https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

MMSeaIce: Multi-task Mapping of Sea Ice Parameters from AI4Arctic Sea Ice Challenge Dataset

Xinwei Chen[1], Muhammed Patel[1], Fernando J. Pena Cantu[1], Jinman Park[1], Javier Noa Turnes[1], Linlin Xu[1], K. Andrea Scott[1], and David A. Clausi[1]

[1]Vision and Image Processing Lab, Department of System Design Engineering, University of Waterloo, ON, Canada

Correspondence: Xinwei Chen (xinwei.chen@uwaterloo.ca)

Abstract. The AutoIce challenge, organized by multiple national and international agencies, seeks to advance the development of near-real-time sea ice products with improved spatial resolution, broader spatial and temporal coverage, and enhanced consistency. In this paper, we present a detailed description of our solutions and experimental results for the challenge. We have implemented an automated sea ice mapping pipeline based on a multi-task U-Net architecture, capable of predicting sea ice concentration (SIC), stage of development (SOD), and floe size (FLOE) using Sentinel-1 SAR data. For model training and evaluation, we utilize the AI4Arctic dataset, which includes SAR imagery, corresponding passive microwave and auxiliary data, and ice chart-derived label maps. Among the submissions from over 30 teams worldwide, our team achieved the highest combined score of 86.3%, as well as the highest scores on SIC (92.0%) and SOD (88.6%). Additionally, our result analysis showcases the effectiveness of various techniques, such as input SAR variable downscaling, spatial-temporal encoding, input feature selection, and loss function selection, in significantly improving the accuracy, efficiency, and robustness of deep learning-based sea ice mapping.

1 Introduction

Automated sea ice mapping using satellite data plays a vital role in understanding and monitoring the Earth's polar regions. Sea ice, a critical component of the cryosphere, undergoes significant spatial and temporal variations, impacting climate, ecosystems, and human activities. Satellite-based automated mapping techniques offer a unique advantage in providing comprehensive and frequent coverage over vast and remote areas. By employing advanced algorithms and machine learning (ML) approaches, these methods enable the efficient detection and characterization of different sea ice parameters (Lyu et al. 2022b). Accurate and timely sea ice mapping aids in climate modeling, facilitating climate change assessments, supporting operational activities such as navigation and resource management (Li et al. 2022), and enhancing our understanding of the intricate dynamics between the atmosphere, ocean, and ice-covered regions (Mahmud et al. 2022). The continuous advancements in automated sea ice mapping techniques using satellite data offer valuable insights into this fragile environment and aid in making informed decisions for sustainable development and environmental stewardship.

Deep learning (DL) has emerged as a powerful tool for sea ice parameter estimation from satellite data especially dual-polarized SAR imagery, revolutionizing the field with its wide-ranging applications and improved performance compared to

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

traditional algorithms or conventional ML methods. DL-based models have demonstrated exceptional capabilities in accurately estimating crucial sea ice parameters such as sea ice concentration (SIC) (Wang et al. 2016, 2017; Cooke and Scott 2019; Radhakrishnan et al. 2021; De Gelis et al. 2021; Stokholm et al. 2022; Malmgren-Hansen et al. 2020), stage of development (SOD) (Jiang et al. 2022; Lyu et al. 2022a; Chen et al. 2023a; Song et al. 2021; Khaleghian et al. 2021a; Liu et al. 2021a; Khaleghian et al. 2021b; Kruk et al. 2020; Boulze et al. 2020; Guo et al. 2023; Zhang et al. 2021b, a), and floe size (Chen et al. 2020; Nagi et al. 2021). These models leverage the ability of deep neural networks to automatically learn complex features and patterns from large volumes of data, enabling more robust and precise parameter estimation.

However, it is important to acknowledge the limitations of previously proposed DL-based methods for sea ice parameter estimation. First, many existing models focus on estimating a specific parameter, which do not address the comprehensive characterization of sea ice in operational use. Second, a significant number of studies rely on data from a single sensor, leading to potential ambiguities and limited information integration. For example, although SAR images are capable of showing the spatial patterns formed by sea ice in high resolution, its backscatter intensities do not always distinguish between open sea in windy conditions and various ice surfaces (Malmgren-Hansen et al. 2020). In contrast, brightness temperature maps collected by radiometers such as the Advanced Microwave Scanning Radiometer 2 (AMSR2) satellite sensor can distinguish well between ice and open water but with coarse spatial resolution. Recent studies have implemented ML and DL-based methods for retrieving SIC from brightness temperature data and achieved promising results (Chi et al. 2019; Soleymani and Scott 2021; Chen et al. 2023b). Third, due to the challenges in obtaining labeled samples for training, DL-based models for sea ice often suffer from limited volume of datasets, which can impact their generalization capabilities. Addressing these limitations is crucial to further enhance the effectiveness and applicability of DL-based sea ice parameter estimation methods.

Therefore, to address these challenges in automated sea ice mapping, the ESA (European Space Agency), DMI (Danish Meteorological Institute), the Technical University of Denmark (DTU), and NERSC (the Nansen Environmental and Remote Sensing Center) collaborated to create a sea ice challenge called AutoIce (Stokholm et al. 2023a). The goal of the challenge is to invite participants worldwide to derive more accurate and robust AI-based solutions of automated retrieval of multiple sea ice parameters, specifically, sea ice concentration (the percentage ratio of sea ice to open water), stage of development (the type of sea ice and its thickness), and floe size (the size and continuity of sea ice pieces). A large volume of multi-source satellite and auxiliary data named as AI4Arctic Sea Ice Challenge Dataset (Buus-Hinkler et al. 2022b) are provided for the training and evaluation of the derived models.

In this paper, we present our methodology and corresponding outcomes that resulted in achieving 1^{st} place in the challenge. Following the Introduction, Section 2 provides an overview of the AI4Arctic dataset used in this work. The methodology for the retrieval of sea ice parameters based on a multi-task U-Net, along with a bag of tricks employed for model performance improvement (e.g., SAR scene downscaling, input variable selection, spatial-temporal encoding, loss function selection) are illustrated extensively in Section 3. Experimental results with ablation studies are analyzed and discussed in Section 4. Finally, conclusions along with future research are summarized in Section 5.

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

2 Data Overview

The AI4Arctic dataset consists of 533 netCDF files, including 513 training files and 20 test files. Each training file contains dual-polarized Sentinel-1 Extra Wide Swath (EW) images, AMSR2 passive microwave radiometer measurements, numerical weather prediction (NWP) parameters from ERA5 reanalysis dataset, and ice charts that follows the World Meteorological Organization (WMO) code for sea ice classes provided by either the Greenland Ice Service or the Canadian Ice Service. The 20 test files have the same parameters as the training files, except for the sea ice chart (label) data. There are two versions of the dataset available: a raw version and a ready-to-train version. The ready-to-train version undergoes additional processing steps to prepare it for deep learning algorithms. To focus on model development and skip the initial preparation steps, we adopt the ready-to-train version to train our models. This version converts the original ice chart shapefile format into the netCDF format.

Each polygon in the ice chart is represented by an ID number and a table containing the ice chart variables for the polygon in the associated netCDF file. The sea ice concentration (SIC) parameter in each polygon represents the ratio of sea ice to open water in a given area, divided into 11 classes with 10% increments, ranging from 0% (open water) to 100% (fully-covered sea ice). In addition to total SIC, each polygon contains partial sea ice concentrations associated with stage of development (SOD) and floe size (FLOE), which sum up to the total SIC. The partial concentrations are normalized by the total concentration to determine if a partial concentration is dominant in each polygon. Dominant parameters are identified based on a threshold of 65%. Therefore, a large portion of polygons do not have a dominant SOD or FLOE and are masked out from the labeling of SOD and FLOE. The SOD serves as an indicator of the sea ice type, which can be interpreted as a proxy for its thickness and ease of traversal. It consists of 5 classes: 0 represents open water, 1 is for new ice, 2 for young ice, 3 for thin first-year ice, 4 for thick first-year ice, and 5 for old ice (older than 1 year). The FLOE characterizes the size and continuity of sea ice floes, and it is defined by 6 classes: 0 for open water, 1 for cake ice, 2 for small floe, 3 for medium floe, 4 for big floe, 5 for vast floe, and 6 for bergs, which include various forms of icebergs and glacier ice. In addition, SAR scenes are downsampled to 80 m pixel spacing (around 5000×5000 pixels) for ease of use and to help reduce barriers to entry. The pixel values in the scenes are normalized within the $[-1,1]$ range, and statistical information and class bins are provided. NaN values in SAR images are replaced with 2, and polygon ice charts are assigned a value of 255 to represent non-data or masked pixels. A detailed description of the dataset can be found in the manual provided by (Buus-Hinkler et al. 2022b). To evaluate the model performance numerically, SIC results are evaluated by calculating the R2 coefficient, while SOD and FLOE maps are both evaluated using the F1 score. The three sea ice parameter scores will be combined into one single final score as defined in the weighting scheme shown in Table 1.

Table 1. The metrics for evaluating the three sea ice parameters and their weights in the final score specified by the competition.

Sea ice parameter	Metric (%)	Weight in total score
SIC (sea ice concentration)	R2	2/5
SOD (stage of development)	F1	2/5
FLOE (floe size)	F1	1/5

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

3 Methodology

3.1 Network Design

U-Net has shown success in many recent research concerning sea ice mapping (Radhakrishnan et al. 2021; Stokholm et al. 2022; Kucik and Stokholm 2023; Nagi et al. 2021; Ren et al. 2021; Huang et al. 2021; Stokholm et al. 2023b). For example, in a recent study by Kucik *et al.* (Kucik and Stokholm 2023), a U-Net architecture (Ronneberger et al. 2015) was trained on the AI4Arctic dataset to accurately retrieve sea ice concentration (SIC). Building upon this success, we extend the model to estimate three sea ice parameters concurrently. Our multi-task U-Net consists of four encoder-decoder blocks, with the first two blocks having 32 filters and the remaining blocks having 64 filters (as shown in Figure 1). Alternative configurations, such as adding more blocks or increasing the number of filters, as well as employing state-of-the-art DL-based models for image segmentation such as the Swin transformer (Liu et al. 2021b), were explored. However, none of these approaches surpassed the performance of our current model.

To predict stage of development (SOD) and floe size (FLOE), we utilize the output feature maps from the final decoder and feed them into separate 1×1 convolution layers. Each convolution layer has a number of filters equal to the number of classes, enabling the generation of pixel-based classification results through segmentation. Regarding SIC estimation, as it can be treated as either a classification or a regression problem, we investigate both convolution and regression layers, employing different loss functions to compare their effectiveness.

3.2 Input Variable Processing and Selection

Despite the resolution of the SAR imagery that is well suited for SAR sea ice monitoring, the polygon egg code data is derived from the knowledge of ice analysts who have to generate coarse spatial resolution assessments. Therefore, to generate predictions consistent with the label maps, it is advantageous for input SAR image patches to encompass a large receptive field, which is achieved through the following operations. Initially, the dual-polarized SAR images, distance maps (DMs), and corresponding ice chart-derived label maps are downsampled by a certain ratio (10 in the proposed model). During the training process, patches of size 256×256 are randomly extracted from the downsampled SAR images. As the AMSR2 and ERA5 inputs have been resampled to the Sentinel-1 geometry, their corresponding data points within the geographical areas covered by these patches are also interpolated to the size of 256×256. Data augmentation operations listed in Table 3 are applied to the extracted patches (with a probability of 0.5 for each operation) to enhance the model's generalization ability. During the validation and testing phases, the complete SAR scenes and DMs are downscaled, combined with other upsampled data inputs, and then fed into the trained model. The outputs are subsequently interpolated to match the original size of the SAR data and ice charts for evaluation purposes. To select suitable inputs for training the model, we conduct experiments using various combinations of data inputs. Table 2 presents the combination of data inputs that yield the best performance. For the AMSR2 data, frequencies of 18.7 GHz and 36.5 GHz are chosen due to their higher spatial resolution in comparison to lower frequency channels, as well as their reduced sensitivity to atmospheric water vapor and cloud liquid water when compared to the 89 GHz channels (Minnett et al. 2019; Chen et al. 2023b). All ERA5 inputs in the AI4Arctic dataset are included, except

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

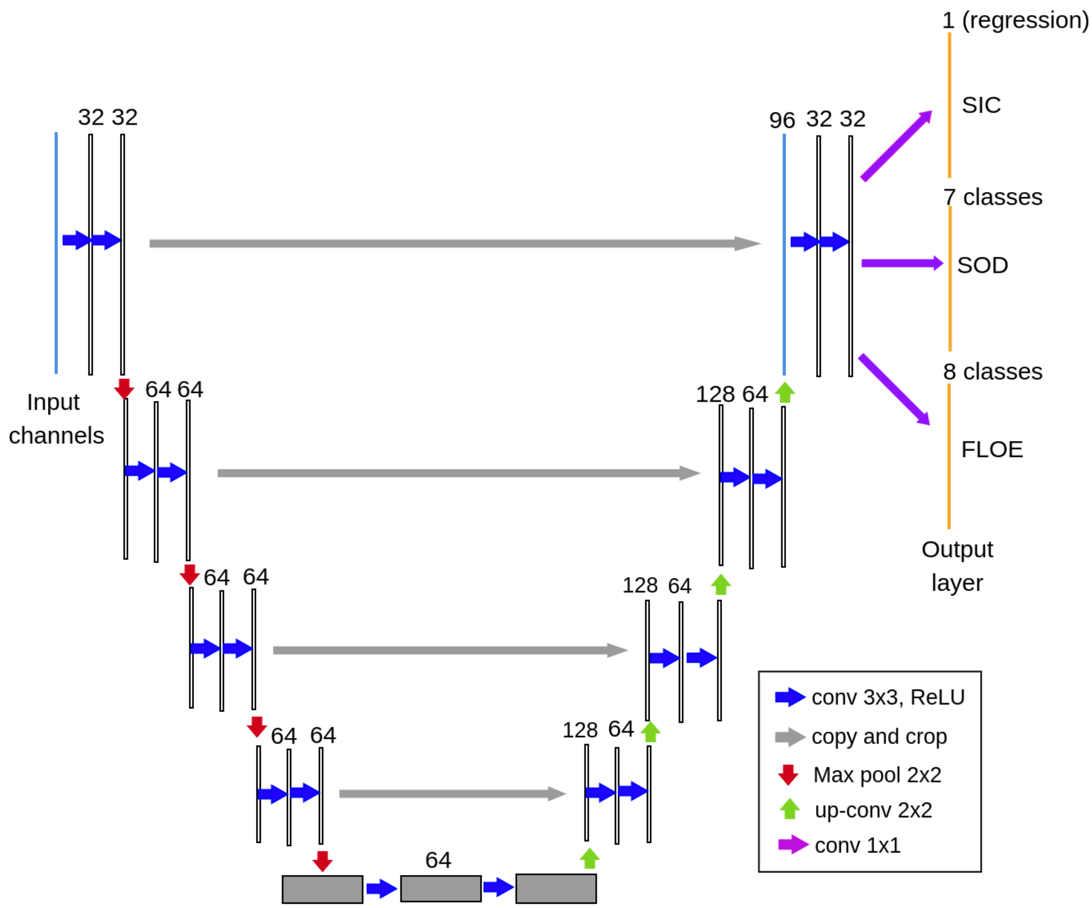

Figure 1. The structure of the proposed multi-task U-Net-based model with output layers in yellow.

for the skin temperature, which exhibits a high correlation with the 2-meter air temperature and does not significantly improve overall accuracy. Detailed results using different combinations of input channels will be demonstrated and discussed in Section 4.

3.3 Spatial-temporal Encoding

In operational sea ice mapping, ice experts not only rely on satellite data analysis but also utilize their domain knowledge, such as understanding typical ice conditions in specific regions during certain months in previous years. Additionally, SAR scenes captured in close proximity and similar time periods tend to exhibit comparable ice conditions. As the DL-based models proposed in this study lack access to such domain knowledge, we incorporate spatial and temporal information of each scene into the input channels, as illustrated in the last row of Table 2. Specifically, the latitude and longitude coordinates of the 21×21

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

Table 2. The combination of data inputs that produces the highest accuracy.

Variable abbreviation	Variable description	Total number of channels
HH, HV	Dual-pol SAR scene	2
AMSR2 subset	Dual-pol AMSR2 brightness temperature data in 18.7 and 36.5 GHz	4
ERA5 subset	10-m wind speed, 2-m air temperature, total column water vapor, total column cloud liquid water	5
Loc, time	Latitude/longitude of each pixel and scene acquisition month	3

Sentinel-1 SAR geographic grid points provided in the dataset are interpolated to match the size of the input SAR image. The time information of each pixel corresponds to the acquisition month (represented by enumeration, e.g., '01' for January) of the respective SAR scene. The effectiveness of spatial-temporal encoding in enhancing accuracy will be demonstrated in the subsequent ablation studies.

3.4 Model Training and Implementation

The specifications of model training are detailed in Table 3, encompassing the combination of hyperparameters that yield the highest validation accuracy. Cosine Annealing (Loshchilov and Hutter 2016) is employed as our learning rate schedule, initially utilizing a large learning rate that gradually decreases following the cosine function to reach a minimum value before rapidly increasing again (every 20 epochs in our model). This approach allows the model to navigate different regions of the loss landscape, potentially avoiding suboptimal local minima and converging to a favorable solution. To ensure sufficient exposure per data sample during training, each epoch comprises 500 iterations, with a batch of patches randomly extracted from training scenes during each iteration. Through exploring various combinations of loss functions, we observe that employing mean square error (MSE) loss for SIC and cross entropy (CE) loss for SOD and FLOE produces the highest testing accuracy. Specifically, the SIC retrieval is treated as a regression task, with a regression layer added before the SIC output in the model. Considering that the magnitude of MSE loss is considerably higher than that of CE loss, we assign a larger weight value (determined empirically) to the CE losses when calculating the total loss. This weight assignment facilitates the convergence of the three scores, as outlined in Table 3.

To validate the generalization capability of the model, for each experiment, 20 SAR scenes from the training data are randomly selected as the validation set. Besides, to prevent the influence of randomness in parameters initialization and training, we train a total of 20 networks for each configuration and obtain the mean and variation of accuracy for more trustworthy performance evaluation. At the conclusion of each epoch, a combined score is calculated from the validation set, utilizing the metrics outlined in Table 1. If the score obtained in the current epoch surpasses all previous epoch scores, the model parameters are updated and saved. The final saved model is subsequently employed to generate predictions for the testing data submissions.

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

Table 3. Specifications of training the proposed model.

Optimizer	Stochastic gradient descent with momentum (SGDM)
Learning rate	0.001
Weight decay	0.01
Scheduler	Cosine Annealing
Batch size	16
Number of iterations per epoch	500
Total epoch	300
Number of epochs for the first restart	20
Downscaling ratio	10
Data augmentation	Rotation, flip, random scale, cutmix
Patch size	256
Loss functions	Mean square error loss for SIC, cross entropy loss for SOD and FLOE
Total loss calculation	SIC×1+SOD×3+FLOE×3
Number of validation scenes	20

All experiments were conducted on the Narval cluster of Compute Canada, Canada's national high-performance computing system. The experiments utilized a NVIDIA A100-SXM4-40GB GPU with 128GB of RAM memory, employing the PyTorch 1.12 library. It takes an average of about 3.5 hours to train the proposed model.

4 Experimental Results

Out of numerous submissions on the leaderboard, we achieved the highest combined score of approximately 86.3%, as well as the highest SIC and SOD scores. As the ice chart-derived labels for the testing data were released subsequent to the conclusion of the competition, we conducted additional model retraining using diverse configurations to obtain more comprehensive statistical outcomes for detailed analysis.

The statistical results obtained from model validation and testing are summarized in Table 4. Different configurations of trained models are represented by distinct model numbers (Model 1 to Model 9) and plotted in different colors, as specified in Table 4. Model 1 corresponds to the proposed model with settings described in Section 3 and Table 3. The remaining models serve as ablation studies to validate the effectiveness of various components/features in Model 1, with modifications detailed in Table 4. Each score in a certain model corresponds to the average score of the 20 networks trained with the same configuration. The relatively large standard deviation (STD) values of the combined scores in validation are caused by the randomness in validation scene selection. In contrast, the STDs of combined scores in testing are much smaller (around 1%). Through comparison, the following findings are validated:

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

Table 4. The statistical results of models with different configurations. Model 1 is developed using the specifications introduced above. Compared to Model 1, Models 2-7 change the combinations of data inputs, Model 8 changes the loss function for SIC, and Model 9 splits the decoder into three separate parts for the three parameters.

Model number	Modifications compared to Model 1	Mean validation accuracy (%)					Mean testing accuracy (%)				
		Combined score	Standard deviation	SIC	SOD	FLOE	Combined score	Standard deviation	SIC	SOD	FLOE
1	N/A	91.6	2.2	93.9	92.2	85.7	86.5	1.2	91.7	87.2	73.7
2	Remove SAR downscaling	85.6	2.5	88.9	86.1	78.1	79.7	1.4	84.4	80.7	68.4
3	Remove all data inputs except HH, HV	87.9	2.6	91.1	86.8	83.5	78.6	1.4	84.8	75.1	73.1
4	Remove AMSR2 data inputs	91.3	2.6	93.1	92.6	85.2	82.2	0.7	85.3	84.4	71.5
5	Remove ERA5 data inputs	91.7	2.4	93.6	92.6	86.3	85.2	0.6	90.4	86.5	72.0
6	Remove spatial-temporal encoding	88.7	2.1	92.9	86.8	83.9	82.5	0.8	91.1	78.1	73.8
7	Add all available data inputs not used in Model 1	91.5	2.3	93.6	91.8	86.6	86.5	0.6	91.3	88.7	73.3
8	Replace MSE loss with CE loss for SIC	90.7	2.3	91.4	92.9	85.0	83.5	1.2	86.7	85.8	72.7
9	Change the shared decoder to separate decoders	91.7	2.1	93.4	92.2	87.2	87.3	0.7	91.7	88.2	76.4

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

- Downsampling the SAR data inputs significantly improves the mapping accuracy (Model 1 vs. Model 2), with improvements of 6.8% in average testing combined score, 7.3% in SIC, 6.5% in SOD, and 5.3% in FLOE. Furthermore, this downsampling enhancement also leads to a substantial increase in computational efficiency. Training the proposed model takes approximately 3.5 hours, while producing a map using the forward model for a SAR scene only requires an average of around 2 seconds. In contrast, without downsampling, the average training time is approximately 15 times longer. Various downsampling ratios were tested, and a value of 10 yielded one of the best results along with high efficiency.

- The inclusion of multi-source input channels is essential, as demonstrated by the comparison between Model 1 and Model 3. Using only SAR data inputs results in lower SIC and SOD scores by 6.9% and 12.1%, respectively. Although the removal of AMSR2 (Model 4) or ERA5 (Model 5) data inputs does not affect validation scores significantly, a drop in testing accuracy can be observed. This is particularly evident in the model without AMSR2 inputs, where the average SIC and SOD testing scores decrease by 6.4% and 2.8% compared to the proposed model. Thus, the inclusion of brightness temperature data plays a vital role in enhancing model accuracy.

- The effectiveness of spatial-temporal encoding in improving accuracy, particularly the SOD score, is evident in the comparison between Models 1 and 6. This is likely because the model in Model 1 can learn the distribution of dominant ice types in different Arctic regions during different months based on the training data, resulting in a 9.1% improvement in average SOD score during testing. Additionally, compared to the model utilizing all available data as inputs (Model 7), the proposed model with feature selection (selecting a subset of AMSR2 and ERA5 data) achieves nearly the same accuracy while improving efficiency.

- Adopting MSE loss for SIC, as opposed to CE loss (Model 8), increases the average SIC testing score significantly by 5.0% and improves the average testing combined score by 3.0%.

- Despite improvements in SIC and SOD scores, the FLOE scores remain relatively low, with a significant gap between validation and testing accuracy. After exploring numerous configurations, we found that only downscaling and separating the decoders for the three parameters (Model 9) might enhance the FLOE score. Visually, it is challenging to distinguish patterns of different floe sizes from SAR imagery. The mapping results of FLOE will be further discussed in the visual analysis below.

In addition to numerical results, visual interpretation is essential for analysis. Sea ice mapping results from two example SAR scenes in the testing data, which were obtained using models with different configurations, are presented in Figs. 2 and 3. Fig. 2 illustrates that implementing input downscaling (including Model 1 and Model 8) enhances the consistency between the ice-water boundaries in the label maps and the model predictions. With a larger receptive field, contextual information is captured by the model, leading to polygon-based predictions. Conversely, without downscaling, the extracted features only contain local intensity information, limiting the model's ability to capture the presence of ice in surrounding areas, as demonstrated in the row corresponding to Model 2. Although larger patch sizes (e.g., 512, 768) also provide a larger receptive field instead of input

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

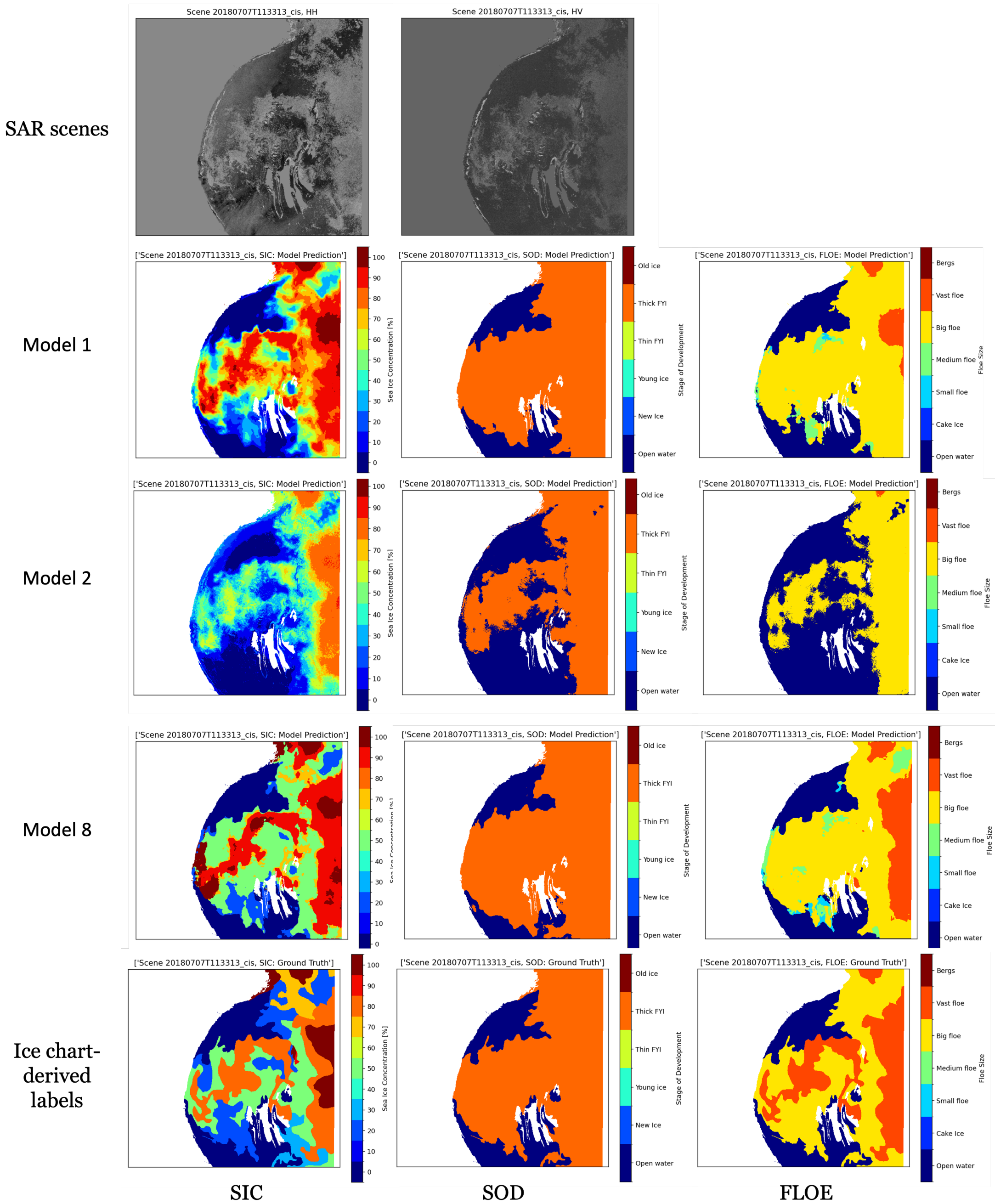

Figure 2. Sea ice mapping results obtained from a SAR scene (ID: 20180707T113313_cis) in the testing data using models trained with different configurations indicated by experiment numbers on the left. The ice chart-derived labels are displayed in the last row for comparison. Areas that are land or without labels are masked in white.

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

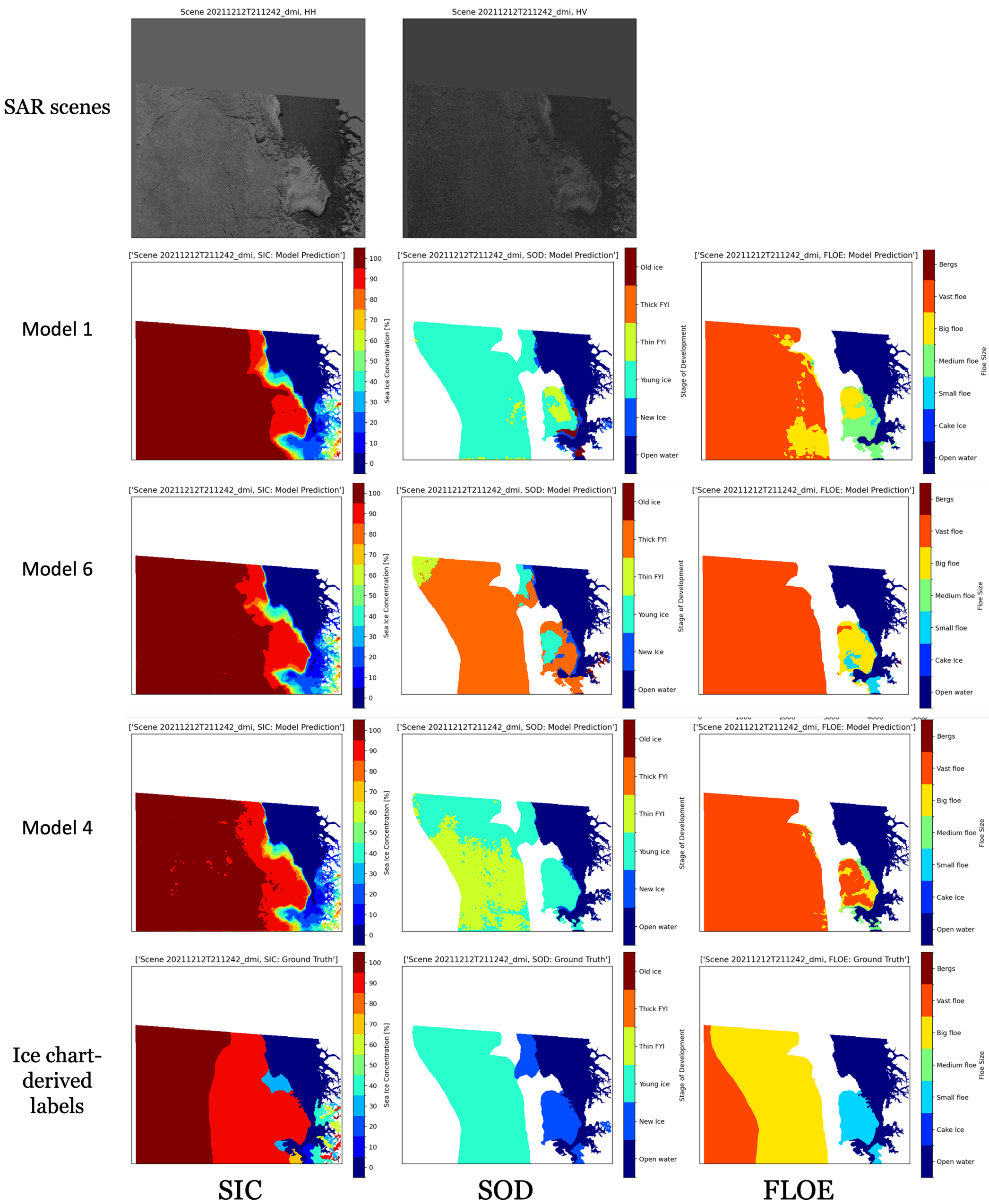

Figure 3. Sea ice mapping results obtained from a SAR scene (ID: 20211212T211242_dmi) in the testing data using models trained with different configurations indicated by experiment numbers on the left. The ice chart-derived labels are displayed in the last row for comparison. Areas that are land or without labels are masked in white.

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

downscaling, we find that these models perform much worse than a patch size of 256. Furthermore, while choosing CE loss for SIC yields lower accuracy than MSE loss, the predictions consist of larger polygons that visually align more closely with the SIC label map, as seen in Model 8. This finding is consistent with the observations in (Kucik and Stokholm 2023).

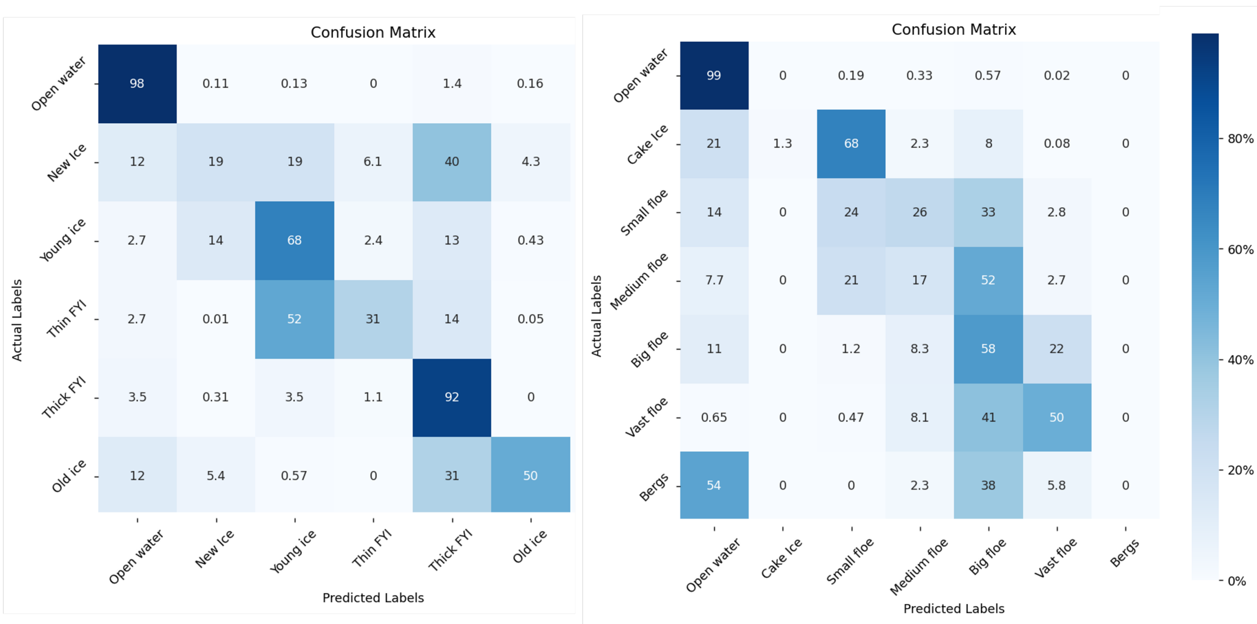

Figure 4. The confusion matrices of SOD (left) and FLOE (right) testing results. The "255" class represents the mask and is not considered for performance evaluation.

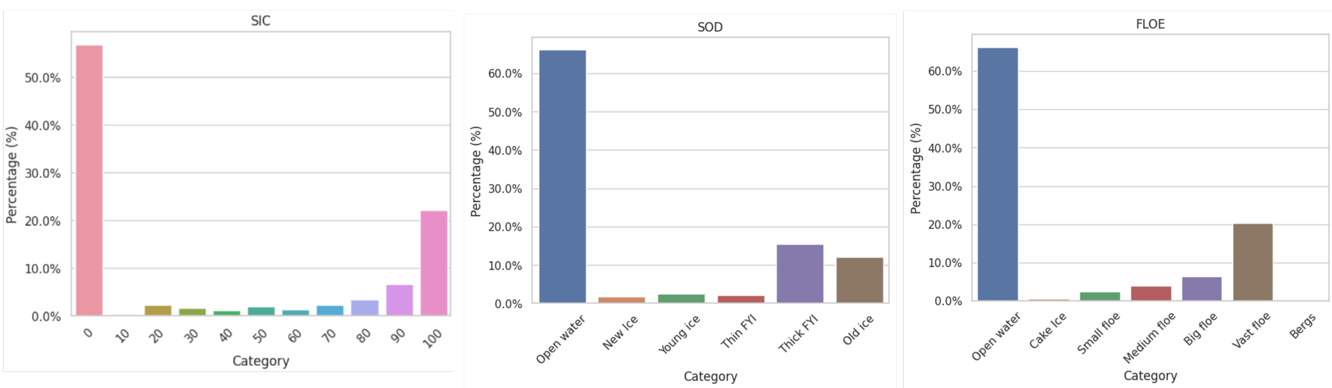

Figure 5. The distribution of training samples for each class in the three parameters (SIC on the left, SOD in the middle, and FLOE on the right). The bars of some categories may be invisible due to very low percentages (e.g., 0.43% for 10% SIC, 0.48% for cake ice, and 0.17% for bergs in FLOE).

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

Furthermore, the effectiveness of spatial-temporal encoding in improving SOD prediction accuracy can be observed by comparing Model 1 and Model 6 in Fig. 3. In Model 6, where spatial-temporal encoding is not applied, a large area labeled as young ice is misclassified as thick first-year ice (FYI). The model without AMSR2 inputs (Model 4) also misclassifies a relatively large area as thin FYI. Despite achieving relatively high SOD accuracy, there are some classes with significant misclassification rates. For instance, as shown in the confusion matrices in Fig. 4, the classification accuracies for new ice and thin FYI are only 19% and 31%, respectively. Misclassifications between ice types with neighboring thickness are also prevalent. For example, 31% of old ice samples are misclassified as thick FYI. These issues may be attributed to various factors, such as the highly imbalanced distribution of samples among ice types. As depicted in Fig. 5, new ice and thin FYI have the most and second least samples in the training data (comprising only around 2% of the total). Additionally, the labeling method of SOD and FLOE in the ready-to-train dataset might contribute to these challenges. Although most polygons in ice charts contain multiple ice types and floe sizes, they are labeled with only the dominant classes due to a lack of pixel-based labels, leading to inevitable labeling errors. During the competition, we attempted several strategies to address the issue of sample imbalance, such as implementing focal loss (Lin et al. 2017). However, none of these approaches significantly improved the accuracy of the minority classes thus far.

5 Conclusions

In this paper, we present our MMSeaIce pipeline, which consists of a multi-task U-Net for automated sea ice parameter retrieval from the ready-to-train version. To enable our model to learn contextual information within a large receptive field, we initially apply a downscaling operation to the SAR data inputs. This enhances the consistency between model predictions and ice chart-derived labels, resulting in a remarkable improvement of 6.8% in the combined score and significant enhancement in computation speed. We conducted ablation studies to investigate the impact of different data inputs on model performance. These studies demonstrate the necessity of including brightness temperature data, which leads to a 4.3% improvement in the average combined score, as well as the importance of incorporating spatial-temporal information, which contributes to a 4.0% improvement in the combined score. Additionally, we show that other modifications to the model, such as applying the MSE loss in SIC retrieval during training and employing separate decoders for the three parameters, also improve the overall performance. The best model we developed achieves an average combined score of 87.3% on the testing dataset, with average individual scores of 91.7%, 88.2%, and 76.4% for SIC, SOD, and FLOE, respectively.

Despite our success in the competition, there are still several areas that require further investigation to derive robust and accurate automated sea ice maps with high resolution. For instance, it is crucial to propose a new labeling method that adequately addresses polygons with mixed ice types or floe sizes. Furthermore, with the upcoming release of an updated AI4Arctic dataset containing a significantly larger volume of data, we recommend retraining our proposed model to improve the predictive accuracy of the minority classes. Additionally, considering the spatial and temporal variation of sea ice in SAR imagery, training models specific to certain regions or seasons, particularly the melting season, would be a preferable approach for enhancing performance.

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

Code availability. The codes will be available after the publication of the paper.

Data availability. The AI4Arctic dataset are available from Buus-Hinkler et al. 2022a (Accessed on 01-Jun-2023).

Author contributions. XC is the team lead for the competition and wrote the manuscript. XC, MP, FPC, JNT and JP built the models and performed the analysis, LX, KS, and DC supervised the competition and provided suggestions for performance improvement. All authors contributed to revised manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors acknowledge the providers of the AI4Arctic dataset and the organizers of the AutoIce competition. This work was supported in part by Environment and Climate Change Canada (ECCC) and the Natural Sciences and Engineering Research Council (NSERC) of Canada. This research was enabled in part by support provided by Calcul Québec's (calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca).

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

References

Boulze, H., Korosov, A., and Brajard, J.: Classification of sea ice types in Sentinel-1 SAR data using convolutional neural networks, Remote Sens., 12, 2165, https://doi.org/10.3390/rs12132165, 2020.

Buus-Hinkler, J., Wulf, T., Stokholm, A. R., Korosov, A., Saldo, R., Pedersen, L. T., Arthurs, D., Solberg, R., Longépé, N., and Brandt Kreiner, M.: AI4Arctic Sea Ice Challenge Dataset, https://doi.org/10.11583/DTU.c.6244065.v2, 2022a.

Buus-Hinkler, J., Wulf, T., Stokholm, A. R., Korosov, A., Saldo, R., Pedersen, L. T., and et al.: AI4Arctic Sea Ice Challenge Dataset, Technical University of Denmark. Collection, https://doi.org/10.11583/DTU.c.6244065.v2, 2022b.

Chen, S., Shokr, M., Li, X., Ye, Y., Zhang, Z., Hui, F., and Cheng, X.: MYI floes identification based on the texture and shape feature from dual-polarized Sentinel-1 imagery, Remote Sens., 12, 3221, 2020.

Chen, X., Scott, K. A., Jiang, M., Fang, Y., Xu, L., and Clausi, D. A.: Sea Ice Classification With Dual-Polarized SAR Imagery: A Hierarchical Pipeline, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 224–232, 2023a.

Chen, X., Valencia, R., Soleymani, A., and Scott, K. A.: Predicting Sea Ice Concentration With Uncertainty Quantification Using Passive Microwave and Reanalysis Data: A Case Study in Baffin Bay, IEEE Trans. Geosci. Remote Sens., 61, 1–13, 2023b.

Chi, J., Kim, H.-c., Lee, S., and Crawford, M. M.: Deep learning based retrieval algorithm for Arctic sea ice concentration from AMSR2 passive microwave and MODIS optical data, Remote Sens. Environ., 231, 111 204, 2019.

Cooke, C. L. and Scott, K. A.: Estimating sea ice concentration from SAR: Training convolutional neural networks with passive microwave data, IEEE Trans. Geosci. Remote Sens., 57, 4735–4747, 2019.

De Gelis, I., Colin, A., and Longépé, N.: Prediction of categorized Sea Ice Concentration from Sentinel-1 SAR images based on a Fully Convolutional Network, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 14, 5831–5841, https://doi.org/10.1109/JSTARS.2021.3074068, 2021.

Guo, W., Itkin, P., Singha, S., Doulgeris, A. P., Johansson, M., and Spreen, G.: Sea ice classification of TerraSAR-X ScanSAR images for the MOSAiC expedition incorporating per-class incidence angle dependency of image texture, The Cryosphere, 17, 1279–1297, 2023.

Huang, Y., Ren, Y., and Li, X.: Classifying Sea Ice Types from SAR Images Using a U-Net-Based Deep Learning Model, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 3502–3505, IEEE, https://doi.org/10.1109/IGARSS47720.2021.9554511, 2021.

Jiang, M., Clausi, D. A., and Xu, L.: Sea Ice Mapping of RADARSAT-2 Imagery by Integrating Spatial Contexture With Textural Features, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., https://doi.org/10.1109/JSTARS.2022.3205849, 2022.

Khaleghian, S., Ullah, H., Kræmer, T., Hughes, N., Eltoft, T., and Marinoni, A.: Sea Ice Classification of SAR Imagery Based on Convolution Neural Networks, Remote Sensing, 13, 1734, https://doi.org/10.3390/rs13091734, 2021a.

Khaleghian, S., Ullah, H., Krmer, T., Eltoft, T., and Marinoni, A.: Deep Semi-Supervised Teacher-Student Model based on Label Propagation for Sea Ice Classification, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., https://doi.org/10.1109/JSTARS.2021.3119485, 2021b.

Kruk, R., Fuller, M. C., Komarov, A. S., Isleifson, D., and Jeffrey, I.: Proof of concept for sea ice stage of development classification using deep learning, Remote Sens., 12, 2486, 2020.

Kucik, A. and Stokholm, A.: AI4SeaIce: selecting loss functions for automated SAR sea ice concentration charting, Sci. Rep., 13, 5962, 2023.

Li, X.-M., Qiu, Y., Wang, Y., Huang, B., Lu, H., Chu, M., Fu, H., and Hui, F.: Light from space illuminating the polar silk road, Int. J. Digit. Earth, 15, 2028–2045, 2022.

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P.: Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988, 2017.

Liu, H., Guo, H., and Liu, G.: A Two-Scale Method of Sea Ice Classification Using TerraSAR-X ScanSAR Data During Early Freeze-Up, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., 2021a.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 10 012–10 022, 2021b.

Loshchilov, I. and Hutter, F.: SGDR: Stochastic gradient descent with warm restarts, arXiv preprint arXiv:1608.03983, 2016.

Lyu, H., Huang, W., and Mahdianpari, M.: Eastern arctic sea ice sensing: First results from the RADARSAT Constellation Mission data, Remote Sens., 14, 1165, 2022a.

Lyu, H., Huang, W., and Mahdianpari, M.: A Meta-Analysis of Sea Ice Monitoring Using Spaceborne Polarimetric SAR: Advances in the Last Decade, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., 2022b.

Mahmud, M. S., Nandan, V., Singha, S., Howell, S. E., Geldsetzer, T., Yackel, J., and Montpetit, B.: C-and L-band SAR signatures of Arctic sea ice during freeze-up, Remote Sens. Environ., 279, 113 129, https://doi.org/10.1016/j.rse.2022.113129, 2022.

Malmgren-Hansen, D., Pedersen, L. T., Nielsen, A. A., Kreiner, M. B., Saldo, R., Skriver, H., Lavelle, J., Buus-Hinkler, J., and Krane, K. H.: A convolutional neural network architecture for Sentinel-1 and AMSR2 data fusion, IEEE Trans. Geosci. Remote Sens., 59, 1890–1902, 2020.

Minnett, P., Alvera-Azcárate, A., Chin, T., Corlett, G., Gentemann, C., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E., Santoleri, R., Saux Picart, S., Steele, M., and Vazquez-Cuervo, J.: Half a century of satellite remote sensing of sea-surface temperature, Remote Sens. Environ., 233, 111 366, 2019.

Nagi, A. S., Kumar, D., Sola, D., and Scott, K. A.: RUF: Effective sea ice floe segmentation using end-to-end RES-UNET-CRF with dual loss, Remote Sens., 13, 2460, 2021.

Radhakrishnan, K., Scott, K., and Clausi, D.: Sea Ice Concentration Estimation: Using Passive Microwave and SAR Data With a U-Net and Curriculum Learning, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 14, 5339–5351, 2021.

Ren, Y., Li, X., Yang, X., and Xu, H.: Development of a Dual-Attention U-Net Model for Sea Ice and Open Water Classification on SAR Images, IEEE Geosci. Remote Sens. Lett., https://doi.org/10.1109/LGRS.2021.3058049, 2021.

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

Soleymani, A. and Scott, K. A.: Evaluation of a Neural Network on Sea Ice Concentration Estimation in MIZ Using Passive Microwave Data, in: IGARSS, pp. 5656–5659, 2021.

Song, W., Li, M., Gao, W., Huang, D., Ma, Z., Liotta, A., and Perra, C.: Automatic Sea-Ice Classification of SAR Images Based on Spatial and Temporal Features Learning, IEEE Trans. Geosci. Remote Sens., https://doi.org/10.1109/TGRS.2020.3049031, 2021.

Stokholm, A., Wulf, T., Kucik, A., Saldo, R., Buus-Hinkler, J., and Hvidegaard, S. M.: AI4SeaIce: Toward Solving Ambiguous SAR Textures in Convolutional Neural Networks for Automatic Sea Ice Concentration Charting, IEEE Trans. Geosci. Remote Sens., 60, 1–13, 2022.

Stokholm, A., Buus-Hinkler, J., Wulf, T., Korosov, A., Saldo, R., Arthurs, D., Solberg, R., Longépé, N., and Kreiner, M.: The AutoICE Competition: Automatically Mapping Sea Ice in the Arctic, Tech. rep., Copernicus Meetings, 2023a.

Stokholm, A., Kucik, A., Longépé, N., and Hvidegaard, S. M.: AI4SeaIce: Task Separation and Multistage Inference CNNs for Automatic Sea Ice Concentration Charting, EGUsphere, 2023, 1–25, 2023b.

https://doi.org/10.5194/egusphere-2023-1297
Preprint. Discussion started: 24 October 2023

Wang, L., Scott, K. A., Xu, L., and Clausi, D. A.: Sea ice concentration estimation during melt from dual-pol SAR scenes using deep convolutional neural networks: A case study, IEEE Trans. Geosci. Remote Sens., 54, 4524–4533, https://doi.org/10.1109/TGRS.2016.2543660, 2016.

Wang, L., Scott, K. A., and Clausi, D. A.: Sea ice concentration estimation during freeze-up from SAR imagery using a convolutional neural network, Remote Sens., 9, 408, https://doi.org/10.3390/rs9050408, 2017.

Zhang, T., Yang, Y., Shokr, M., Mi, C., Li, X.-M., Cheng, X., and Hui, F.: Deep Learning Based Sea Ice Classification with Gaofen-3 Fully Polarimetric SAR Data, Remote Sensing, 13, 1452, https://doi.org/10.3390/rs13081452, 2021a.

Zhang, Y., Zhu, T., Spreen, G., Melsheimer, C., Huntemann, M., Hughes, N., Zhang, S., and Li, F.: Sea ice and water classification on dual-polarized Sentinel-1 imagery during melting season, The Cryosphere Discussions, 2021, 1–26, https://doi.org/10.5194/tc-2021-85, 2021b.