Dear TC editor and authors of the revised manuscript egusphere-2023-295,

The manuscript topic is interesting and ice floe information is useful for multiple purposes. The other two reviewers have already provided good comments to improve the manuscript. I hope my comments will complement the comments of the other reviewers.

We thank for reviewer for careful review and helpful feedback on our manuscript. Please find our responses to your comments below.

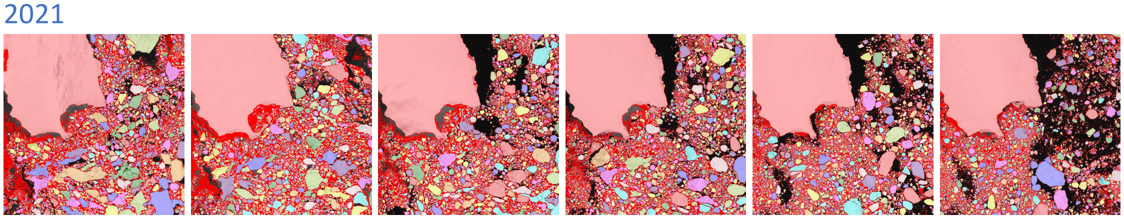
As suggested by another reviewer the abstract needs to be updated to include more on the applied method and results provided by the method instead of general level information. The abstract also begins with FSD and in the manuscript FSD is in the section named as "Case study:FSD" and FSD does not appear in the manuscript title. I suggest at least to remove "Case study:" from the section 6 title, also consider including FSD in the manuscript title. FSD is also significantly present in the introduction section.

Thanks for the comments. We will revise the abstract, reorganize manuscript structure, and include ''floe size distribution'' in manuscript title.

The amount of data is very restricted, Why only four Sentinel-2 images have been used? There exist a lot of Sentinel-2 data. It should be emphasized that with such limited data sets this is a case study and the results can possibly not be generalized.

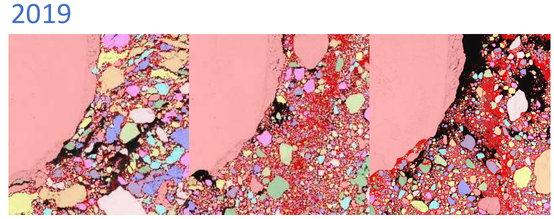
Our approach has been applied to extract the ice floes spatially and temporally from Sentinel-2 images, e.g., below show the floe segmentation results for S2-1 region from April to June in 2021 and other years.

2021



2020

2019



It is not practical to present so many images in the paper, and instead it is common to use some typical images to demonstrate segmentation performance of the methods. Therefore, in addition to four airborne MIZ images, we also used these four typical Sentinel-2 floe images with ice concentration from low to high to show the floe segmentation results.

Please note that the DL models were trained only on local-scale airborne MIZ images, and they did not encounter global-scale satellite data during the training. The Sentinel-2 images were extra data used to investigate the generalization abilities of the trained DL models from local-scale MIZ images to global-scale images (please see our response to reviewer 1's comment ''Line 98'').

Please note also that, how many individual floes (especially those that tightly connected with many other floes) are successfully identified from an image without being over- or under-segmented is also an important measure of ice floe segmentation method.

A proper cloud mask is required to be able to automatically segment ice floes. As "manual-free" is mentioned in the manuscript title I think cloud masking should be discussed in the manuscript. Does there exist automated methods for reliable cloud masking or at least excluding images with clouds? Give references of possible cloud masking approaches or suggestions for improved automated cloud masking. Could "manual-free" in the title be "automated" instead?

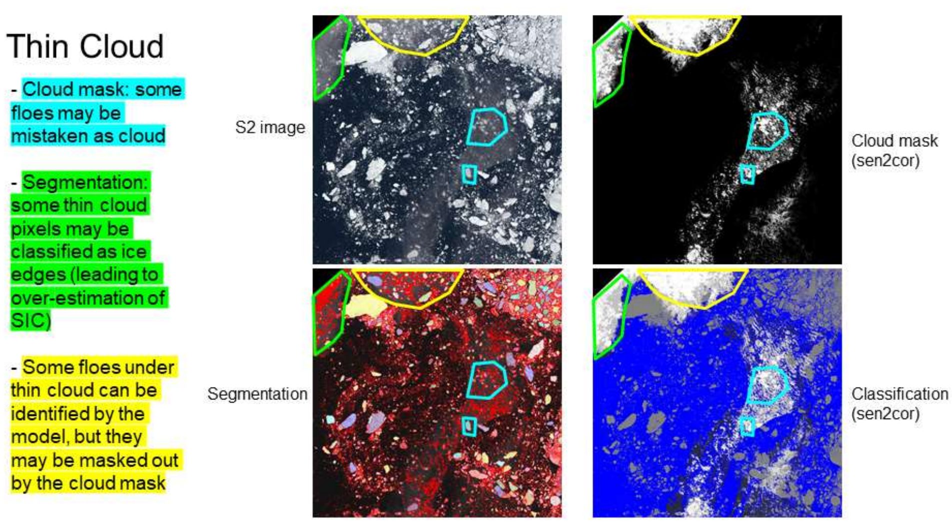
There is an option to choose cloud coverage when downloading Sentinel-2 data from the Copernicus Open Access Hub. So it can exclude images with clouds.

In addition, we have applied and compared the existing methods, Sen2Cor (via ESA Snap software or running on linux terminal) and Fmask, to mask clouds in Sentinel-2 images by means of cloud masking and classification. For us, Fmask works better than Sen2Cor. Please see below for reference.

Thanks for the comment, we will add a brief discussion on cloud masking in the revised manuscript, and also revise manuscript title.

=====

Cloud masking for Sentinel-2 floe image segmentation

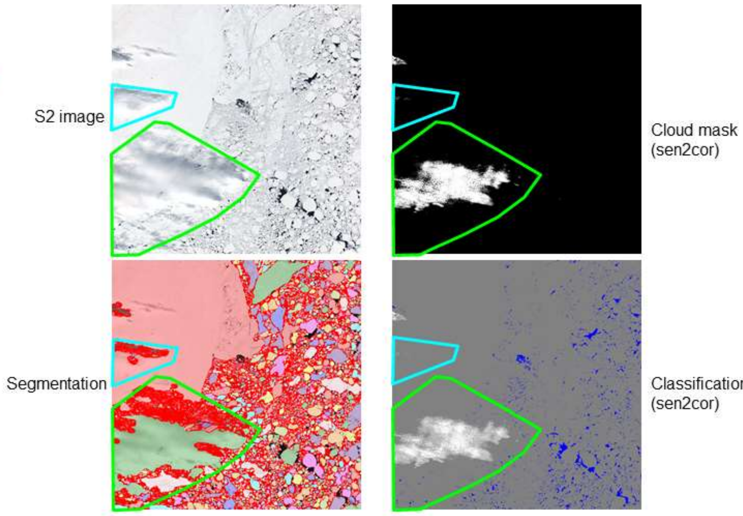


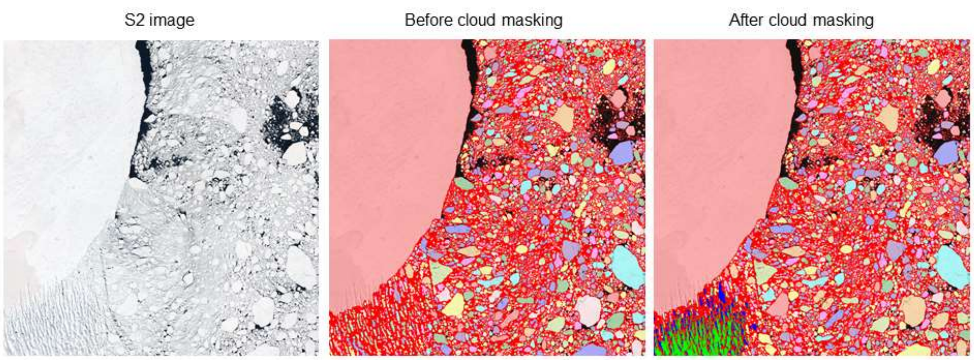
Thick Cloud

- Cloud mask: some cloud regions may not be detected

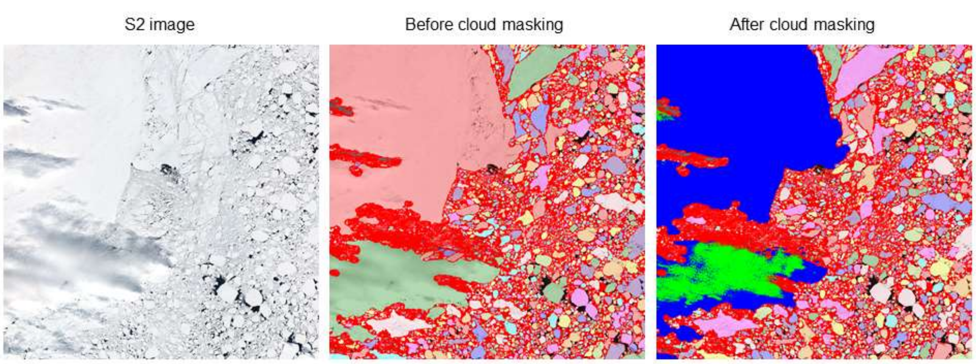
- Segmentation: it is hard to find floes and floe boundaries in and around the regions covered by thick cloud

Suggestion:
The detected floes adjacent to cloud regions will not be considered when determining FSD, since they are likely to be incorrectly segmented





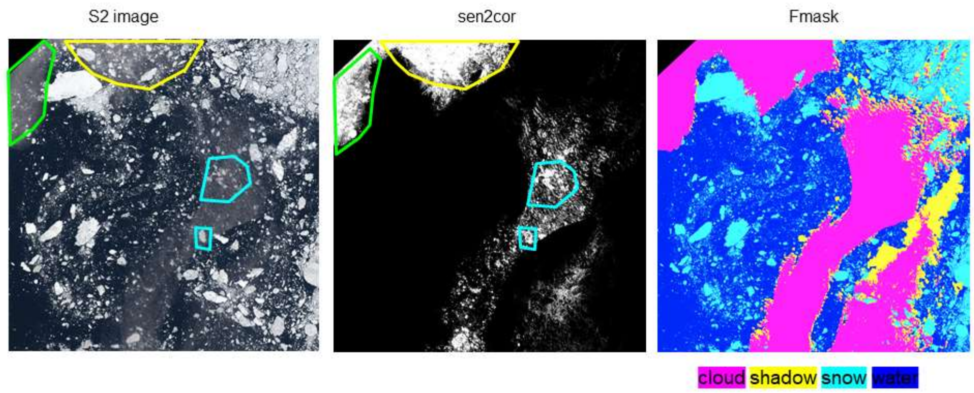
Green: clouds, detected by sen2cor
Blue: floes adjacent to the detected clouds



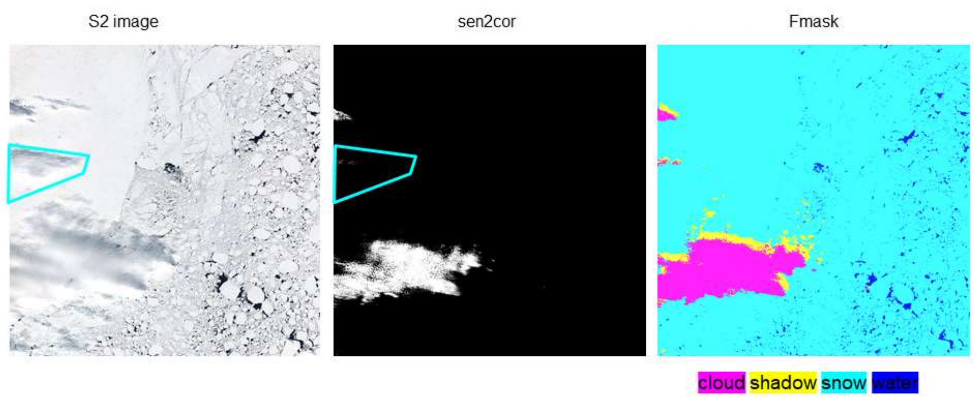
Green: clouds, detected by sen2cor
Blue: floes adjacent to the detected clouds

The current model doesn't consider the category of cloud/cloud shade, some cloud/cloud shade pixels are mistaken for ice edges.
Also due to the insufficient detection of cloud by the sen2cor module, it's hard to correct the this type of error.

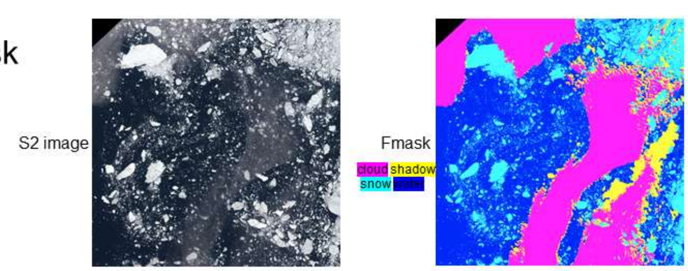
Sen2Cor vs Fmask
Via cloud mask - thin cloud

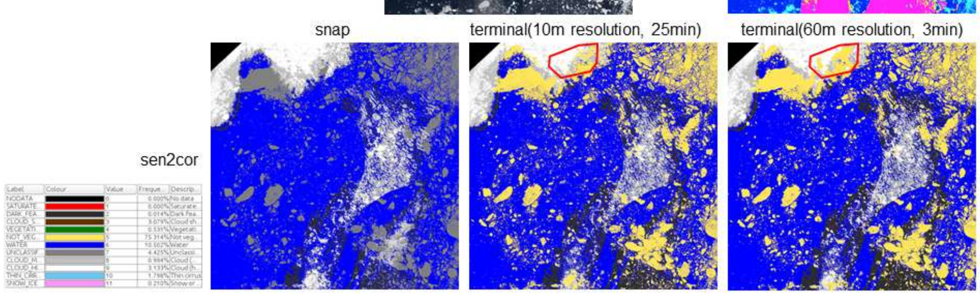


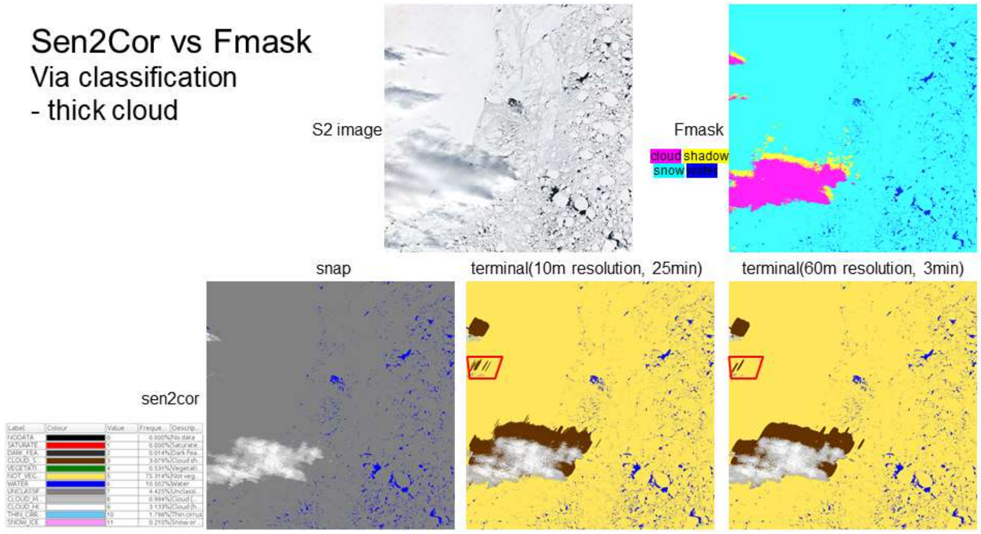
Sen2Cor vs Fmask
Via cloud mask - thick cloud



Sen2Cor vs Fmask
Via classification
- thin cloud







The results and discussion have now been presented in the same section. I suggest to make a separate “Discussion” section or rather combine “Discussion” with the “Conclusions” section which is very short now.

Thanks for the suggestion, we will adjust manuscript structure.

P2 L36 “Copernicus”: give a reference

The link: https://scihub.copernicus.eu will be added there.

P3 Dataset → Airborne data: What is the number of airborne images used in this study? What is “a large amount”? Rather give numbers. Were the images used as long strips, mosaics, or single shots?

The total number of airborne images we got was 254, of which 52 were selected to be labelled for this study.

The airborne image data were used as single shots.

P3 Table 1: Give flight altitude(s) and surface area covered, or their ranges, for the images used. These could possibly be included in the table.

We apologize that we cannot provide the information you suggested.

Neither of the authors of this manuscript was on board the expedition 8 years ago. The airborne images were kindly provided from other research group at a university. Table 1 and the average resolution for a pixel were the only information we got about the airborne images.

P3 L83 “Cophub”: Give reference (URL).

The URL was given in the 1st ref. It will be moved to the end of ‘‘Copernicus Open Access Hub’’.

P4 Table 2: Include location information and covered area in the table, e.g. by given center latitude and longitude (and covered area e.g. in km2).

We will add these information to the table.

P6 L110-112: Hypothesis of improvement by widening the boundaries would require some evidence. Would it be possible to show test results with a small set of imagery and some numeric evidence based on these tests?

Yes, we can do that when submitting the revised manuscript.

Please note, the widened boundaries (2-pixel wide) are inner boundary (1-pixel wide) and outer boundary (1-pixel wide), which are the two types of boundaries for an object/floe. Please also see our response to reviewer 1’s comment ‘‘Line 111-112’’.

P8 S3.2. Deep learning model: Give at least a short description of U-Net++ giving best results or a diagram of the network. Now this subsection is very short and it is essential for the study.

We apologize for the lack of descriptions about the models, and we will add the descriptions and diagram of the models in the revised manuscript.

P8 Post-processing: Applying morphological opening and closing seems a bit heuristic to me. Are there any references or if not would it be possible to demonstrate the benefits of the morphological processing? What is the shape and size of the morphological operator (often a disk is used)? Could this step be included in the ML algorithm somehow, i.e. could the NN learn the post-processing?

Please see (Banfield, 1991; Banfield 30 and Raftery, 1992; Soh et al., 1998; Steer et al., 2008; Wang et al., 2016) for reference. We also mentioned their work in the Introduction section:

‘‘Morphological operations can be used with different improvements to determine individual ice floes, but the methods operate directly on binarized floe images and thus cannot separate out the floes that had no or few gaps with any surrounding floes after binarization (Banfield, 1991; Banfield 30 and Raftery, 1992; Soh et al., 1998; Steer et al., 2008; Wang et al., 2016)’’

Due to the problems with the morphological operations described above, we used the morphological operations in a post-processing step to refine the floe segmentation. Fig. 5 and Fig. 6 in the manuscript can demonstrate the benefit of the processing. A disk-shaped structuring element with a radius of 4 pixels was used in the morphological operations. The disk shape was chosen because it is non-directional and can handle ice floes more uniformly than other shapes without being aware of floe's irregular shape and orientation.

The next task after the work presented in the manuscript is to use the trained DL model together with the post-processing to label more ice floe images and create more dataset, and then train more robust DL models for ice floe segmentation, as we mentioned in the Conclusion section: ''it can also be utilised as a "higher version" of "annotation tool" and produce more "ground truth" from a wide variety of ice image data sources to further train more robust DL models for obtaining more accurate ice parameters from images.''

P9 Training: GPU memory is referred on line 170 and this information is then given later on page 10. The hardware (and software) used should be given before it is referred. The used HW and SW could e.g. be included in the dataset section and changing the title to something like "Datasets and computational resources". Also include the used SW with reference in the same section. Also mention there that all the execution times given later are given for this specific configuration.

Thanks for the suggestion! We will adjust them in the revised manuscript.

P10 L174: Does this distribution of classes correspond to their distribution in general? Then it can be used in training. What happens if the distribution of classes is balanced (33% of samples for each class) for the training? Does balancing degrade the classification?

The distribution of classes depends on the extent of sea ice coverage and the amount of ice floes in the image. The higher the ice concentration in the image, the higher the proportion of ''ice'' class (the lower the proportion of ''water'' class); the more ice floes, the higher the proportion of ''floe boundary'' class. MIZ images generally have higher proportion of ''floe boundary'' class than other ice floe images.

Balanced classes will not degrade the classification. Instead, it makes model easier to train because it helps the model learn the features of each class equally.

In ice floe segmentation, ''floe boundary'' is a hard-to-train class because: 1) the pixel intensity of floe boundaries, especially the boundaries between connected floes, is usually similar to that of ice; 2) the proportion of floe boundaries in ice floe images is usually much smaller than other two classes of "ice" and ''water'' (MIZ images are less affected by this issue). Therefore, it is necessary to increase the proportion of ice floe boundaries in the training data set, and using MIZ images as training images and widening floe boundaries can help with this.

Although the DL model was trained on MIZ images with a relatively high proportion of ice floe boundaries, it also has a good generalization ability to other ice floe images with low proportion of floe boundaries (e.g., the largest ice and water regions in Sentinel-2 images). It demonstrates the model trained on restricted datasets that can also generalize to wider datasets.

P10 L175: The number of test samples is not very large. What is the effect of reducing samples in training and validation data sets and increasing of the test data set? Are these numbers of samples selected based on some kind of performed tests?

Reducing the training and validation sets to increase the test set often leads to a decrease in the performance of DL models.

A common ratio between train, validation, and test is 80:10:10, and 70:15:15, 60:20:20, etc. are also practical ratios for splitting datasets.

Regarding the number of test samples, please see our response to your previous general comment ‘‘The amount of data is very restricted, …’’.

P11 Section 5: Would be good to have some kind of related introductory text under “5 Experimental results and discussions” and “5.1. DL model evaluation”, now they are empty.

We apologize for the lack of descriptions, and we will put some descriptions in the revised manuscript.

P13 Section 5.1.3. Inference time: Could this be “Segmentation time” instead, it would be more informative. The HW (and SW) used for segmentation could be given already in an earlier section, e.g. jointly with the introduction of data sets.

Thanks for the suggestion and we will change them.

P15-15 Figure 10: Fig. 10 is nor in two parts and in two pages. Would it be possible to compress a little and make it to fot on one page?

We will put the figures on one page.

P19 “6 Case study: floe size distribution”: I recommend to drop “Case study:” because FSD is the essential parameter to be estimated by the method and it is also in essential role in the abstract and introduction and the whole manuscript is actually a case study because the datasets are quite limited.

Floe size distribution: Now FSD is estimated in two different resolutions (airborne and satellite data). It would be interesting to see how well FSD can be extrapolated from resolution to another (both from larger to smaller and smaller to larger) based on a fit distribution model.

This would be very valuable information and this theme could be included in the discussion section.

Thanks for the suggestion! Multi-scale FSD is actually part of our ongoing project. We will try to give a brief explanation on multi-scale FSD in the revised manuscript.

P24 “Conclusions”: This section is very short. Possibly it could be combined with a discussion section. Here could also be some conclusions on how close to automated FSD estimation the proposed method is? Could it be used for operational monitoring and what will still be required before possible. At least cloud masking should be discussed and also the annual period of possible operation (lighting conditions, what is the fraction of cloudless time in suitable lighting conditions in different sea ice covered areas). What are the ways forward in automated ice floe analysis?

We apologize for the short conclusion and thanks for the valuable suggestion!

In the revised manuscript we will 1) discuss the cloud masking 2) discuss the limitations of the method 3) give the application prospects of the method 4) give suggestions on the way forward in automated floe analysis.