Response to Reviewer

We would like to thank the reviewer for the detailed and helpful comments, suggestions, and careful checking. Comments are responded on point-by-point basis. The comments from the reviewer are in black while the authors' responses are in blue. Where we refer to line numbers, the revised version with changes highlighted in blue is applicable.

Overall comment by the reviewer: This paper presents an interesting exploration of using U-NET to segment Sentinel-1 datasets. 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.

Response:

Thank the reviewer for his/her positive comment on this study. The raised comments and questions are addressed below.

Comments 1: 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 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).

Response 1:

In our research, the Arctic sea ice cover product is generated based on more than 28,000 S1 images and we intend to evaluate the product case by case. However, there are no other available (at least, easily accessible) sea ice cover and concentration products that have comparable spatial resolution and time-space coverage to the S1 EW data, for the case-by-case comparison. So we dialed it back and evaluated our data by comparing with the AMSR2 sea ice concentration data and IMS sea ice cover data, which have a lower resolution but are both widely used for sea ice mapping in polar regions.

According to the suggestion, we conducted a comparison between the S1-derived Arctic sea ice cover data and the pixel-level visual inspection results. We manually labeled 96 cases of objectively selected S1 data as ground truth, and the validation based on those 96 cases shows an overall accuracy of 96.10%. A section "4.1 Comparison with Visual Interpretation Results" has been added to the revised manuscript and please refer to this section for more details.

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

Response 2:

A more detailed illustration of the parameter determination has been added in the revised manuscript.

For the "fixed thresholds of [2dB, 7dB] and [-2dB, 3.5dB]":

"These thresholds ensure approximately 95% of the HH/HV and HH-HV values falling into the range, which were determined according to the statistics of more than 200 S1 EW images acquired under different scenarios of sea ice and open water." (page 9, line 202-204)

For the "batch size",

"In our case, the performance of the model improves with the increasing of the batch size. However, due to the limitation of the memory capacity (16G), the batch size of 8 is the maximum value the device can handle." (page 11, line 254-256)

For the "patch size of 256 × 256 pixels":

"The patch size of 256 × 256 pixels is an empirical value weighed between manageable model size and sufficient information one patch has." (page 10, line 224-225)

For other unexplained hyperparameters of the U-Net model, we followed the original settings by the authors of U-Net (Ronneberger et al., 2015). There is no automatic hyperparameter optimization.

Reference:

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical image computing and computer-assisted intervention, 234-241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Comments 3: A minor concern is the use of a small training dataset (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?

Response 3:

The size of a S1 image is about 1,000 × 1,000 pixels after down-sampling (from original size of about 10,000 by 10,000). But the inputs of U-Net are image patches with dimensions of 256 × 256 pixels. This means numbers of small patches can be extracted within one S1 image. After patches extracting and training sample augmentation, the total number of training and evaluation dataset reaches more than 8000, which is a reasonable number for the U-Net model. To make it more clear, we have further illustrated the number of datasets in our revised manuscript. Please refer to subsection

3.4 (page 10, line 229-230).

Moreover, the selection of the training samples considered both the data acquisition location and the data acquisition season. We believe they covered most of the complex ice and water conditions. Considering that pixel-level manual labeling is quite time-consuming, not cost-effective to add more training samples for the moment.