

Data augmentation for SAR sea ice and water classification based on per-class backscatter variation with incidence angle

Qiang Wang, Johannes Lohse, Anthony P. Doulgeris, *senior member, IEEE* and Torbjørn Eltoft, *member, IEEE*

Abstract—Monitoring sea ice in polar regions is critical for understanding global climate change and supporting marine navigation. Recently, researchers started to utilize machine/deep learning methodologies to automate the separation of sea ice and open water in synthetic aperture radar imagery. However, this requires a large amount of reliably labeled training data. We here propose an augmentation routine for Sentinel-1 data which incorporates physical principles of radar backscatter into the augmentation procedure. Firstly, we apply an incidence angle aware algorithm to segment Sentinel-1 images into separate clusters. We compute the corresponding slopes of backscatter intensity with the incidence angle for each cluster. Secondly, the slopes are used as prior information to project labeled pixels and segments to different incidence angles and thus further enrich the labeled data. We then apply a simplified U-Net for pixel-wise classification of Sentinel-1 images into sea ice or open water. The performance of our model is evaluated by visual inspection as well as comparison with an available product from the Chinese Academy of Science (CAS). The results indicate that the physics-based data augmentation improves the model performance compared to training with data without augmentation. The inferred ice edge is in line with the inference for other available data sets (CAS), but with a finer spatial resolution. Finally, we also found the inference of our model is highly correlated with the visual interpretation of overlapping optical observations. Overall, the proposed methodology provides an alternative for the automated separation of sea ice/open water at fine spatial resolution.

Index Terms—Sea ice classification, data augmentation, deep learning, polar region

I. INTRODUCTION

SEA ice is an important component of the earth climate system and a sensitive indicator of climate change. It is a threat to industrial offshore operations, as well as marine navigation in polar areas. Since 1979, Arctic sea ice coverage has decreased for all months of the year, with the largest decline in the summer and early fall. In September, at the end of melt season, the ice extent is now about half of the 1979 extent [1]. This significant sea ice loss has several climatological consequences: For instance, it leads to wetter European summers [2] and more extreme Northern Hemisphere winters [3]. Furthermore, it affects the Arctic marine ecosystems, which is for example indicated by the physiological adaptations of polar bears and narwhals [4], and it makes the Arctic more easily accessible for shipping and offshore operations [5]. Accurate and timely information

about sea ice conditions is therefore crucial for safe Arctic operations, for improving weather and climate models, and it can be used to support policy makers in making local, regional as well as global regulations to prevent deterioration of the Arctic ecosystem.

Sea ice extent (SIE) and sea ice concentration (SIC) are two main parameters to describe the state of the Arctic sea ice cover. At present, passive microwave radiometer (PMR) observations are the main data source for pan-Arctic mapping of SIE and SIC. For instance, researchers from the University of Bremen have developed a SIC product based on Advanced Microwave Scanning Radiometer 2 (AMSR-2) data, using an algorithm developed during the Arctic radiation and turbulence interaction study (ARTIST) [6]. Similarly, the Bootstrap (BT) algorithm [7] is being employed by NASA to provide a near-real-time NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration (<https://nsidc.org/data/g10016>), utilizing observations since 1978.

Both of these data sets are widely used by researchers for sea ice climatology studies. However, the spatial resolution of the PMR sensors is coarse, on the order of several kilometers. This is not suitable for operational ice charting to support marine navigation, which requires continuous information on a much finer spatial resolution. Such information can only be reliably provided by spaceborne synthetic aperture radar (SAR) observations, which are independent of sun illumination and cloud coverage. The weather independent and reliable imaging capability makes SAR the main data source in national sea ice services around the world. At present, operational sea ice charts are produced manually by visual interpretation of the imagery by experienced sea ice analysts [8], a process that is both subjective and time-consuming. Furthermore, the manually drawn ice charts do not always reflect the fine pixel-level spatial detail of the underlying SAR imagery, as is for example shown in the ice charts produced by the Danish Meteorological Institute (DMI) in Fig.1) (https://data.dtu.dk/articles/dataset/AI4Arctic_ASIP_Sea_Ice_Dataset_-_version_2/13011134).

Hence, a vast number of scholars have made efforts towards automating the ice charting process from SAR at the pixel-level. For instance, Lohse et al. [9] utilized a Bayesian classifier for sea ice type mapping based on Sentinel-1 HH and HV intensities, incidence angle (IA), and selected grey level co-occurrence matrix (GLCM) texture features. Cristea et al. [10] proposed an unsupervised sea ice segmentation algorithm by integrating the target-specific IA-dependent intensity decay rates into a non-stationary statistical model.

With the recent advancement of artificial intelligence technology, researchers also started to utilize machine learning methods for sea ice mapping. Zakhvatkina et al. [11] utilized a support vector machine (SVM) classifier, based on Radarsat-2 HH and HV backscatter intensities as well as texture features, to separate sea ice and open water. Park et al. [12] developed a semi-automated sea ice classifier based on a random forest algorithm, using sea ice charts to extract training data. It is known that the input features used to train classification algorithms (e.g. backscatter intensities or texture features with selected parameter settings) affect the generalization and performance. Traditionally, these features are manually selected. In deep neural networks (DNNs), which are now being extensively explored also for sea ice applications, the networks automatically learn the optimal features through a hierarchy of computational layers. For instance, Khaleghian et al. [13] developed a modified convolutional neural network (CNN) architecture based on Visual geometric Group 16-layer (VGG-16) for binary sea ice versus open water classification. They also studied multi-class sea ice type classification using the same architecture. Malmgren-Hansen et al. [14] proposed a CNN architecture for fusing Sentinel-1 SAR imagery and AMSR-2 PMR data for ice charting, while Stokholm et al. [15] used a modified U-Net architecture, where they included the SAR noise correction scheme developed by the Nansen Environmental and Remote Sensing Center (NERSC) [16]. Wang and Li [17] utilized multiple U-Net classifiers to map the Arctic sea ice cover from Sentinel-1 imagery and found a 5.55% difference compared with AMSR-2 sea ice concentration data.

Successful sea ice classification based on DNN algorithms requires a large number of high-quality training data with known class labels. Several such training data sets are publicly available. However, most of these data sets are based either on manual/visual analysis of the SAR imagery or on sea ice charts produced by the national ice services (which are eventually also based on manual analysis). Consequently, the class labels often lack detailed spatial information at the SAR pixel level. This is the case for the previously mentioned DMI data set (Fig.1) as well as for example for the UiT data set (<https://zenodo.org/record/4683174#.Yk7kVshBwuV>) shown in Fig.2. To increase the objectivity and robustness of DNN sea ice mapping algorithms, more high-quality training data at pixel-level resolution is required. While the manual selection of such training data is possible, it will be even more time-consuming than the generation of "traditional" training data sets with less spatial detail (Fig.1 and Fig.2). Hence, less human intervention in training set generation is desirable.

Data augmentation is a method to increase the quantity of training samples without human intervention. It is a fairly new methodology, which has proven successful to overcome the challenge of small data sets [18] or data sets with imbalanced class abundances [19] [20]. Such imbalance can be a problem in the case of ice type mapping, where sea ice and water may not be evenly distributed or the number of pixels for various ice types may vary. Data augmentation also reduces the overfitting phenomenon and increases the generalization ability of CNNs. Hence, it provides an alternative option in cases when

a small labelled data set already exists. Many researchers have explored various data augmentation methods. For instance, operations such as rotating, flipping, enlarging the contrast among pixels, and blurring have been widely implemented in computer vision related tasks, and were adopted by Khaleghian et al. [13] for sea ice applications. Zhang et al. [21] utilized the convex combination of multiple samples and labels to create new samples, increasing the diversity of the sampling and the robustness of the model performance. Furthermore, variants of generative adversarial networks (GAN) have been proposed by Salimans et al. [22] and been adopted to data augmentation for imbalanced data sets via transformation [23], [24]. Apart from the above methods for data augmentation, a number of researchers have utilized physics-based data augmentation in training deep learning networks [25], [26], [27].

Motivated by the aforementioned works, we propose in this manuscript a new physics-based data augmentation approach for extending the training data used for training a DNN to separate sea ice and open water at the individual SAR pixel level. In the model, we take advantage of the physics that the IA dependence of the radar backscatter intensity is different for sea ice and open water, which is being demonstrated by several research papers [9], [10], [28].

This article is organized as follows: Section II describes the selected study area, as well as the utilized data set. Section III presents the methodology, including the training data generation with physics-based augmentation, the CNN model architecture, and the configuration of the training process. Section IV presents the performance of the proposed augmentation approach by depicting classification results, with error metrics of the model, as well as comparison with other ice chart products. Section V discusses the problem of noisy pixels over transition areas in the scenes as well as the open water inference. Meanwhile, possible future solutions to these issues are proposed. The main findings are concluded in section VI, giving several perspectives on the current model performance and further steps to take in future work.

II. STUDY AREA AND SATELLITE DATA

A. Study area

In this study we focus on the geographic region between Greenland and Svalbard (22°W-25°E, 69°N-83°N). The air temperature at 2m above the surface increases from March to June and then decreases afterwards. The lowest temperature can reach to -39.1°C on March 10th, while the temperature reach at a peak of 13.4°C on June 30th according to Danmarkshavn station's measurements carried out by DMI over 2020 (<https://confluence.govcloud.dk/display/FDAPI>).

This region is characterized by large open water areas as well as variable and highly dynamic sea ice conditions. From East to West (Svalbard to Greenland), the region generally shows a transition from predominantly open water conditions close to Svalbard, via a marginal ice zone (MIZ) with SIC between 0.2 and 0.8, towards dense drift ice and finally landfast sea ice close to the East Greenland coast. The landfast ice area largely consists of rough and strongly deformed rubble fields, which are built up from drift ice that was locked in

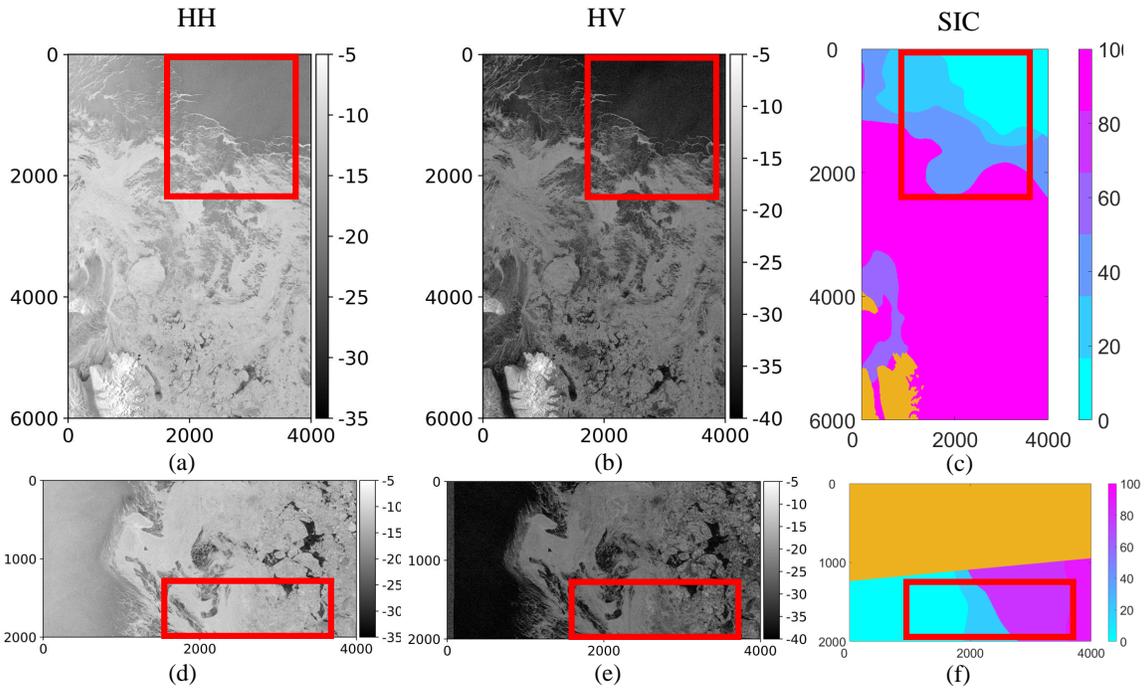


Fig. 1. HH (a,d) and HV (b,e) intensity (in dB) for selected regions of two S1 example images (scene NO.9 and 10 in Table I), together with SIC (c,f) from the corresponding DMI ice chart. Note the differences in spatial detail (indicated by red rectangles) between the original SAR images and the manually drawn ice charts. The SAR images contain more spatial information than what is captured in the ice chart. If the classifier is expected to reproduce these details, they must be included in the training.

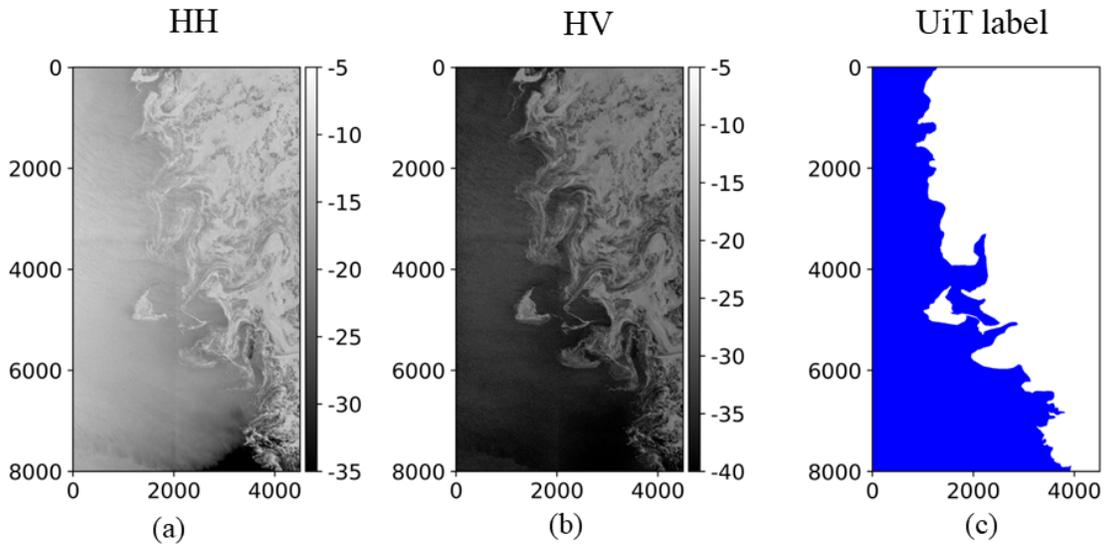


Fig. 2. HH and HV intensity (in dB) for a selected region of an S1 example image (scene NO.11 in Table I), together with the corresponding ice-water labels from the UiT data set. While the ice-water mask captures a fair amount of spatial detail, it does not reflect the pixel-level spatial information that is visible in the SAR imagery, especially in the area close to the ice edge.

place and became landfast. These rubble fields are intersected by regions of very smooth and level landfast ice, which formed in-situ under wind- and wave-protected conditions. In the drift ice area, a mix of first- and multi-year ice can be found, intersected by ridges and rubble fields as well as leads that can either show open water conditions or be covered by thin nilas or young ice. The open water area closer to Svalbard provides a large fetch for wind and can therefore show variable SAR backscatter intensity due to changing sea surface state caused by varying wind and ocean current conditions. Overall, with its diverse sea ice conditions in combination with its strategic importance for Arctic operations, the selected area provides a suitable and challenging test region for our present study.

B. Satellite data

1) *Sentinel-1 SAR data:* Sentinel-1 is a SAR mission with global observation coverage and is part of the European Space Agency’s (ESA) Copernicus program. It currently consists of two satellites, with Sentinel-1A launched on 3 April 2014 and Sentinel-1B on 25 April 2016, respectively. Each individual satellite has a 12-day repeat cycle at the equator, which results in a combined 6-day repeat cycle for both satellites. The revisit frequency increases from low to high latitudes, resulting in a temporal resolution of 1-3 days in Arctic regions. It should be noted that the Sentinel-1B mission ended on 23 Dec 2021 due to a failure in the instrument electronics power supply. Sentinel-1 data is freely available and can for example be downloaded via the Copernicus open access hub (<https://scihub.copernicus.eu/>).

Sentinel-1 operates at C-band frequency and offers four exclusive acquisition modes with different swath widths: Stripmap (SM), Interferometric Wide swath (IW), Extra-Wide swath (EW), and Wave (WV) mode. In this study, we use EW data in ground range detected (GRD) format at medium resolution. The EW mode is the standard observation mode over open water and has a swath width of approximately 410 km. It covers roughly $10,000 \times 10,000$ pixels in total with each pixel corresponding to approximately 40×40 m on the ground. Meanwhile, the incidence angle across the swath ranges from 18.9° at near-range to 47.0° at far-range. Several studies have shown that the noise floor has a strong influence on Sentinel-1 observations of sea ice [29]. To mitigate noise effects, we apply the de-noising procedure proposed by the Nansen Environmental and Remote Sensing Center (NERSC) [16]. After noise correction, we clip the HH and HV intensity values to intervals ranging from -30 to 0 dB and -35 to -5 dB, respectively, to remove the outliers.

Each Sentinel-1 scene can be uniquely identified by its name string. For the easier readability of this paper, we use a running ID number for all S1 images that we explicitly refer to in the text. A list that matches the running IDs with the unique name strings of the corresponding S1 images is given in Table I of the supplementary information.

In this work, 8 S1 EW images acquired between January and March 2021 (see NO.1-8 of Table I for scene IDs) are selected to generate the training data for later usage (see green rectangles in Fig.3). The selection of these images is based

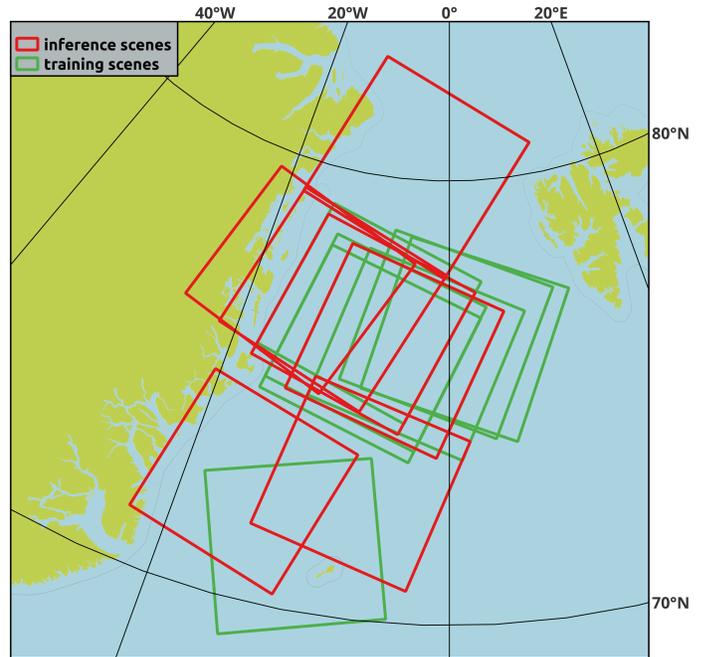


Fig. 3. Overview of all image footprints used for either training (green) or inference (red) in this study.

on the criteria that they should cover different geo-locations, varying sea ice conditions, as well as both ascending and descending image acquisitions. The footprints of all training images as well as all images used for inference and validation in this study are shown in Fig. 3.

2) *Sentinel-2 optical data:* Like Sentinel-1, Sentinel-2 is part of ESA’s copernicus program and currently consists of two satellites, Sentinel-2A (launched on 23 June 2015) and Sentinel-2B (launched on 7 March 2017). Both satellites carry a MultiSpectral Instrument (MSI) that samples 13 spectral bands, with the central wavelength of the individual bands ranging from 443 nm to 2190 nm at a spatial resolution of 10 m, 20 m, or 30 m. Sentinel-2 provides a swath width of 290 km and the twin satellites give a combined revisit frequency of five days at the equator. Detailed configuration can be found via (<https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>) and data can freely be download via the Copernicus open access hub.

Operating in the visual and near-infrared part of the electromagnetic spectrum, Sentinel-2 is affected by natural sunlight and cloud conditions. It is therefore not suitable as a primary sensor for operational sea ice charting, but can be highly useful as complementary data. In this study, we use Sentinel-2 data as auxiliary data to evaluate and judge the performance of our deep learning algorithm, which is based purely on Sentinel-1 imagery.

III. METHOD

In this section we describe the details of our study method. We first give a short and concise overview of the experiment setup, then we describe the details of the different experiments, including training data selection and data augmentation, model

TABLE I
S1 SCENE IDS AND CORRESPONDING INTERNAL SCENE NUMBER IN THIS MANUSCRIPT

scene number	purpose	scene ID
1	train	S1B_EW_GRDM_1SDH_20210101T074539_20210101T074639_024958_02F86C_D497
2	train	S1A_EW_GRDM_1SDH_20210102T073823_20210102T073923_035956_043634_DE5A
3	train	S1A_EW_GRDM_1SDH_20210104T072154_20210104T072254_035985_043735_D4FF
4	train	S1B_EW_GRDM_1SDH_20210105T071250_20210105T071350_025016_02FA36_6592
5	train	S1B_EW_GRDM_1SDH_20210209T175419_20210209T175519_025533_030AD7_84BD
6	train	S1A_EW_GRDM_1SDH_20210301T070437_20210301T070537_025818_031425_6693
7	train	S1B_EW_GRDM_1SDH_20210304T072924_20210304T073024_025862_03159B_D26A
8	train	S1B_EW_GRDM_1SDH_20210309T073735_20210309T073835_025935_0317F0_F5F5
9	evaluation	S1A_EW_GRDM_1SDH_20190110T183531_20190110T183635_025419_02D0D0_E2AC
10	evaluation	S1B_EW_GRDM_1SDH_20180421T080246_20180421T080346_010579_0134BA_95A0
11	evaluation	S1A_EW_GRDM_1SDH_20181016T072958_20181016T073058_024158_02A460_DA8F
12	evaluation	S1B_EW_GRDM_1SDH_20190110T080052_20190110T080152_014429_01ADC3_DDBE
13	evaluation	S1A_EW_GRDM_1SDH_20190417T075536_20190417T075636_026827_0303EB_ADF6
14	evaluation	S1A_EW_GRDM_1SDH_20190717T074617_20190717T074717_028154_032E13_D688
15	evaluation	S1A_EW_GRDM_1SDH_20191019T080241_20191019T080341_029525_035C27_393
16	evaluation	S1A_EW_GRDM_1SDH_20220404T081913_20220404T082013_042621_0515B_D75F
17	evaluation	S1B_EW_GRDM_1SDH_20200601T072921_20200601T073021_021837_02972D_1C76
18	evaluation	S2B_MSIL1C_20220404T143749_N0400_R039_T29XMH_20220404T151309

structure selection, hyper-parameter settings, and performance evaluation metrics.

A. Experiment setup

We investigate three scenarios for different training data sets:

- (1) Scenario 1: Full training data
- (2) Scenario 2: 10% of the original training data set
- (3) Scenario 3: Expand the 10% of the original training data set by using our physics-based data augmentation

In scenario 1, we use the full set of training data, which is generated by manual labelling of the physics-aware Gaussian Mixture Model (GMM) ([10], [30]) clustering results. In scenario 2, we use only 10% of the full training data set to simulate a situation where just a small number of labelled training samples is available, e.g. an ice analyst has just delineated a few polygons. Finally, in scenario 3, we augment the training set from scenario 2 to investigate if we are able to achieve similar accuracy as in scenario 1.

Since both scenario 2 and scenario 3 use only 10% of the original data set, we are able to carry out ten independent experiments and provide an uncertainty estimation for the algorithm performance. More information about the different scenarios is given in Table II. Meanwhile, the detailed procedure for generating the training data for these three scenarios is described in the following section.

B. Training data generation

1) *Incidence angle dependence of SAR backscatter*: The radar backscatter from any surface is determined by both radar and surface parameters. The most important surface parameters are the dielectric properties of the material and the surface roughness, the main radar parameters include frequency, polarization, and local incidence angle (IA). While frequency and polarization are fixed for a given sensor and imaging mode (in our case Sentinel-1 EW mode) the local

IA varies across the swath. It is known that backscatter intensity decreases with increasing IA, and several studies have shown that this decrease is approximately linear for backscatter intensity in decibel (dB) (for example [28], [31], [32]). The linear slope for σ_{HH}^0 versus IA is generally larger than the slope for σ_{HV}^0 versus IA.

Furthermore, it has been shown that the variation of backscatter intensity with IA is dependent on the surface type [28], [31]. In particular, sea ice and open water can have significantly different slopes, on the the order of 0.2 dB/1° for sea ice and up to 0.7 dB/1° for open water [32]. Hence, a good IA correction must take the surface type into account. Lohse et al. [9] and Cristea et al. [10] have proposed classification and clustering methods for sea ice classification that include the per-class IA dependence.

2) *Training data generation stage 1 - GMM clustering*: Cristea et al. [10] proposed an automatic and unsupervised segmentation algorithm for SAR image clustering by integrating the IA dependence. Each cluster is described by a "Gaussian tube" in the intensity-IA space, as the mean value of the Gaussian distribution decreases linearly with IA. A GMM is sequentially applied, using the expectation-maximization (EM) algorithm, with an increasing number of the clusters. After each EM convergence, the goodness-of-fit (GoF) of the individual clusters is tested using Pearson's chi-squared test and the least well-fitting one is split. When all clusters are considered good-fits, or a maximum number of clusters is reached, the algorithm stops. The final hard cluster labels are assigned according to their associated maximum posterior probabilities. Several parameters can be tuned to affect the final segmentation result. In our work, we used 0.99 as the confidence level, maximum number of clusters is 15, sub-sampling to 9000 samples during the iterations, as this usually achieved around ten clusters which were considered simple enough to interpret. These unsupervised segmentation clusters are manually labelled as either "sea ice" or "open water" with the help of expert analysis, visual interpretation, and interpretation of the slope values per cluster.

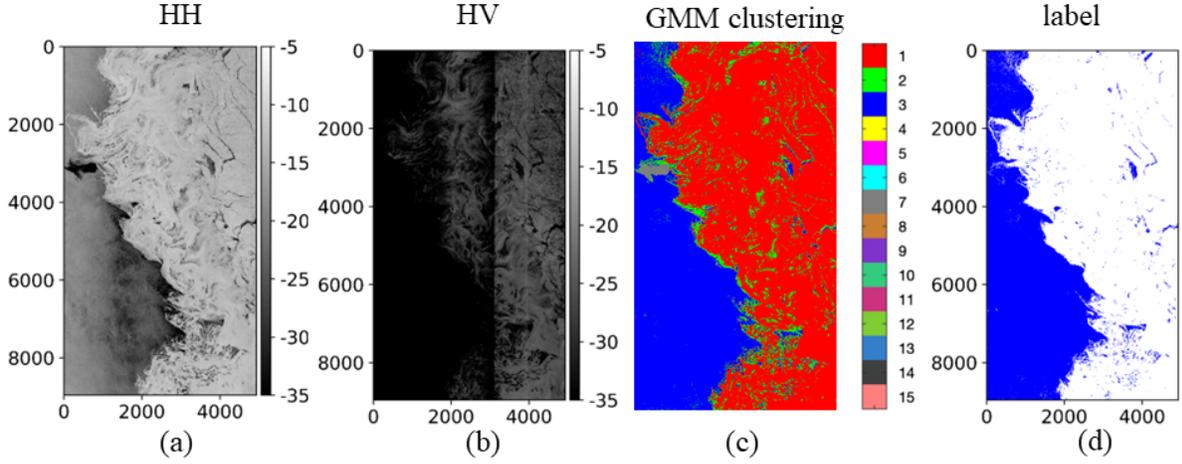


Fig. 4. S1 example image (scene NO.3 in Table I), showing HH (a) and HV (b) backscatter intensity (in dB), GMM segmentation result (c), and the corresponding classes (d) after manually labelling segments as sea ice (white) or open water (blue).

An illustrative example of this clustering and labelling process is given in Fig.4 and Fig.5. Fig.4 shows HH and HV backscatter intensity together with the corresponding GMM segmentation result and the manually assigned cluster labels for a S1 scene captured on January 4th, 2021. For this scene, 15 total clusters are obtained by the GMM, although only the main four are easily visible. In Fig.5, the HH intensity in dB is shown against against IA for three main clusters and the corresponding IA linear slopes (obtained by linear least-square fitting) are identified as well. Both the visual interpretation of the intensity images in Fig.4 and the interpretation of the IA linear slope steepness in Fig.5 lead us to the labelling of the green and the red classes as sea ice and the blue class as open water.

As we are targeting binary ice-water classification, separate ice clusters (see Fig.4 (c)) are merged into one single sea ice class. Similarly, different open water clusters (see Fig.4 (c)) are merged into one single open water class (Fig.4 (d)). The criterion for merging classes is mainly based on (1) the linear slope (shown in Fig.5) where ice has a shallow linear slope while water has a steep linear slope, and (2) further consulting with ice experts. Meanwhile, the IA linear slopes of all the clusters as found by the segmentation algorithm are recorded independently. For the following steps, the segmentation results are further sub-divided into patches of size of 224×224 pixels (roughly 9×9 km) in our study. The patch size of 224×224 pixels is selected as this patch size is used by most CNN-based models (for instance, AlexNet developed by Krizhevsky et al. [33], VGG16 proposed by Simonyan and Zisserma [34], Resnet developed by He et al. [35]).

3) *Training data generation stage 2 - IA linear slope-based data augmentation:* In the next step, the individual IA linear slopes captured in stage 1 of the training data generation are utilized to augment the training data set. For this purpose, we define a transformation function that adjusts the IA values of the original reference patch to any target IA range between 19° and 47° . Taking the original patch labels, clusters, and cluster-dependent IA linear slopes into

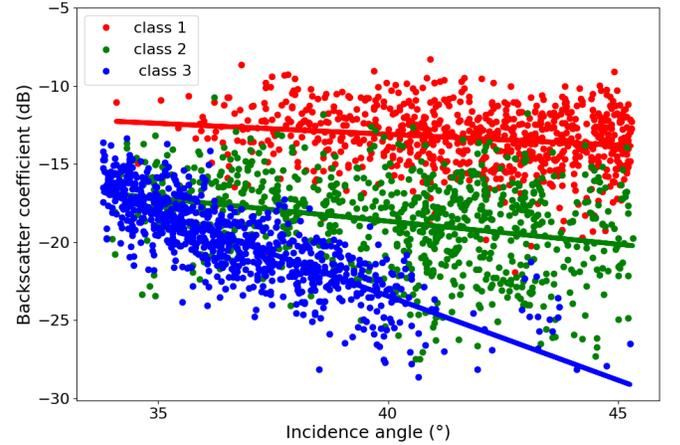


Fig. 5. Example of slopes of backscatter intensity with IA for different clusters. The clusters correspond to the GMM segmentation shown in Fig.4c (here, red and green indicate different ice classes, while blue is the water class)

account, each pixel's backscatter values are then projected along the corresponding slope to the new IA value. Since this transformation makes use of the physics-based principle of per-class backscatter variation with IA, we refer to it as physics-based data augmentation. The augmentation is defined by the following equations:

$$IA_{i,j}^{new} = IA_{i,j}^{ref} + m * int \quad (1)$$

$$HH_{i,j}^{new} = HH_{i,j}^{ref} + (IA_{i,j}^{new} - IA_{i,j}^{ref}) * S_{HH}^k \quad (2)$$

$$HV_{i,j}^{new} = HV_{i,j}^{ref} + (IA_{i,j}^{new} - IA_{i,j}^{ref}) * S_{HV}^k \quad (3)$$

where int is the predefined IA interval (here we choose 1°) and m is the multiplication factor for the shifting the IA

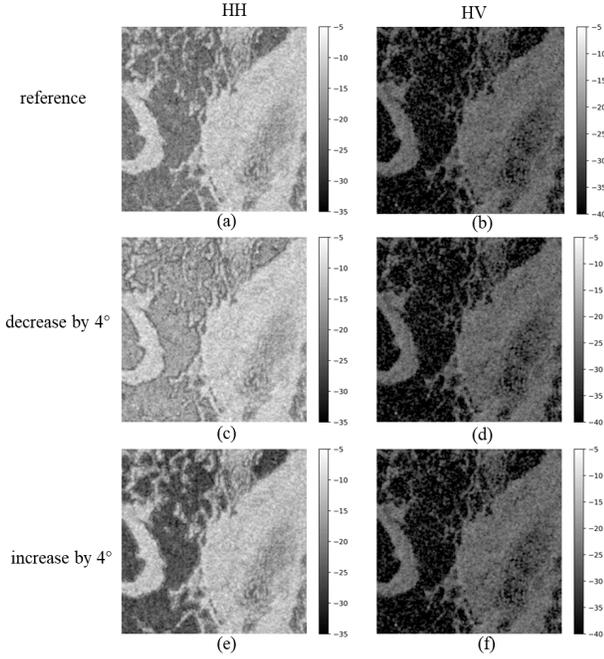


Fig. 6. Example of a 224×224 pixel patch before and after physics-based augmentation. The top row shows backscatter intensities and IA from the original image (stage 1 training generation). Rows 2 and 3 show backscatter intensities and IA after data augmentation (stage 2 training generation). Backscatter values are projected along the respective slopes of their corresponding clusters to IA values with 4 degrees below (row 2) and 4 degrees above (row 3) the reference IA.

values. $IA_{i,j}^{ref}$ and $IA_{i,j}^{new}$ represent the IA of the reference pixel and its corresponding pixel after the augmentation. Similarly, $HH_{i,j}^{ref}$ (and $HV_{i,j}^{ref}$) and $HH_{i,j}^{new}$ (and $HV_{i,j}^{new}$) represent the backscatter intensities of the reference pixel and its corresponding pixel before and after augmentation. S_{HH}^k and S_{HV}^k denote the IA linear slopes of the cluster k to which the pixel (i, j) belongs. In Fig.6, a sample for original reference patch and new patch after physics-based augmentation is illustrated.

We perform the proposed two-stage training data generation routine for eight S1 EW images acquired between January and March 2021. 500 boundary pixels in the upper, lower, left and right edges of the scenes have been removed before the procedure, as these pixels are more likely to be affected by noise or boundary effects. In the end, we are able to generate 5910 patches for stage 1, based on the aforementioned procedure of GMM clustering and expert analysis of the clusters. Following the 80% vs 20% rule for training and validation separation, we randomly select 4731 patches for training and 1179 for validation, as shown in Table II. The patches allocated in the training set are used to learn the hyper-parameters of the classifier, whereas the validation data is used to independently assess its performance.

This data set is being considered the complete data set and will be used for training in our scenario 1 experiment. Meanwhile, to simulate a situation with fewer available training patches, we split the 4731 training patches into 10 equal-sized partitions, such that each partition consists of 470 training patches. Taking each of these 10 sub-sets in combination

with the complete 1179 patches for validation, we generate 10 training and validation data set couples. These form the basis for our experiments in scenario 2.

Finally, we apply the physics-based data augmentation on the 10 training sets from scenario 2 to generate the training sets for scenario 3. Hereby, we are able to generate a large amount of training data and make sure that all example training classes are evenly distributed across the whole IA range of the S1 EW mode. We are able to generate approximately 11,000 training patches for each training sub-set by moving the patches at 1° steps within the IA range between 19° to 47° . The resulting 10 enriched training and validation patch couples are the basis for our scenario 3 training configuration. Detailed information for three scenarios are listed in Table II.

C. Model structure selection

Various models for semantic segmentation applications are being developed in the community, including FCN developed by Long et al. [36], U-Net [37], Deeplabv3 [38], and their variants including U-Net++ [39], deeplabv3+ [40], etc. Among these methods, FCN normally shows the worst performance compared with others, since it employs the most simple architecture with only two or three levels of up-sampling, which might not be able to capture enough details. Similarly, Deeplab V3 and Deeplabv3+ only implement one or two up-sampling levels, which is sufficient for images with large homogeneous areas. However, in the case of SAR imagery, pixel backscatter values often vary strongly. Hence, using U-Net or its variants appears to be most beneficial. As our starting point for selecting the model architecture, we therefore choose the U-Net, which has more encoder (down-sampling) and decoder (up-sampling) layers.

U-Net employs a symmetric encoder-decoder structure in which the encoder part is trying to extract the features in various levels and the decoder part is utilized to re-construct the segmentation masks based on integrating the features in different levels. The encoder part can be divided in to 4 similar blocks where each block consist of a sequence of two 3×3 convolutional layers followed by a batch normalization (BN) procedure and the rectified linear unit (ReLU) activation function is added in the end. Following these, max-pooling operations with 2×2 window are used for down-sampling to reduce the feature size. Correspondingly, in the decoder part, a bi-linear up-sampling operation is employed. In the last step, the kernel size is set as 1×1 to ensure it has the same size as original image. The original number of filters are 64, 128, 256, and 512 for the four levels. Using this architecture as a starting point, we further designed a simplified U-Net for our study by reducing the number of layers as well as the filters from the original U-Net, which leads to a three layers U-Net with the numbers of 32,128,256 filters in the three levels. The detailed architecture for our simplified U-Net architecture is shown in Fig.7. Note that we utilize three channels (HH, HV, IA) as input into the modified U-Net.

D. Model performance evaluation metric

Various metrics can be considered to evaluate the performance of separation of sea ice and open water. For instance,

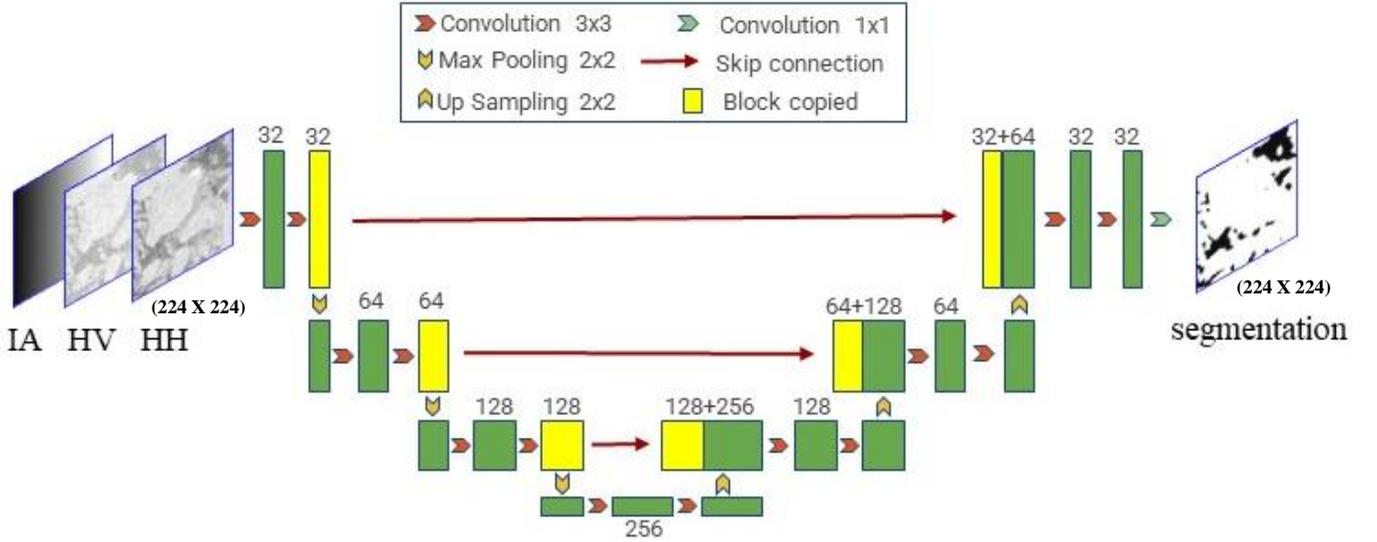


Fig. 7. Illustration of the modified U-Net architecture for sea ice/water segmentation used in this study.

Park et al. [12] utilized the confusion matrix for the results of their random forest classifier, while Stockholm et al. [15] adopted R^2 to assess the accuracy of their output SIC maps. In this manuscript, we will follow the commonly used error metrics in the field of semantic segmentation, which is mean Intersection over Union (MIoU). The computation of MIoU is given in the following equations, where k is the number of classes and p_{ij} is the element of the confusion matrix at row i and column j .

$$IoU_i = \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (4)$$

$$MIoU = \frac{\sum_{i=1}^k IoU_i}{k} \quad (5)$$

E. Hyper-parameter settings

For all the scenarios, the same configuration of the hyper-parameters is implemented for the training process of the simplified UNET. The detailed configuration for the training is: Learning rate is set to $1e-4$, maximum training epoch is set to 200, and the batch size is configured with 64 which is maximum size for the environment (NVIDIA Quadro RTX 5000 graphic processing unit (GPU)). Note that we utilize the focal loss [41] to compensate the class imbalance as well as the difference of difficulty to distinguish sea ice and open water classes. In particular, the parameters α and γ are set to 0.9 and 2, respectively. Using this configuration, the training procedure takes 181, 31 and 396 minutes for scenario 1, 2 and 3, respectively (shown in Table II). The complete processing and inference of a new S1 image takes approximately 10 minutes, most of which is used for feature extraction and cropping the of the image into 224x224 pixel patches. The

actual inference of these patches takes only approximately 20 seconds.

IV. RESULTS

A. General model performance

The performance results (based on the validation patches) of the three different scenarios are summarized in Table II. Recall that we have split the complete training set and carried out 10 experiments each for scenario 2 and scenario 3. Clearly, the highest accuracy is achieved in scenario 1, where we use the complete training set. Furthermore, we find that the physics-based data augmentation of training data improves classification accuracy. This is indicated by the MIoU improved from 91.72% - 92.71% to 92.48% - 92.96%. Meanwhile, the standard deviation for the scenario 2 results is 0.24% for MIoU, while the standard deviation of MIoU for scenario 3 is 0.16%. The lower variance for scenario 3 accuracy compared to scenario 2 accuracy indicates that the model performance is more robust and stable with scenario 3, owing to the benefit of physics-based data augmentation. The overall higher score achieved in scenario 1 can be attributed to the fact that the reduced training data of both scenario 2 and scenario 3 covers fewer and less diverse sea ice situations than the full training of scenario 1. Hence, while physics-based data augmentation improves the results, it cannot fully substitute the manual labelling of additional training data.

To illustrate changes during the training process of the model, we plot the evolution of error metrics of the training for scenario 1 together with experiment 5 of both scenario 2 and scenario 3 in Fig.8. The upper panel (Fig.8 (a)) shows the change of MIoU with training epochs, the lower panel (Fig.8 (b)) shows the loss evolution with training epochs. It can be seen that when using physics-based augmentation, the loss for the training and validation process drops faster compared to the training without data augmentation. Furthermore, the training converges to 0.018 when we include physics-based

TABLE II
SETTINGS AND MIOU (CALCULATED BASED ON THE VALIDATION
PATCHES) FOR THE DIFFERENT SCENARIOS

scenario	time (")	training patches	MIOU
scenario 1	10847	4731	94.29%
scenario 2	1	1879	92.30%
	2	1868	91.72%
	3	1853	92.12%
	4	1855	92.27%
	5	1833	92.18%
	6	1810	92.29%
	7	1809	92.28%
	8	1821	92.71%
	9	1828	92.32%
	10	1830	92.28%
mean			92.25 ± 0.24%
scenario 3	1	23695	92.62%
	2	23898	92.69%
	3	23875	92.82%
	4	23939	92.79%
	5	23562	92.96%
	6	24155	92.48%
	7	23935	92.79%
	8	23929	92.84%
	9	23706	92.49%
	10	22668	92.83%
mean			92.73 ± 0.16%

data augmentation (experiment 5 of scenario 3), which is close to the value of 0.014 when using the complete data set (scenario 1) and better than the convergence of experiment 5 from scenario 2, which converges to 0.021. This shows that we are able to achieve a better performance during training by including physics-based data augmentation. Lastly, the training process of scenario 3 is stabilized at the epoch of 130-140, which is same as using the complete data set in scenario 1, but much earlier than for the training without data augmentation in scenario 2, which stabilizes at the epoch 170-180.

B. Comparison with CAS product

To further evaluate the performance of our model both with and without physics-based data augmentation, we compare our model inference with other existing products. Several studies, for example Andersson et al. [42] or Stokholm et al. [15], generate automated ice charts which provide SIC at a rather coarse resolution (in the order of km). These products are difficult to use for comparison with our model inference since the spatial resolution of our inference is 40 m. We therefore use the sea ice mapping produced by Wang and Li from Chinese academic science (CAS) for the comparison in this study. The CAS data set is based on an ensemble prediction of multiple U-Net models [17]. Its spatial resolution is 400 x 400 m, resulting from a 10 x 10 pixel averaging window that was applied during processing. In contrast, our algorithm provides output labels at the original S1 EW pixel level of 40 x 40 m. Meanwhile, the CAS data set is based on training from 167 S1 scenes distributed over all seasons, while we use only 8 S1 scenes acquired during winter conditions.

A visual comparison of the inference results for 4 different S1 images (scene NO.12-15, Table I) is shown in Fig.9. Note that we compare to the results from our model in scenario 1

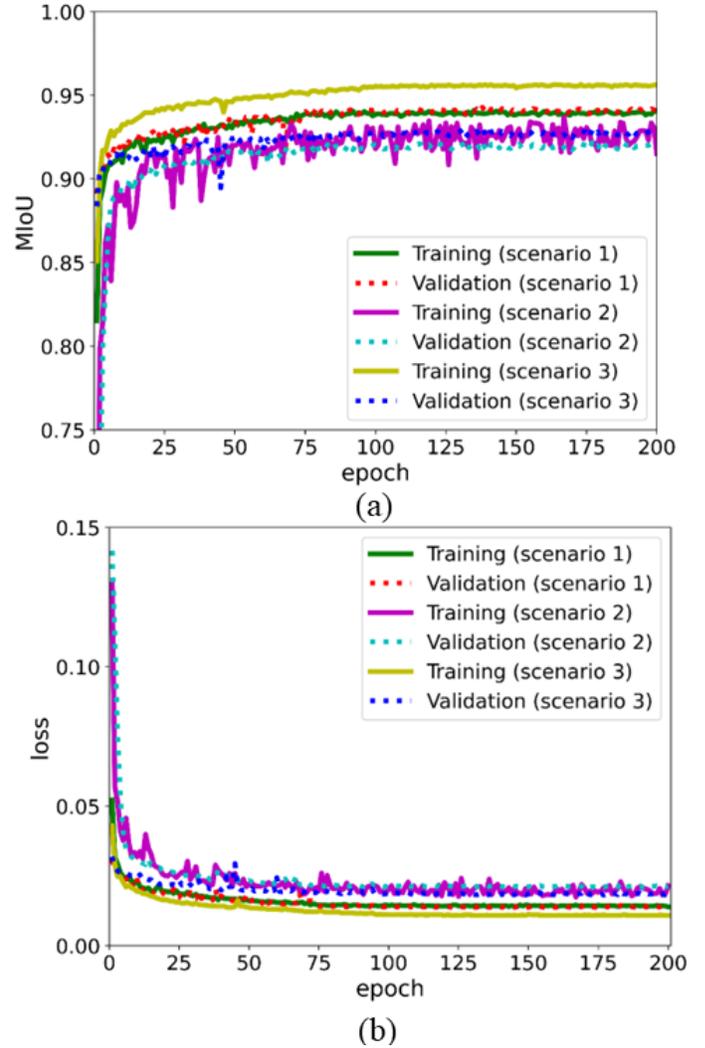


Fig. 8. Evolution of MIOU (a) and loss (b) for training and validation of different scenarios (scenario 1, experiment 5 of scenario 2 and scenario 3)

(full data set), as well as one selected sub-set from scenario 2 and scenario 3 (experiment 5). Overall, our model inference matches well with the CAS classification. However, several differences in the inference can be found, some of which are highlighted by colored rectangles in Fig.9.

For example, in all 4 examples shown here, artifacts from the patch edges (224x224 pixels) are visible in the inference from scenario 2 (red rectangles in Fig.9 (d), (j), (p), (v)). At these edges, open water is mis-classified as sea ice. The issue is resolved in the inference from scenario 3, where physical data augmentation is applied.

Furthermore, there is considerable mis-classification of open water as sea ice at the boundary between the EW1 and EW2 swaths of the S1 images, highlighted by the purple rectangles in Fig.9(d),(e),(j),(k). This is a commonly known issue in ice-water classification caused by the noise properties of S1 imagery [9] and has already been recognized as problematic by different researchers, who suggested several methods to address this issue ([16], [43], [44]). As can be seen in Fig.9(c) and (i), the CAS algorithm is able to classify

correctly around the swath boundary, while both our models from scenario 2 (Fig.9(d), (j)) and scenario 3 (Fig.9(e), (k)) output mis-classified pixels. Hereby, we explore the potential of adding more training samples to overcome these noise issues. Instead of adding new training patches by physics-based data augmentation, we can also re-train our model using the complete data set (scenario 1), shown in the last column of Fig.9. It can clearly be seen that using the complete training set that contains more diverse sea ice situations, the noisy pixels in transitional area can be removed significantly (see purple rectangle in Fig.9(f), (l)). It appears that in this case the physics-based data augmentation cannot improve the inference as much as additional, completely new and independent training labels. Meanwhile, the remaining noisy pixels across the swath boundary can be easily removed by applying a spatial filter. Generally, better noise estimation as well as de-noising methods by using physical models or deep learning models could be considered in the future. It can also be seen in Fig.9(f), that using the complete training set (scenario 1) enables us to delineate the ice edge even better (see purple rectangle). The resulting ice edge matches better with the edge from the CAS data set (Fig.9(c)).

Thus, we conclude that a diverse training set is essential to resolve noise-related classification errors for the separation of sea ice and open water. Furthermore, post-processing techniques to clean the false alarms after inference are recommended.

Further differences between the CAS data set and our model inference are highlighted by the green and blue rectangles in the 1st, 2nd, and 4th row of Fig.9. For example, the green rectangle in 2nd row of Fig.9, highlights an area close to the ice edge, which is identified as open water by the Fig.9(i), while both scenario 2 (Fig.9(j)) and scenario 3 (Fig.9(k)) inference find more sea ice. Without additional information, such as optical data or in-situ ground truth observations, it is not straightforward to conclude which algorithm is more accurate. The area is characterized by lower backscatter values than the surrounding sea ice, which can be an indication of either relatively calm open water areas or comparatively smooth newly-formed sea ice. The discrepancy between the CAS labels and our inference in this case may be emphasized by the difference in spatial resolution. Similar considerations apply for the area marked by the green and the blue rectangles in the 4th row of Fig.9, where our model inference shows slightly more open water compared to the CAS labels. In green rectangle in particular, the CAS labels (Fig.9(u)) shows a small portion of water with a sharp vertical line, whereas our inference from both scenario 2 and scenario 3 shows a smooth ice-water boundary.

Lastly, areas of landfast ice close to the East Greenland coast (Fig.9(c)-(f), blue and green rectangles) prove to be challenging for all presented algorithms. At the time of the image acquisition (January 2019), the highlighted area is entirely covered by landfast sea ice. However, when the landfast ice forms thermodynamically in-situ, it will produce a very smooth surface that results in low backscatter intensity of the SAR signal. This can be mis-interpreted as calm open water. This issue is demonstrated in more detail in Fig.10 and the

following section.

Overall, our model inference matches well with the CAS labels. Noise issues and transitions at swath boundaries, as well as large areas of young ice and very smooth landfast ice, can cause mis-classification. Without additional information or complementary data, it is not always possible to assess which model is more accurate. The CAS model performs better in noisy open water areas, but it requires more computational resources (recall that it is based on ensemble prediction of multiple UNET models) and offers coarser spatial resolution. Some network-related artifacts (patch boundaries) from our model can be mitigated by physics-based data augmentation.

C. Comparison with optical imagery

We also compare the model inference with optical imagery from Sentinel-2. During cloud-free conditions, optical data allow for straightforward separation of sea ice and open water based on image brightness. However, in the marginal ice zone, sea ice drift and differences in image acquisition time make it difficult to use optical images for pixel-wise validation of our model inference. In Fig.10, we therefore show an area along the edge of the landfast ice close to East Greenland. The Sentinel-1 SAR image was acquired on April 4th 2022 at 08:19:13 UTC, the optical Sentinel-2 image at 14:37:49 UTC in the same day.

For most of the image, the model inference is correct. Some errors occur, indicated by the red rectangles in Fig.10, where very smooth and level landfast ice is mis-classified as open water. This classification error was already observed in the previous section. While much of the landfast ice in the region is formed in other areas and gets locked in place as highly deformed rubble fields, some of the ice forms in-situ under very protected conditions. This results in unusually smooth and undeformed sea ice, which is classified incorrectly by our model. The green rectangles in Fig.10 mark polynya regions, in which new sea ice is formed. As the overall ice drift is towards the south, the northern parts of the polynyas is ice free, and ice thickness gradually increase in southern direction within the polynya. The model inference captures the transition between open water and young ice in the polynya very well.

Additional comparisons with more optical data would be desirable, in particular in the marginal ice zone. However, sea ice conditions close to the ice edge are highly dynamic and the structural patterns of the ice change so quickly that ice drift compensation in this area is not feasible. In this section, we therefore restrict ourselves to the one example discussed above. For future work, a better approach for more and easier comparison of classification results and optical data close to the ice edge would be beneficial.

V. DISCUSSION

A. Seasonal variation for sea ice and water segmentation

Apart from some remaining ambiguities (young ice and very smooth and level landfast ice), the results of this study show that our model is able to correctly classify open water and sea ice during the winter season for both calm and windy water conditions (Fig.9 (g)-(r)). However, as we only include training

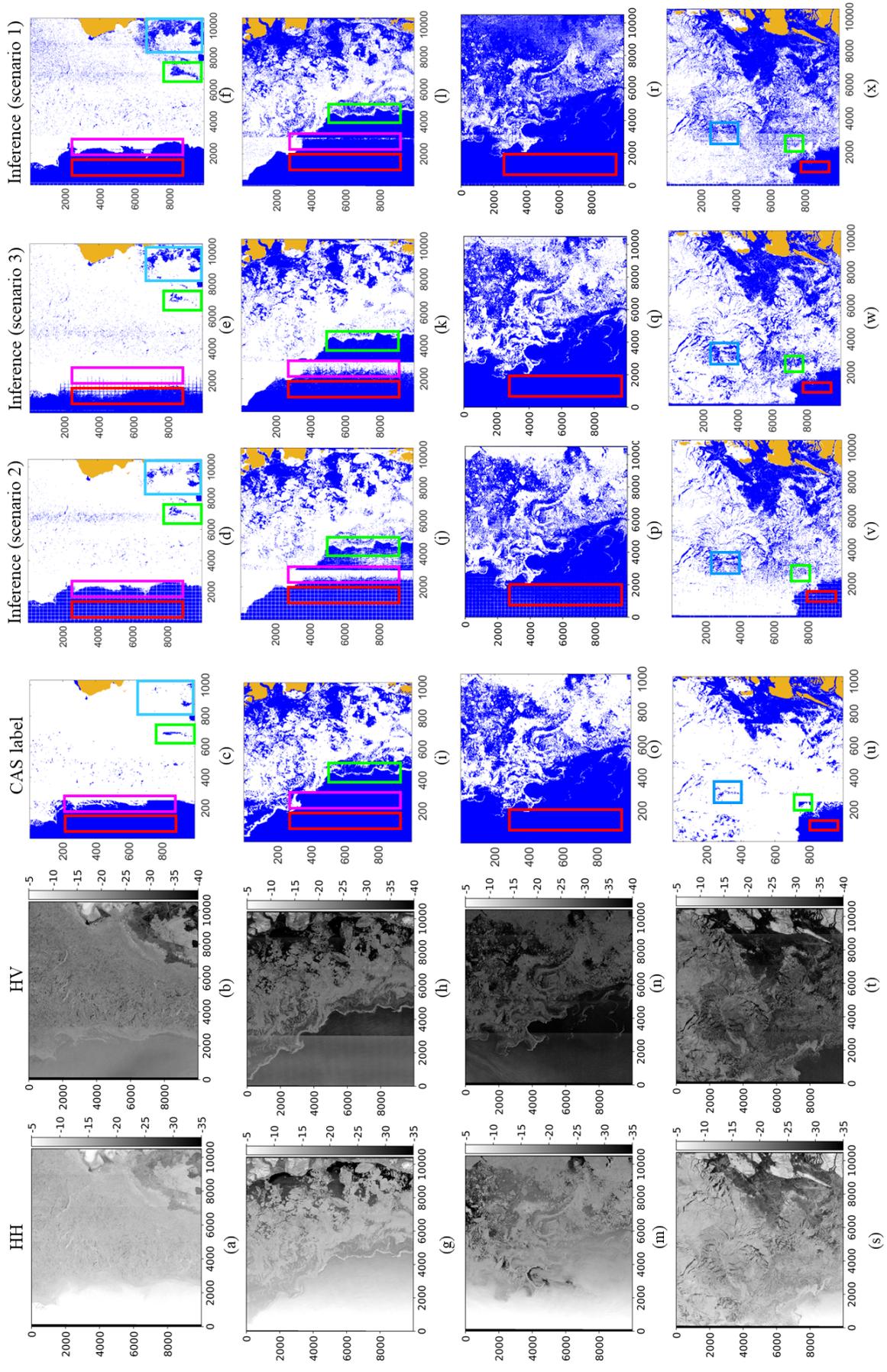


Fig. 9. From left to right: HH, HV, CAS labels, model inference (scenario 2), model inference (scenario 3), and model inference (scenario 1) for four S1 example images (scene NO. 12-15, Table I). Note the coarser spatial resolution of the CAS data set (400m compared to 40m). Regions of particular interest are highlighted by colored rectangles (artifacts from the patch edges are identified by the red rectangle, purple box indicates the transition are between swath EW1 and EW2, green and blue boxes indicate other areas which are being identified differently by CAS label and our inference) and referred to in the text. Water is blue, ice is white, and land is masked yellow.

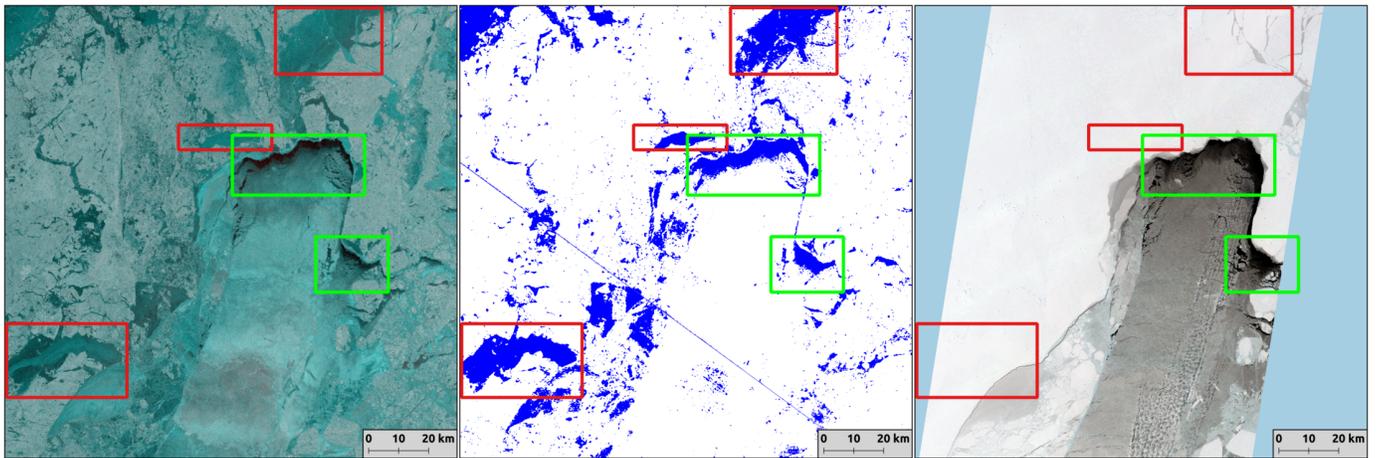


Fig. 10. Comparison of Sentinel-1 (left, false-color composite of R:HV, G:HH, B:HH), inference (middle) (scene NO.16 in Table I), and Sentinel-2 visual channels (right) (scene NO.18 in Table I). Red rectangles indicate areas where very smooth and level landfast ice close to East Greenland is mis-classified as open water. Green triangles indicate areas where the algorithm successfully identifies polynya areas that are opening up and partly covered with young ice.

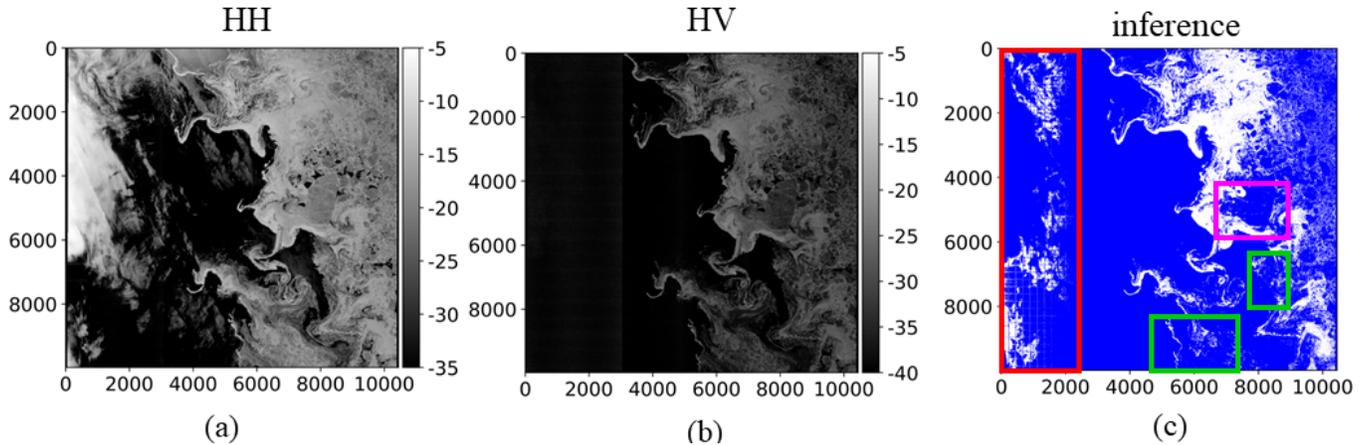


Fig. 11. HH (dB) and HV (dB) intensity together with the inference result for a selected example image that was acquired in the summer (scene NO.17 in Table I). The colored rectangles in (c) indicate areas with incorrect inference and referred in the text. Such errors are expected, as the current version of the model is trained on winter scenes only.

data from the winter (January to March), the inference for the melt and summer season is expected to be more challenging. An example image acquired on June 1st 2020 (scene NO.17 in Table I) is shown in Fig.11, with wrongly classified areas highlighted by colored rectangles.

For this example image, the surface properties of the ice and snow have been altered by warm temperatures, resulting in mis-classification of sea ice areas as open water (green and purple rectangles in Fig.11). Additionally, there is considerable mis-classification of wind-affected water as sea ice (red rectangle in Fig.11). Overall, the ice edge is not well delineated. The most straightforward way to obtain a reasonable inference for images acquired during melting conditions is to include similar imagery captured in the summer in the training data set and retrain the model carefully.

B. Young ice and smooth landfast ice areas

As clearly indicated in Fig.10, areas of smooth and level landfast ice are challenging for automated separation of sea ice and open water. Although it achieves slightly better results,

this is also true for the CAS algorithm (Fig.9). While further investigation of this issue is beyond the scope of the present study, there are several approaches that could help to improve the classification of these regions. The high-resolution SAR observations can be merged with low-resolution PM observations to a multi-modal data set. PMR is less sensitive to surface roughness and would help to correctly identify landfast ice regions as sea ice. However, merging these different data sources comes with the challenge of their different spatial resolution, as well as differences in acquisition time, which will be problematic in drift ice areas. Alternatively, the SAR-based classification could potentially be combined with ice drift algorithms, which are capable to identify landfast ice areas. Further research in this area is required in future studies.

C. Effect of the training size on physics-based augmentation

A main focus of the present study was the investigation of our proposed physics-based data augmentation for classification of sea ice and open water. The results presented in Table II show that the data augmentation leads to a significant

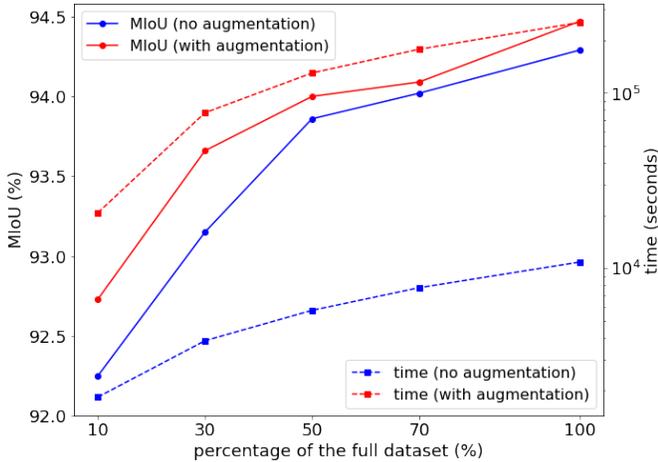


Fig. 12. MIoU and processing time for training with different fractions of the full data set. (Note that the MIoU is plotted against first y axis and processing time (logarithmic scale) is plotted against second y axis.)

TABLE III
SETTINGS AND MIOU (CALCULATED BASED ON THE VALIDATION PATCHES) FOR TRAINING WITH DIFFERENT FRACTIONS OF THE FULL DATA SET, WITH AND WITHOUT DATA AUGMENTATION

scenario	fraction (%)	time (")	training patches	MIOU
no augmentation	10	1839	473	92.25%
	30	3855	1458	93.15%
	50	5749	2309	93.86%
	70	7740	3303	94.02%
	100	10847	4731	94.29%
with augmentation	10	20739	10889	92.73%
	30	77231	36315	93.66%
	50	130212	60525	94.00%
	70	177877	84736	94.09%
	100	253255	121051	94.47%

improvement in classification accuracy, when a small subset of 10% of the full training set is used. However, it remains unclear whether the augmentation is also beneficial when the original data set is already larger. To investigate this question, we performed our proposed data augmentation repeatedly on differently sized training sets, corresponding to subsets of 30%, 50%, 70%, and 100% of the full training data set (4731 patches). The resulting performances are shown in Fig.12 and summarized in Table III.

Fig.12 shows that the performance with data augmentation is always better than without data augmentation. However, the MIoU gain achieved by the data augmentation is larger when only a small fraction of the full data set is used. For a larger fraction of the full data set the performance gain is smaller. More specifically, the gain in MIoU of training with augmentation using 10% of the full data set is 0.48% (from 92.25% to 92.73%), while the gain of augmentation with larger training sets, i.e. 50%, 70%, and 100% of the available training data, will result in a gain of 0.14%, 0.07% and 0.18%, respectively. We also note that when training with the full data set (100%, i.e. all 4731 patches), the model achieves an MIoU of 94.29%. The training procedure for

this case takes approximately 3 hours. However, to achieve a similar performance with data augmentation, we need to use already 70% of the full training set. Hence, we need a very large number of training patches, consequently resulting in a long training process, taking more than 49 hours. Note the training time (in seconds) on the logarithmic y-axis in Fig.12. Considering the large amount of computer resources required for such long training processes, we conclude that it is preferable to manually label additional training data, whenever this is possible. However, as the manual labeling is time-consuming for an expert, this is not always a viable option. In such cases, our proposed physics-based data augmentation provides an alternative way to improve the algorithm performance.

VI. CONCLUSION

In this study, we have introduced a novel, physics-based data augmentation routine to increase the amount of training data for ice-water separation in SAR imagery. Our proposed method utilizes the per-class dependency of backscatter intensity (in dB) with IA to shift training patches with known class labels to different IA ranges. To test this method, we have implemented a modified U-Net model to classify sea ice and open water in S1 images. We have trained different versions of the model, first using our full training data set and then using subsets of the full data set, both with and without data augmentation. Overall, we find that our proposed method for physics-based data augmentation increases the accuracy of the classification. Using a subset training size of 10% of the full training set, we find that MIoU increases by 0.48% (92.73% compared with 92.25%). We also find that including the new data augmentation results in earlier convergence and stabilization during the training phase of the model (at epoch 130-140 compared to epoch 170-180). Further experiments with larger subsets of the full training set show that the effect of the data augmentation is largest when the amount of original training data is small, and decrease with an increasing amount of labeled samples. Finally, it should be noted that the data augmentation leads to a highly increased training time of the model. Meanwhile, the best accuracy overall is still achieved when using the full training set, as it covers most diverse sea ice situations. Hence, adding additional manual training patches is preferable to data augmentation whenever possible. When the training set is limited or more manual labeling is not possible though, our proposed method can help to increase classification performance.

We tested the overall performance of our model by comparing the inference for several S1 images with corresponding ice-water labels from the CAS sea ice data set. The comparison shows that our model produces similar results to the CAS data set, indicating that it can provide detailed separation of sea ice and open water at fine spatial resolution (original 40x40 m pixel spacing of the S1 EW GRDM format). While the overall delineation of the sea ice edge corresponds well with the CAS labels, our training process is much less computationally expensive. We also find several limitations of our current

model, the most important ones being the mis-classification of open water as sea ice at the boundary between the S1 sub-swaths EW1 and EW2, and the mis-classification of very smooth and level landfast ice with low backscatter values as open water. Thus, more effort is required to handle these situations. Suggestions for future improvement of the model include adding more training data for ice situations that are known to be particularly difficult to classify and applying post-processing filters to smooth results in parts of the image that are affected by noise. Alternatively, different input features (for example texture features or intensity ratios) can be integrated into the model for better separating ambiguous classes. In addition, transformer-based network architectures, which can exploit long-range information effectively with the self-attention mechanism, should be further considered. These could be combined with pure CNN based networks, which are more robust in understanding short-range information due to the intrinsic locality of the convolution operations. For instance, TransUNet [45] and Swin-UNet [46] will be implemented and tested in the future.

ACKNOWLEDGMENTS

This work is funded in part by Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) and the Research Council of Norway (RCN Grant no. 237906), the European Union's Horizon 2020 research and innovation programme ExtremeEarth project, grant agreement no. 825258. We thank the anonymous reviewers for their insightful comments and suggestions to improve the paper.

REFERENCES

- [1] M. C. Serreze and W. N. Meier, "The arctic's sea ice cover: trends, variability, predictability, and comparisons to the antarctic," *Annals of the New York Academy of Sciences*, vol. 1436, no. 1, pp. 36–53, 2019.
- [2] B.-M. Kim, S.-W. Son, S.-K. Min, J.-H. Jeong, S.-J. Kim, X. Zhang, T. Shim, and J.-H. Yoon, "Weakening of the stratospheric polar vortex by arctic sea-ice loss," *Nature communications*, vol. 5, no. 1, pp. 1–8, 2014.
- [3] M. Kretschmer, D. Coumou, J. F. Donges, and J. Runge, "Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation," *Journal of climate*, vol. 29, no. 11, pp. 4069–4081, 2016.
- [4] A. M. Pagano and T. M. Williams, "Physiological consequences of arctic sea ice loss on large marine carnivores: unique responses by polar bears and narwhals," *Journal of Experimental Biology*, vol. 224, no. Suppl 1, p. jeb228049, 2021.
- [5] F. Lasserre and S. Pelletier, "Polar super seaways? Maritime transport in the Arctic: an analysis of shipowners' intentions," *Journal of Transport Geography*, vol. 19, no. 6, pp. 1465–1473, 2011, Special section on Alternative Travel futures.
- [6] G. Spreen, L. Kaleschke, and G. Heygster, "Sea ice remote sensing using amsr-e 89-ghz channels," *Journal of Geophysical Research: Oceans*, vol. 113, no. C2, 2008.
- [7] J. Comiso and C. Sullivan, "Satellite microwave and in situ observations of the weddell sea ice cover and its marginal ice zone," *Journal of geophysical research: Oceans*, vol. 91, no. C8, pp. 9663–9681, 1986.
- [8] R. Saldo, M. Kreiner, J. Buus-Hinkler, L. Pedersen, D. Malmgren-Hansen, and A. Nielsen, "Ai4arctic/asip sea ice dataset—version 2," 2020.
- [9] J. Lohse, A. P. Doulgeris, and W. Dierking, "Incident angle dependence of sentinel-1 texture features for sea ice classification," *Remote Sensing*, vol. 13, no. 4, p. 552, 2021.
- [10] A. Cristea, J. Van Houtte, and A. P. Doulgeris, "Integrating incidence angle dependencies into the clustering-based segmentation of sar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2925–2939, 2020.
- [11] N. Zakhvatkina, A. Korosov, S. Muckenhuber, S. Sandven, and M. Babiker, "Operational algorithm for ice–water classification on dual polarized radarsat-2 images," *The Cryosphere*, vol. 11, no. 1, pp. 33–46, 2017.
- [12] J.-W. Park, A. A. Korosov, M. Babiker, J.-S. Won, M. W. Hansen, and H.-C. Kim, "Classification of sea ice types in sentinel-1 synthetic aperture radar images," *The Cryosphere*, vol. 14, no. 8, pp. 2629–2645, 2020.
- [13] S. Khaleghian, H. Ullah, T. Kræmer, N. Hughes, T. Eltoft, and A. Marinoni, "Sea ice classification of sar imagery based on convolution neural networks," *Remote Sensing*, vol. 13, no. 9, p. 1734, 2021.
- [14] D. Malmgren-Hansen, L. T. Pedersen, A. A. Nielsen, M. B. Kreiner, R. Saldo, H. Skriver, J. Lavelle, J. Buus-Hinkler, and K. H. Krane, "A convolutional neural network architecture for sentinel-1 and amsr2 data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1890–1902, 2020.
- [15] 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," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.
- [16] J.-W. Park, J.-S. Won, A. A. Korosov, M. Babiker, and N. Miranda, "Textural noise correction for sentinel-1 topsar cross-polarization channel images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 4040–4049, 2019.
- [17] Y.-R. Wang and X.-M. Li, "Arctic sea ice cover data from spaceborne synthetic aperture radar by deep learning," *Earth System Science Data*, vol. 13, no. 6, pp. 2723–2742, 2021.
- [18] M. Olson, A. Wyner, and R. Berk, "Modern neural networks generalize on small data sets," *Advances in Neural Information Processing Systems*, vol. 31, 2018.
- [19] B. Rok and L. Lusa, "Smote for high-dimensional class-imbalanced data," *BMC Bioinformatics*, vol. 14, no. 1, pp. 106–121, 2013.
- [20] R. Hasibi, M. Shokri, and M. Dehghan, "Augmentation scheme for dealing with imbalanced network traffic classification using deep learning," arXiv preprint arXiv:1901.00204, 2019.
- [21] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," arXiv preprint arXiv:1710.09412, 2017.
- [22] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, 2016.
- [23] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," arXiv preprint arXiv:1711.04340, 2017.
- [24] G. Mariani, F. Scheidegger, R. Istrate, C. Bekas, and C. Malossi, "Bagan: Data augmentation with balancing gan," arXiv preprint arXiv:1803.09655, 2018.
- [25] A. O. Omigbodun, F. Noo, M. McNitt-Gray, W. Hsu, and S. S. Hsieh, "The effects of physics-based data augmentation on the generalizability of deep neural networks: Demonstration on nodule false-positive reduction," *Medical physics*, vol. 46, no. 10, pp. 4563–4574, 2019.
- [26] N. Dahiya, S. R. Alam, P. Zhang, S.-Y. Zhang, T. Li, A. Yezzi, and S. Nadeem, "Multitask 3d cbct-to-ct translation and organs-at-risk segmentation using physics-based data augmentation," *Medical physics*, vol. 48, no. 9, pp. 5130–5141, 2021.
- [27] Z. Fabian, R. Heckel, and M. Soltanolkotabi, "Data augmentation for deep learning based accelerated mri reconstruction with limited data," *International Conference on Machine Learning. PMLR*, 2021, pp. 3057–3067.
- [28] M. Makynen and J. Karvonen, "Incidence angle dependence of first-year sea ice backscattering coefficient in sentinel-1 sar imagery over the kara sea," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 11, pp. 6170–6181, 2017.
- [29] Y. Sun and X.-M. Li, "Denoising sentinel-1 extra-wide mode cross-polarization images over sea ice," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2116–2131, 2021.
- [30] Q. Wang, A. P. Doulgeris, and T. Eltoft, "Physics-aware training data to improve machine learning for sea ice classification from sentinel-1 sar scenes," in IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2022, pp. 4992–4995.
- [31] M. Makynen, A. T. Manninen, M. Simila, J. A. Karvonen, and M. T. Hallikainen, "Incidence angle dependence of the statistical properties of c-band hh-polarization backscattering signatures of the baltic sea ice," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 12, pp. 2593–2605, 2002.
- [32] J. Lohse, A. P. Doulgeris, and W. Dierking, "Mapping sea-ice types from sentinel-1 considering the surface-type dependent effect of incidence angle," *Annals of Glaciology*, vol. 61, no. 83, pp. 260–270, 2020.
- [33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

- [35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [36] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [37] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [38] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv preprint arXiv:1706.05587, 2017.
- [39] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2018, pp. 3–11.
- [40] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoderdecoder with atrous separable convolution for semantic image segmentation," *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
- [41] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [42] T. R. Andersson, J. S. Hosking, M. Perez-Ortiz, B. Paige, A. Elliott, C. Russell, S. Law, D. C. Jones, J. Wilkinson, T. Phillips et al., "Seasonal arctic sea ice forecasting with probabilistic deep learning," *Nature communications*, vol. 12, no. 1, p. 5124, 2021.
- [43] P. Q. Lee, L. Xu, and D. A. Clausi, "Sentinel-1 additive noise removal from cross-polarization extra-wide topsar with dynamic least-squares," *Remote Sensing of Environment*, vol. 248, p. 111982, 2020.
- [44] A. Korosov, D. Demchev, N. Miranda, N. Franceschi, and J.-W. Park, "Thermal denoising of cross-polarized sentinel-1 data in interferometric and extra wide swath modes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.
- [45] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.
- [46] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," arXiv preprint arXiv:2105.05537, 2021.



ter retrieval as well as computer vision tasks.

Qiang Wang received the B.S. degree in geodesy from Changsha university of science and technology in 2008 and PhD. Degree in geoscience and remote sensing from the University of Twente in 2018. Currently, he is a postdoc with the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA), Department of Physics and Technology, UiT—The Arctic University of Norway. His research interests include machine learning and deep learning algorithm for sea ice classification, embedding physics into neural network for parameter



classification.

Johannes P. Lohse received the B.Sc. and M.Sc. degree in geophysics from the University in Hamburg in 2009 and 2012, respectively. After overwintering in Antarctica in 2014, he joined the Department of Physics and Technology, UiT The Arctic University of Norway, in 2016 and received his PhD degree in 2021. Currently, he is a postdoc with the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) at UiT. His research interests focus on sea ice physics and remote sensing, in particular radar remote sensing and sea ice type



developing generic algorithms for remote sensing, pattern recognition, and multidimensional statistical modeling, in particular for polarimetric synthetic aperture radar (SAR) applications in sea ice and glaciers.

Anthony P. Doulgeris (Senior Member, IEEE) received the B.Sc. degree in physics from The Australian National University, Canberra, ACT, Australia, in 1988, the M.Sc. degree and the Ph.D. degree in physics from the Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway, in 2006 and 2011, respectively. He joined the Department of Physics and Technology, UiT The Arctic University of Norway, in 2007, and is currently a professor and head of the Earth observation group. His research interests focus on



include multidimensional signal and image analysis, statistical modeling, neural networks, and machine learning, with applications in multichannel synthetic aperture radar and ocean color remote sensing. Prof. Eltoft was a recipient of the year 2000 Outstanding Paper Award in Neural Networks awarded by IEEE Neural Networks Council, the Honorable Mention for the 2003 Pattern Recognition Journal Best Paper Award, and the 2017 UiT Award for Research and Development. He served as an Associate Editor for Pattern Recognition (Elsevier) from 2005 to 2011. He was a Guest Editor of Remote Sensing on the Special Issue for the PolInSAR 2017 Conference.

Torbjørn Eltoft (Member, IEEE) joined the Faculty of Science and Technology, UiT—The Arctic University of Norway, Tromsø, Norway, in 1988, where he is employed as a Professor in remote sensing with the Department of Physics and Technology. He is also the Director of the Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA), a center for research-based awarded by the Norwegian Research Council in 2014. He has a significant publication record in the area of signal processing and remote sensing. His research interests