*In the text below, **reviewer comments**, author comments, original manuscript text and updated manuscript text are color-coded as shown here.*

Review of ” Pan-Arctic Sea Ice Concentration from SAR and Passive Microwave” by Wulf et al.

Robust all-weather multi-sensor SIC estimates for the Arctic are an important topic in cryospheric Earth Observation. Here, the authors present a deep learning-based retrieval framework for SIC which combines SAR and PMW observations, trained with regional ice charts from Danish and Canadian Ice Services.

The authors mainly do a good job of explaining the technically complex training and retrieval. The idea of applying the full classification probability vector’s information to estimate SIC is clever and commendable. Still, some gaps in presentation remain. One result figure was missing from the pdf, and some aspects of the retrieval and training could be better justified.

Reviewer comment: Also, the pan-Arctic applicability discussion, while logical and relevant, seemed to focus on resolution aspects and operator-dependent biases. I was missing some critical thought on the broader aspects of deriving SIC across the full width of the Arctic Ocean. Once these issues are remedied, however, I see this paper as a very worthwhile addition to the body of SIC retrieval literature, with clear advances in several aspects and a clean delivery in written form.

Author comment: Thanks for the comment. We added a couple of paragraphs in the Discussions section (Section 5) addressing broader aspects of deriving SIC from SAR across the full width of the Arctic Ocean.

We included a small section on currently poor Arctic coverage of Sentinel-1 imagery and the prospect of future sensor-agnostic SAR-based sea ice retrieval algorithms with greater Arctic coverage, e.g. by including SAR imagery from the RADARSAT Constellation Mission in addition to Sentinel-1.

Added to the “Future Work” section of the manuscript:

Improving the Arctic coverage of SAR-based sea ice retrievals

Since the loss of Sentinel-1B in December, 2021, the acquisition of Sentinel-1 imagery in the Arctic has been severely reduced, with no images being acquired in the central Arctic at all. The reduced coverage of Sentinel-1 imagery in regions of high maritime traffic has impacted the national ice services that now are more reliant on other SAR missions to meet user demands on the update frequency of their sea ice products. For example, the Radarsat Constellation Mission from the Canadian Space Agency provides C-band SAR imagery in the Arctic. A natural next step in the development of SAR-based sea ice retrieval algorithms, such as ASIP, is the adaptation to SAR imagery from multiple sensors. The development of sensor-agnostic SAR-based sea ice retrieval algorithms would greatly improve the coverage of the derived sea ice products in the Arctic, which would benefit not only the national ice services, but the sea ice modeling community as well.

We added a few sentences about the representation of Arctic sea ice conditions in the ASIDv2+ training dataset, which only covers a part of the Arctic. This also relates to the reviewer's very last comment about "the applicability of the near-coastal training data to the full range of ice behaviour across the broad swath of the Arctic Ocean".

Original text: ….The ice charts are often produced on the basis of SAR images, enabling the compilation of training datasets consisting of very timely - if not exact - match-ups between ice chart and SAR image, which is important due to the high spatial resolution of the SAR sensor and the continuous movement of drifting sea ice. However, as ice charts are drawn by ice analysts by manual interpretation of satellite observations, there are bound to be inherent uncertainties in the ice charts, such as analyst subjectivity, and inter - and intra-analyst variability…..

Updated text: ….The ice charts are often produced on the basis of SAR images, enabling the compilation of training datasets consisting of very timely - if not exact - match-ups between ice chart and SAR image, which is important due to the high spatial resolution of the SAR sensor and the continuous movement of drifting sea ice. However, while the ice charts in the ASIDv2+ dataset cover diverse sea ice conditions in the Greenland waters as well as the Canadian Arctic, the dataset might not be representative of all possible sea ice conditions across the full width of the Arctic Ocean. As such, there might be sea ice conditions that are not represented in the ASIDv2+ dataset and thus not available for ASIP to learn. Furthermore, as ice charts are drawn by ice analysts by manual interpretation of satellite observations, there are bound to be inherent uncertainties in the ice charts, such as analyst subjectivity, and inter - and intra-analyst variability…..

Reviewer comment: The regional DMI and CIS coverages appear to have overlap near west Greenland. Did the authors assess the similarity of the ice charts as a measure for subjective analyst's classification uncertainty? Section 4.2 suggests the results do reflect some operator dependence, but did you try to quantify it?

Author comment: Thanks for the comment – we did not quantify the ice analyst uncertainty in this study. When we compiled the ASIDv2+ training dataset, we did find a smaller number of Sentinel-1 scenes (<100) in the Baffin Bay region for which we were able to find both DMI and CIS regional ice charts that met our match-up criteria. We agree that it would be beneficial to measure analyst uncertainty by assessing the similarity between ice charts drawn by different analysts on the basis of the same Sentinel-1 acquisition, but we did not include it in the (already wide) scope of the present paper. Such studies on the assessment of inter-analyst variation exist, and we refer to these studies in section 4.2.

Reviewer comment: Why were the validation and test datasets very, very small compared to the training dataset? Often at least 10% of all data are assigned to validation and test groups, here it's ~1%. How much does this influence the results?

Author comment: Our focus in this study is the pan-Arctic application of a deep learning model trained on a regional dataset covering only the Greenland waters and parts of the Canadian Arctic. As we cannot assess the quality of ASIP in the pan-Arctic region using a regional dataset of ice charts, we consider the Arctic-wide comparison of the ASIP SIC to the OSI SAF SIC product (OSI-408) as the main result in this study and the main evaluation of the presented methodology. The initial quantitative evaluation of the ASIP SIC against ice charts (section 4.2 and Fig. 5) is included in the manuscript to show the result of the training of the ConvNets, as well as the improvement in SIC accuracy (measured against ice charts) over previous studies when using our proposed SIC retrieval from the calibrated confidence vectors. As ASIP is trained with ice charts as labels, the qualitative evaluation in section 4.2 (Fig. 6) is included to show how well (if at all) the model imitates ice charts – both in terms of spatial resolution and quality/accuracy. For these purposes, we would argue that a test set of 50 samples is sufficient, especially given that the test samples have been carefully selected to be representative of the Greenland and Canadian sea ice regimes.

That being said, an 80%/10%/10% split is a good rule of thumb in many cases. However, when the dataset becomes increasingly large, the ratio of test data to training data often diminishes. The ASIDv2+ dataset (5382 samples) is many times larger than the similar datasets ASIDv2 (461 samples) and AI4Arctic Sea Ice Challenge Dataset (533 samples). Other studies using these previous, smaller versions of the ASIDv2+ dataset, e.g. Stokholm et al. (2022) and Chen et al. (2023) use test sets of 23 and 20 scenes, respectively.

Reviewer comment: Section 3.2. answers well the question of “what”, but offers little for the question “why”. Did you test alternative deep learning methods than the ConvNet chosen? What was the key reason to choose it over other alternatives?

Author comment: Thanks for the comment. We did not test other deep learning methods. Our focus in this study is the generalization of a regionally trained model to the pan-Arctic region as well as the introduction of a new method for quantifying the uncertainty of the sea ice products inferred by the trained model - two subjects missing in the current corpus of literature on ML-based sea ice retrievals from SAR. For this reason, we chose a fairly simple UNet model, which is widely used in Earth Observation. We “modernized” the original UNet (which is an “old” architecture now, from 2015) with findings from two widely recognized papers from the field of computer vision, namely Sandler et al. (2018) and Liu et al. (2022). We added a sentence in the beginning of Section 3.2 to provide the reader with the reasoning for choosing the chosen ConvNet architecture.

Original text: The ConvNet we employ in this study follows a modified U-Net (Ronneberger et al., 2015) structure…

Updated text: As the focus of this work is on the generalization of deep learning-based sea ice retrieval algorithms and the uncertainty quantification of their outputs, rather than on the architectural optimization of the predictive performance of the algorithm, we carry out all experiments in the subsequent sections using a fairly simple ConvNet architecture. The architecture follows a modified U-Net (Ronneberger et al., 2015) structure…

Reviewer comment: Figure 9 seemed to be missing entirely from the pdf?

Author comment: We are sorry to hear about this inconvenience. However, when we download the .pdf from the Cryosphere MS records, Fig. 9 is there on page 23, as well as in the following link from the discussion page: https://egusphere.copernicus.org/preprints/2024/egusphere-2024-178/egusphere-2024-178.pdf.

Reviewer comment: Fig 10 and similar – the difference plot color range would be better constrained to typical observed ranges rather than the physical maxima of +/- 100%.

Author comment: We agree that the dynamic range should be lower to highlight the spatial variation in the difference plots. While the majority of the values in the difference plots (ASIP SIC – OSI SAF SIC) in Figs. 9 and 10 are positive and between 0% and ~60%, there are both positive and negative extremes in all plots. We believe it's important to include both negative and positive values in the differences plots for the reader to study. We changed the dynamic from the physical maxima of +/- 100% to +/- 50%.

Minimum and maximum values for all difference plots are listed below.

Figure – date: minimum, maximum

Fig. 9 – Jan 24th: -90, 100

Fig. 9 - May 28th: -97, 100

Fig. 9 – Sep 24th: -100, 100

Fig. 10 - Jan 13th: -89, 99

Fig. 10 - April 21st: -96, 100

Fig. 10 – Aug 3rd: -87, 100

Reviewer comment: While I agree that the pan-Arctic SIC estimates here appear reasonable, I hold some reservations about the applicability of the near-coastal training data to the full range of ice behaviour across the broad swath of the Arctic Ocean. For example, were there sufficient leads and melt ponds in summer in the training data w.r.t. the innermost AO? Were ridges present in the full height range encountered?

Author comment: Thanks for the comment. The applicability of a regionally trained model to the pan-Arctic region is indeed one of the main research questions in this work. The ASIDv2+ dataset, which is used for the training of ASIP, covers sea ice conditions in the Greenland waters as well as the Canadian Arctic. This geographic region is diverse in terms of sea ice conditions, covering the full seasonal cycle of sea ice freeze-up and melt, and multi-year ice represented in parts of the Canadian Archipelago, north of Greenland and along the eastern coast of Greenland. Still, we cannot be sure that all sea ice behaviors are represented in the training dataset and thus available for the model to learn. In the comparison against OSI-408-a (Figs. 9, 10 and lines 426-447 in the main text), our objective was to utilize OSI SAF SIC as a benchmark for our ASIP SIC product at a pan-Arctic scale. This comparison aimed to demonstrate that our SAR-based sea ice concentration product aligns well with an established and trusted standard. The OSI SAF Sea Ice Concentration (SIC) algorithm is a highly reliable and time-tested tool that has been in operational use for decades. OSI SAF SIC products are integral to renowned operational and climate models.

Regarding the reviewer's point on the representation of leads and melt ponds during summer in the training dataset, we attempt to address some of these issues in the discussion on the limitations of ice charts as label data for training these algorithms (Sec. 5, lines 519-520). While leads and melt ponds are definitely represented in the Sentinel-1 imagery contained in the ASIDv2+, these ice features are not delineated in the corresponding ice charts, and therefore, these features are not explicitly learned by the model. In the future, ice features, such as leads, could be manually added to the ice charts in ASIDv2+ to allow the model to map these features as well.

References:

A. Stokholm, T. Wulf, A. Kucik, R. Saldo, J. Buus-Hinkler and S. M. Hvidegaard, "AI4SeaIce: Toward Solving Ambiguous SAR Textures in Convolutional Neural Networks for Automatic Sea Ice Concentration Charting," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-13, 2022, Art no. 4304013, doi: 10.1109/TGRS.2022.3149323.

Chen, X., Patel, M., Pena Cantu, F., Park, J., Noa Turnes, J., Xu, L., Scott, K. A., and Clausi, D. A.: MMSeaIce: Multi-task Mapping of Sea Ice Parameters from AI4Arctic Sea Ice Challenge Dataset, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-1297, 2023.

M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. -C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.

Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., and Xie, S.: A ConvNet for the 2020s, CoRR, abs/2201.03545, https://arxiv.org/abs/2201.03545, 2022