2021

This paper presents an interesting exploration of using U-NET to segment Sentinel-1 data sets. The paper provided detailed methodology on the preprocessing of the image data, the construction of the model, and the training of the model. The resulting model has an impressive accuracy of 90%+.

I work mainly in machine learning, so my questions for this paper are about its methodology.

While I understand the difficulty of obtaining ground truth data for Arctic sea ice, the evaluation against lower resolution images is still strange. The S1 data set has a resolution of 400m, whereas the AMSR2 and IMS have resolutions in km. Validating the model against low-resolution data would potentially mask some of the errors of the model. It also limits the impact of the paper – as the paper seems to be motivated by

C1

a need for high-resolution segmentation. What makes things worse was that the sea ice cover data from AMSR2 and IMS were also generated by other models, rather than real ground truth. This evaluation also shows a mismatch between the labels used for training and testing.

Can the authors validate the model against a higher resolution data set? Also, is it possible to validate using other S1 images with manual labels just as the authors did with the training data? In the paper, 1/3 of the 251 images were used as evaluation. It may be helpful to provide the evaluation results on those 1/3 images (I apologize if I missed it).

My other question is about the hyperparameters used in the paper, such as the "fixed thresholds of [2 dB, 7 dB] and [-2 dB, 3.5 dB]" and the batch size (there are quite a few more). How are these hyperparameter selected? Is there hyperparameter optimization?

A minor concern is the use of a small training data set (about 167 or 2/3 of 251). While U-Net was designed for a small training set, 167 seems to be too small, especially given that there are 28k images available. Will increase the training set or use a full CNN increase accuracy?

---

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-332>, 2020.

C2