Robust Multiseasonal Ice Classification From High-Resolution X-Band SAR

Karl Kortum , *Graduate Student Member, IEEE*, Suman Singha , *Member, IEEE*,
and Gunnar Spreen , *Member, IEEE*

Abstract—Automated solutions for sea ice-type classification from synthetic aperture radar (SAR) imagery offer an opportunity to monitor sea ice, unimpeded by cloud cover or the arctic night. However, there is a common struggle to obtain accurate classifications year round, particularly in the melt and freeze-up seasons. During these seasons, the radar backscatter signal is affected by wet snow cover, obscuring information about underlying ice types. By using additional spatiotemporal contextual data and a combination of convolutional neural networks and a dense conditional random field, we can mitigate these problems and obtain a single classifier that is able to classify accurately at 3.5-m spatial resolution for five different classes of sea ice surface from October to May. During the near year-long drift of the Multidisciplinary Drifting Observatory for the Study of the Arctic Climate (MOSAiC) expedition, we collected satellite scenes of the same patch of Arctic pack ice with X-band SAR with a revisit time of less than a day on average. Combined with *in situ* observations of the local ice properties, this offers up the unprecedented opportunity to perform a detailed and quantitative assessment of the robustness of our classifier for level, deformed, and heavily deformed ice. For these three classes, we can perform accurate classification with a probability $>95\%$ and calculate a lower bound for the robustness between 85% and 88%.

Index Terms—Machine learning, robustness, sea ice, synthetic aperture radar (SAR).

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) enables the monitoring of sea ice, unimpeded by cloud cover, weather effects, or the absence of sunlight. To this day, operational ice charting from SAR scenes is still largely carried out manually. This places a restriction on the resolution and frequency of updates. A solution to finding suitable automatic counterparts has obvious advantages, in both time investment and detail of classification. It is not feasible for a human to segment pixel
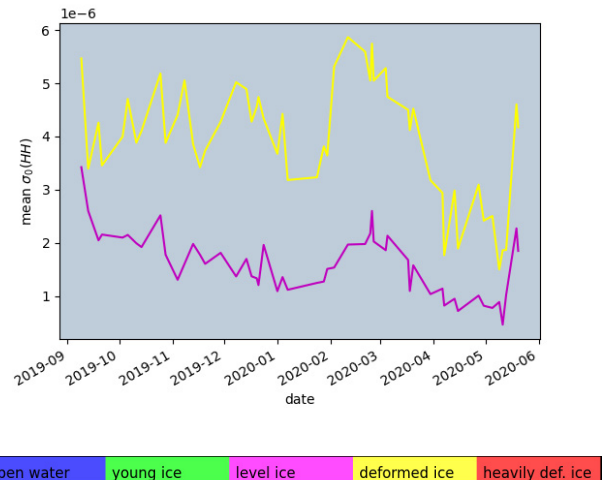
Fig. 1. Chart showing the mean sigma nought calibrated backscatter for the HH band, for LI and DI from October to May. The data are extracted from 52 manually labeled TerraSAR-X scenes.

by pixel, while this poses no problem to an autonomous algorithm. Such autonomous solutions have been proposed as early as 1986 [1]. Early algorithms were based mainly on extracting texture and polarimetric features from the image and then performing classifications using lookup tables [2] or Bayes classifiers [3]. In parallel, the idea of using neural networks for the same task was also investigated [4]. Similar data-driven algorithms are since becoming more attractive, as the volume of available data and the computational power are steadily increasing. Algorithms vary from simpler artificial neural networks using initially computed texture features [5]–[8] or unsupervised segmentation, with manual segment selection [9], [10] to convolutional neural network (CNN) and deep learning techniques [11]. Historically, these approaches have yielded good results for winter and spring seasons [12], [13], where the pack ice is largely dry and changes in ice characteristics are usually minimal. During freeze-up and melt periods, however, classification becomes increasingly difficult. The main challenges here are wet snow lowering radar penetration depth, snow metamorphism, and increased ice dynamics [14]. We observe increased backscatter in those transitional seasons and a general downward trend in radar response after freeze-up (Fig. 1). However, the warmer seasons also bring a loss of contrast between the ice types.

Due to the decreased penetration depth, the SAR texture features, essential to most autonomous classifications, become decreasingly reliable as the backscatter signal becomes more uniform across the different ice types. A possible approach to tackling this is the inclusion of more contextual image data, for example, with larger sliding windows around the ice to be classified. Then, using automated feature extraction and classification with a CNN is especially helpful because the neural network can learn to relate all the information in the window to only the center pixel one that is trying to classify. Thus, it handles large contextual windows better than texture feature-based classification. Recently, it has, for example, been successfully employed for C-Band Sentinel 1 imagery [11] with good results at lower resolution. In this work, we further develop such a classifier, to be able to deal with high-resolution X-band data.

C- and L-band SARs have historically been preferred for sea ice classification. Not only is there greater coverage, with large satellite missions such as Sentinel-1 and Radarsat, but also longer wavelengths offer bigger penetration depths [15] and make it easier to discriminate between ice classes from backscatter and texture features alone. The classification from X-band SAR consequently has more to gain from the inclusion of additional contextual image data and can provide ice types at high resolution.

We propose a scheme that allows to classify pixelwise at high accuracy by using context windows at various zoom levels. In a postprocessing step, we reintroduce some spatial awareness by using a dense conditional random field (DCRF). This concept of adding spatially aware boundary refinement has been implemented in image segmentation as early as 2014 [16]. Random fields have been used successfully in the past for automatic ice charting [17] and have shown promise as a postprocessing step with sparse labels in image processing [18]. The combination of a CNN and conditional random field has also recently been shown to be successful for ice concentration mapping [19].

A central challenge in producing a robust classifier lies in the absence of comprehensive ground-truth data. Despite large amounts of satellite data being collected daily, *in situ* observations are rare. Some approaches have used manually labeled operational ice chart data as ground truth, but the polygon size of ice classes is quite large, so the effective resolution is low and it is known that the same patches of ice are therefore sometimes labeled inconsistently in different scenes. This problem is discussed in [11], where a 50×50 pixel patch of Sentinel 1 imagery is classified using a CNN. At such patch size, the effects of the rough labeling are still manageable, but when classifying smaller patches, it becomes increasingly impactful. Classification at high resolution thus relies on manually labeled data. Here, however, due to the dediousness of the labeling process, there is naturally less data. Thus, it often struggles to capture the span of possible backscatter across different ice types—especially across different seasons.

To truly obtain a measure of robustness of a classifier, one has to show continuity in the classification of overlapping and near coincident SAR scenes, which demonstrates that a patch of pack ice is predicted to be of the same class across different scenes. For large lower resolution scenes, this is feasible, but even here, much research (e.g., [20]) has been focused only on few such overlapping scenes and robustness across a greater range of conditions is still a challenge. At high resolution, coverage is small, and thus, it is an even more complex task to image a small region of pack ice for an extended time. Not only does the drift of the ice have to be tracked, but it also needs to be predicted, due to the delay of the ordering and the capturing of a scene. Over the course of Multidisciplinary Drifting Observatory for the Study of the Arctic Climate (MOSAiC), this task has been tackled by a variety of spaceborne SAR sensors and coordinated by the authors. In this investigation, we use such a dataset captured by the TerraSAR-X satellite in Dual-pol StripMap (HH, VV) mode. It presents the opportunity to validate the robustness of a classifier over an extended time period and a large number of scenes.

While this research stands alone as to the applicability of deep learning techniques for high-resolution SAR ice classification, it also serves as a preliminary study for further investigations using the quantitative *in situ* data collected during the MOSAiC mission, which is currently being processed and quality checked [21] The aim is to use the robust techniques described in this article in further research with high-resolution airborne measurements [22].

II. DATA

The training and test datasets used in this article are comprised of 44 and 8 TerraSAR-X Dual-pol StripMap scenes, respectively. The test scenes contain one randomly chosen scene from every month of the drift. The data points extracted from the 44 training scenes were split into two disjoint training and validation sets, with a size ratio of 9:1. The classifier is trained on the training set, while performance on the validation dataset is used to stop the training in time to prevent overfitting. The data used for robustness analysis are made up of 162 scenes. We will henceforth refer to it as the robustness evaluation dataset. The scenes were acquired between September 2019 and May 2020 over or near the Polarstern vessel, during its drift with the Arctic pack ice. The two channels acquired are the HH and VV polarizations. The images have a row and column spacing of 3.5 m and are typically around $16\,000$ pixels $\times$ 4000 pixels in size. This corresponds to 56 km $\times$ 14 km.

Labeling was done by hand on the basis of the X-band SAR data for five classes chosen to be in line with qualitative *in situ* observations made by members of the MOSAiC expedition. First, we found suitable classes and respective areas to be labeled using the *in situ* observations as a guide. Then, the established logic was extrapolated to the rest of the areas manually using the SAR data only. The five classes are shown in Table I. The color coding used can be found there as well.

All scenes in the robustness analysis dataset entirely contain the immediate area around Polarstern (a 3 km $\times$ 3 km square). To keep similar time spacing between scenes, not more than one scene was used per day.

Due to the drift with the Arctic pack ice, the RV Polarstern entered very high latitudes at the beginning of 2020.

TABLE I
TABLE SHOWING CLASS DEFINITIONS AND LABELS

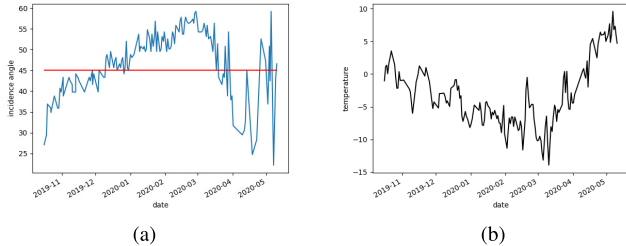| label	meaning	colour
OW	open water	blue
TI	thin ice	green
LI	level ice	magenta
DI	deformed ice	yellow
HDI	heavily deformed ice	red

(a)   (b)

Fig. 2. (a) Average incidence angle that the scenes were acquired at. The red line at 45° indicates the limit of the full performance range of the satellite sensor. (b) Mean ambient temperature measured on the Polarstern vessel over the time period. (a) Incidence angle θ. (b) Temperature T.

Fig. 3. Flowchart showing the pipeline for the proposed ice classification algorithm.

In Fig. 2(a), we see that in this time, the SAR images were consistently taken outside of full performance range, which is between 20° and 45° for StripMap images. The SAR measurements for such high incidence angles have significantly lower signal-to-noise ratios, making it increasingly difficult to differentiate ice types. Weather conditions varied throughout the mission, including events such as storms and warming periods [Fig. 2(b)]. Their effects in regard to this study are constrained to the contribution to increased ice dynamics, as the radar signal is not susceptible to atmospheric conditions at X-band.

III. METHODOLOGY

First, we will give an overview of the general approach, and then, the individual parts are described in detail. The core of the classifier is a CNN. For added robustness, a discriminator and a DCRF are used for further processing. The algorithm assigns one of five classes (Table I) to a given 5×5 pixel patch of an SAR imagery.

Fig. 3 shows the classification pipeline for the algorithm used. After preprocessing, features of varying scope and resolution (zoom levels) are supplied along with each 5×5 image slice that is to be classified (Table II). A CNN is fed these features and makes an initial prediction for that patch (see Fig. 10 in the Appendix for details). The predictions are then checked by a second discriminating network that removes some labels, deemed to be misclassifications. Finally, a DCRF smooths over the labels by relating the spatial context of the labeled data and the underlying image. This also fills the missing values left by the discrimination step.

1) Preprocessing: In the initial step of data preprocessing, the original dual-polarized SAR scene is calibrated to slant range (β_0) and a false-color composition of the data is constructed. The composition consists of four channels: HH, VV, HH-VV, and HH/VV. The difference and ratio are common for manual ice charting and visualization of SAR
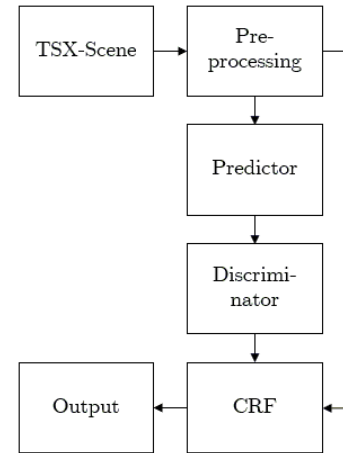
scenes, as they promote contrast across ice types and open water (OW). In addition, they have been shown to be useful for classification in the past [23]. The raw backscatter channels HH and VV are rescaled with a tanh function. The composite features HH-VV and HH/VV are also scaled with a tanh function and an additional offset. The exact parameters are manually selected to give a good contrast. While the network is in principle able to learn these features, feeding them directly alleviates some of the workload of the network and gave improved results in preliminary testing.

2) Convolutional Neural Network: The core of our classification approach is a CNN (see Fig. 10). It predicts one of five classes for each 5×5 pixel patch of the SAR scene. These patches ("local features") are appended by some additional information of the surroundings. The first of these additional features is a 16×16 pixel patch ("superlocal feature") of the surrounding area, which is taken from the SAR scene rescaled by a factor of 5. Thus, moving to the right one 5×5 patch in the original image moves one pixel to the right in the rescaled product the superlocal patch is taken from. This patch gives some insight into the surrounding area, allowing the algorithm to take advantage of surface features nearby, such as ridges or leads, to gain some spatial context. For example, the CNN might learn that heavily deformed ice (HDI) is more likely to occur with well-defined edges, such as the edge of a multiyear ice floe, in the surrounding area. The patch sizes of 5×5 and 16×16 were established empirically. In general, both of these are compromises of resolution and accuracy. The larger the windows will get, the better the accuracy will become (as there is more information in the image to use). However, the more difficult also becomes for the classifier to relate all this information to only the data in the center of the patch. This leads to a lack of effective resolution in the classified product.

To give a more complete picture, the entire scene (or the largest possible near quadratic slice of it) is additionally resized to 64×64 pixels and input to the model ("global feature"). The StripMap data used here are captured in rectangular strips, typically around four times longer than wide. In such a case, we split the scene into four near quadratic slices along the azimuth axis (the long axis). The global input

feature allows some insight into large-scale features, such as the general brightness of the scene, interfaces between ice masses, or (not important for this dataset) the ice–water edge. These can then be related to the high-resolution features and particularly helped with classification of scenes that had very high backscatter (e.g., melt onset) or low radar response (e.g., high incidence angle). As we parse the entire range domain of the scene here, it is no longer possible to ensure that the region to be classified lies in the center of the image. Consequently, a fourth input ("extra feature") consisting of four parameters is provided, containing the position of the region to be classified in the larger 64×64 input. It also contains the incidence angle of the patch and the time at which the product was acquired. A summary of these input features is included in Table II.

The training dataset consists of 44 scenes. Most often, the scenes are split into four near quadratic slices for the global feature. Thus, the number of different inputs for that feature is only $\approx 44 \times 4 = 176$. A substantial risk that the algorithm overfits to the training data lies here. It might memorize where, in each scene, which ice class is located, rather than deduce the ice class from a combination of the inputs. To combat this potential problem, some data augmentation techniques are applied [24]. Specifically, we use random crops, rotations, and flips.

The effect of the incidence angle on radar backscatter is well researched, and thus, its inclusion in the model is easy to motivate. Incidence angle normalization to σ_0 has been shown to be useful in the past, but this does not account for different gradients across the ice classes, which are reported, for example, by Mahmud *et al.* [25] and Lohse *et al.* [26]. In our study, we included the incidence angle as an input to the classifier. This gives the model the opportunity to learn these differing incidence angle dependencies of the sea ice backscatter, similar to the classifier in [26], provided that the range of incidence angles is covered well enough by the training data. However, as we do not force the network to make use of the incidence angle information, this is only done implicitly. Because our classifier spans multiple seasons that have strong correlations to ice-type distribution and the radar response, including the acquisition time also proved to be a helpful parameter for the model. As we only have 44 different acquisition times, we applied strong artificial noise to reduce the risk of overfitting. The random noise added is sampled from a normal distribution with a standard deviation of one week.

The exact details of the parameters of our network are largely based on heuristics and our experiments. The 3×3 kernel size has proven most useful, as our input features are not that large themselves. We found more success in downsampling with convolutions with step size 2, instead of max-pooling layers—in tests, it seems that the network lost a little information upon in the max-pooling layer that was still useful for classification. We apply LeakyReLU as an activation function, which introduces some necessary nonlinearity and does not suffer from the problem of vanishing gradients. The latter property is especially useful for deeper networks. We used strong regularization with multiple dropout layers with a dropout rate of 0.3, as our data are still quite sparse

TABLE II
INPUT FEATURES AND THEIR CONTENT. THE RESOLUTION r_{rel} IS GIVEN RELATIVE TO THE FULL RESOLUTION PRODUCT. THE FOUR CHANNELS USED ARE (HH, VV, HH-VV, AND HH/VV). THE COORDINATES x_{global} AND y_{global} ARE THE NORMALIZED PIXEL COORDINATES OF THE REGION TO BE CLASSIFIED IN THE RESCALED GLOBAL PATCH

input name	dimensions	contents
local	5x5x4	4 channels, $r_{rel} = 1$
superlocal	16x16x4	4 channels, $r_{rel} = 1/5$
global	64x64x4	4 channels, $r_{rel} \approx 1/64$
extra	4	x_{global}, y_{global}, θ_{inc}, t

in contrast to the scope of possible backscatter signatures from sea ice, and thus, overfitting is a concern. In addition, we opted for small spatial dimensions after convolutions (before flattening) and lower number of neurons to force the network to parameterize the input features—which lead to better extrapolation to unseen data. Batch normalization in the early layers slightly sped up the convergence of our network. The Adam optimizer was used to update weights during training.

The network uses a categorical cross-entropy loss appended by an additional term from the FEature and Spatial relaTional regulArization (FESTA) loss [18], specifically the distance of the softmax outputs. The additional term encourages the separation of labels independent of correct classification, which helped with the convergence of our classifier.

In addition, we make use of smoothed labels. Instead of feeding a one-hot vector as a label—where the correct label is denoted as 1 all others as 0, uncertainties are integrated into the labeling in a rudimentary way. The idea is to treat the label vector as a set of probabilities rather than as a Boolean vector. This is particularly useful for the ice classes where manual labeling is most error-prone. In our case, distinguishing deformed ice (DI) and level ice (LI) benefitted most of this treatment because it is partially nonlocal property. Explicitly, ice has varying deformations across larger regions so that individual pixel-sized areas might be smooth, but it is apparent from the surrounding ice that the area is deformed. We found that including uncertainties only minimally lowered the accuracy. However, it leads to significantly increased robustness, which we preferred in this case. This is in line with observations made in [27]. To smooth the labels, we sample a random number from a uniform distribution in a given interval that conveys the uncertainties. The intervals used across the different classes are listed in Table III and were chosen qualitatively through testing and in line with experience of which areas are difficult to label. Note that after random sampling, each output vector is normalized.

The network was implemented using the TensorFlow library for python [28]. On an Intel i7-9850H, a commercially available midrange CPU, inference for an entire scene consisting of $\approx 2.5 \times 10^6$ classifications takes around 8.5 min.

3) Discriminator: The discriminator model has a near-identical structure to the classifier (Fig. 11), except for the additional input layer containing a proposed label and the output being 1-D. Its task is to check whether the proposed label is correct or not. This binary classification is

TABLE III

Label Smoothing Parameters Used. $U_{[a,b]}$ Denotes a Random Sample From a Uniform Distribution From the Interval $[a, b]$. The Label Vectors Were Normalized After Random Sampling

	ow	ti	li	di	hdi
ow	$U_{[0.7,1]}$	0	$U_{[0,0.3]}$	0	0
ti	0	1	0	0	0
li	0	0	$U_{[0.9,1]}$	$U_{[0,0.1]}$	0
di	0	0	$U_{[0,0.3]}$	$U_{[0.7,1]}$	0
hdi	0	0	0	0	1

fundamentally easier than predicting one of five classes and can correct for some systematic errors the classifier makes. The discriminator is trained on randomly mislabeled data as ground truth for mislabeled patches, which performed better than a discriminator trained specifically on the correctly labeled and mislabeled data of the classifier. We suspect this to be the case because specific training promotes an overfit to training data. This step particularly helps with mitigating OW and thin ice (TI) misclassifications.

4) Conditional Random Field: The pixels deemed to be wrongly classified by the discriminator are removed from the classified product. Then, a DCRF is applied, which has a bilateral kernel next to the unary potential. This fills the missing values as well as clears up some noise-like mislabels, such as single pixels classified differently than all their surrounding pixels. The application of the DCRF is straightforward using a Python implementation [29] for an algorithm published in [30]. A bilateral approach is used, with the energy function given by a unary ϕ_u and a bilateral term ϕ_b, such that for N feature vectors y_i and labels x_i

$$E = \sum_i^N \phi_u(x_i, y_i) + \sum_{i,j}^N \mu(x_i, x_j)\phi_b(y_i, y_j). \quad (1)$$

The label compatibility function μ is learned and describes the relationship of how likely labels are to occur next to each other. Thus, incompatible labels close to one another are penalized by the energy function. In this case, the features consist of a vector of color intensities $\vec{I} = (I_r, I_g, I_b)$ across the RGB channels of a color composite image as well as the position of the pixel $\vec{P}$. The RGB channels in the color composite image are VV, HH-VV, and HH/VV, appropriately scaled to capture the relevant dynamic range. The unary potential is given as the logarithm of the probability $p(x_i|y_i)$ of label x_i given feature y_i. It is modeled by the softmax output of our classifier. The bilateral term consists of weighted differences in position and color. Thus, the energy function can be expressed as

$$E = \sum_i^N \log p(x_i|y_i) + \sum_{i \neq j}^N \left[\frac{\|\vec{P}_i - \vec{P}_j\|}{s_{xy}} + \frac{\|\vec{I}_i - \vec{I}_j\|}{s_c} \right] \quad (2)$$

where $\|\ldots\|$ denotes the Euclidian norm. The weights s_{xy} and s_c were adjusted manually to balance smoothness and classification accuracy.

A. Evaluation

As was mentioned in Section I, we have a large number of scenes to test the robustness of our classifier across eight

months of different conditions in the arctic ice. The idea is to test the ice distribution of the same patch of sea ice over this entire time period and investigate how it changes over this timeframe. This should give some insight into how stable the classifier performs at high resolution. Given the positioning of the RV Polarstern, there is no pack ice that we can track more accurately than the ice around the research vessel itself. Thus, this is the region we will use. We evaluate all scenes in the robustness analysis set and then calculate the probability of pixels not changing class, which we can use as a measure of robustness. The window chosen is approximately 3 km × 3 km in size. Of course, we do not expect the ice to stay static over the entire time period; ice dynamics and new ice growth will change the ice-type distribution. A rapid change in ice type is constrained to the OW and TI classes. Change due to shifting of the floe is easily spotted by looking at the individual images and thus can be considered qualitatively during the evaluation of our model. It should also be noted that care was taken not to label the area used for robustness analysis in the training set, so the classifier has not "seen" these regions. Using the ship's GPS information, we can correct for the drift of the ice using a coordinate transformation to local ship coordinates [31] and identify the same area for each scene.

To obtain a quantitative measurement of robustness, we define a robustness criterion for our analysis. We will deem a (pixel-sized) area of ice to be classified *robustly* in one scene if the same prediction is made for the previous and the following scene. As we want to define this criterion for a single scene, note that the computed probability $P_i^3(c)$ of finding the same class c at the same spot for a scene i and its two nearest neighbors is a product of the probabilities $P_i(c)$ of having robustly classified in each of the scenes

$$P_i^3(c) = P_{i-1}(c)P_i(c)P_{i+1}(c). \quad (3)$$

With the assumption that $P_i(c) \approx P_{i+1}(c)$, we will approximate the probability $P_i(c)$ of having classified robustly for scene i and class c as

$$P_i(c) = \left(P_i^3(c)\right)^{\frac{1}{3}}. \quad (4)$$

The regions of ice we use to test this are the pixels in the stabilized images, such as seen in Fig. 7. In the following analysis, we make statements based on the assumption that the same pixel over three scenes actually maps to the same physical area of ice for three consecutive scenes. However, we note that this is not universally true, as the stabilization we use is not perfect and ice dynamics are entirely neglected in this assumption. We treat these phenomena as some underlying noise in the analysis and must hence satisfy ourselves with computing a lower bound for robustness.

IV. Results

The classifier was trained on 44 scenes and tested on 8, which does not contribute to training data. The eight scenes that make up the test set are randomly selected scene from each month of October–May. We will be looking at the performance of the classifier across the two datasets. As a comparison, we also use a simple VGG16-inspired [32] architecture as an
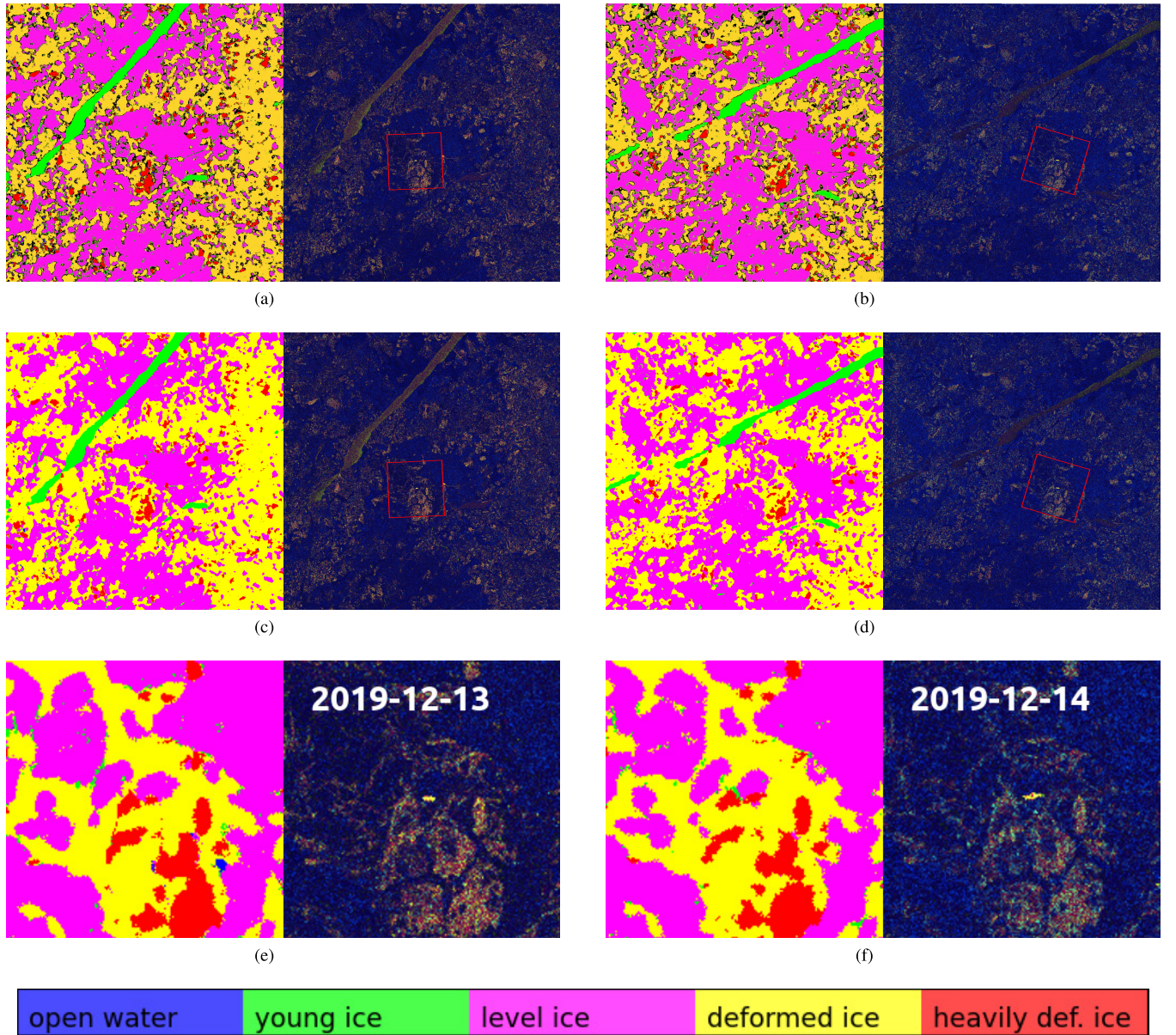
Fig. 4. Illustration of robustness analysis for two consecutive days in December 2019. (a) and (b) Subscenes after classification and discrimination. In (c) and (d), DCRF has been applied. (e) and (f) Scenes cropped and rotated to the region of interest. The false-color compositions are using (VV, HH-VV, and HH/VV) in the (R, G, and B) channels, respectively. The top two rows additionally show the area used for robustness analysis, which is cropped to in the final row. (a) December 13, 2019, $\theta_{inc} = 45.69°$. (b) December 14, 2019, $\theta_{inc} = 49.55°$. (c) December 13, 2019, $\theta_{inc} = 45.69°$. (d) December 14, 2019, $\theta_{inc} = 49.55°$. (e) December 13, 2019, $\theta_{inc} = 45.69°$. (f) December 14, 2019, $\theta_{inc} = 49.55°$.

TABLE IV

CONFUSION MATRIX SHOWING THE PERCENTAGE OF THE VGG16'S PREDICTED CLASSES (COLS) FOR ALL GROUND-TRUTH LABELS (ROWS) ON THE TRAINING DATA. CLASSES ARE OW, TI, LI, DI, AND HDI

	OW	TI	LI	DI	HDI
OW	90.89%	1.85%	7.21%	0.04%	0.01%
TI	0.1%	78.84%	11.34%	6.25%	3.46%
LI	0.8%	6.48%	89.14%	3.5%	0.08%
DI	0.04%	6.82%	17.22%	75.1%	0.82%
HDI	0.01%	1.82%	0.24%	0.62%	97.85%

TABLE V

CONFUSION MATRIX SHOWING THE PERCENTAGE OF OUR NETWORK'S PREDICTED CLASSES (COLUMNS) FOR ALL GROUND-TRUTH LABELS (ROWS) ON THE TRAINING DATA. FOR EXAMPLE, 0.11% OF DATA POINTS OF OW WERE INCORRECTLY PREDICTED TO BE TI. CLASSES ARE OW, TI, LI, DI, AND HDI

	OW	TI	LI	DI	HDI
OW	99.29%	0.11%	0.38%	0.21%	0%
TI	0.28%	93.72%	1.36%	4.11%	0.53%
LI	0.16%	0.54%	97.38%	1.89%	0.02%
DI	0%	0.58%	1.47%	97.84%	0.1%
HDI	0%	0.11%	0%	0.62%	99.26%

alternative classifier (see Fig. 12 in the Appendix for details), which classifies based only on the superlocal 16×16 input data. The classifiers' performances are shown in Tables IV and V for the training set and Tables VI and VII for the test set.

The two classes on which our and the VGG16 style classifier perform worst are OW and TI. The classification across the

TABLE VI

Confusion Matrix Showing the Percentage of the (Used for Comparison) VGG16 Style Network's Predicted Classes (Cols) for All Ground-Truth Labels (Rows) on the Test Data. Classes Are OW, TI, LI, DI, and HDI

	OW	TI	LI	DI	HDI
OW	80.89%	4.6%	14.23%	0.27%	0.01%
TI	2.4%	54.48%	17.42%	15.36%	10.34%
LI	0.54%	11.15%	81.85%	6.44%	0.02%
DI	0.17%	6.95%	13.64%	78.78%	0.46%
HDI	.010%	1.36%	0.21%	0.4%	98.02%

TABLE VII

Confusion Matrix Showing the Percentage of Our Networks' Predicted Classes (Cols) for All Ground-Truth Labels (Rows) on the Test Data. For Example, 8.12% of Data Points of OW Were Incorrectly Predicted to be TI. Classes Are OW, TI, LI, DI, and HDI

	OW	TI	LI	DI	HDI
OW	74.86%	8.12%	16.4%	0.61%	0%
TI	5.05%	71.07%	7.03%	14.32%	2.52%
LI	0.49%	0.62%	95.51%	3.38%	0%
DI	0.02%	0.55%	2%	97.37%	0.05%
HDI	0%	0.02%	0%	0.92%	99.06%

other three ice classes is more accurate, particularly for our classifier. HDI stands out as being especially easy to classify accurately. OW and young ice examples are scarcest in the training and test scenes. Despite balancing the amount of samples per class upon training, this means that the dataset is a lot less diverse for these samples. In other words, the number of regions with OW or TI is significantly lower than that of the other classes. This can be seen (albeit at a smaller scale) in Fig. 5.

To illustrate the entire process of assessing our classifier, we give an example with two consecutive scenes from December 13 and 14, 2019 in Fig. 4. In the first step, the StripMap scene is cropped along the longer range axis to a near quadratic slice, which is needed for the global input feature (Table II). This slice is then labeled using the classifier and checked by the discriminator. The results of this step are shown in Fig. 4(a) and (b). The next step is to apply the DCRF to refine labels and fill missing values left by the discriminator. The results of the DCRF for the two example scenes are shown in Fig. 4(c) and (d). Finally, the image is rotated and cropped to the surroundings of the RV Polarstern, allowing us to image the same region of ice continuously. For the two scenes from December used as an example, the cropped images are shown in Fig. 4(e) and (f).

By executing the procedure shown in Fig. 4 for all scenes from October to May, we have the data necessary to perform the quantitative robustness analysis. Fig. 5 shows the evolution of the predicted ice-type distribution over the investigated time span.

The distribution chart allows some insight into the performance of the classifier over large timescales and shows lower stability, especially in the discrimination of level and DI. Spikes of OW and TI are generally tied to some ice dynamics. To gain some additional insight into the variance of
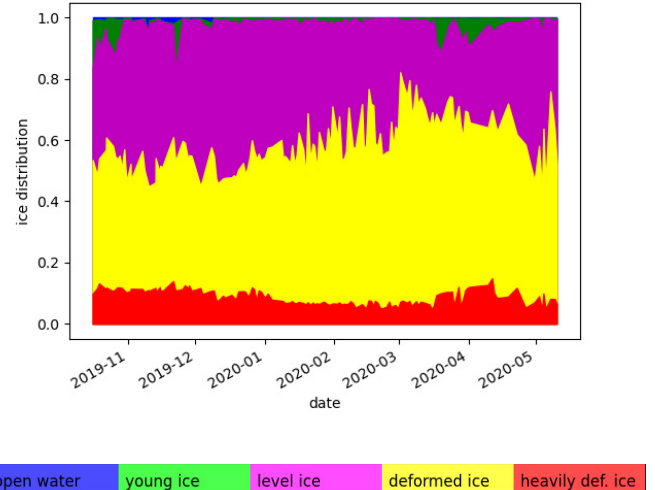

Fig. 5. Chart showing the predicted ice distribution in the 3 km × 3 km area surrounding the Polarstern vessel from October to May.
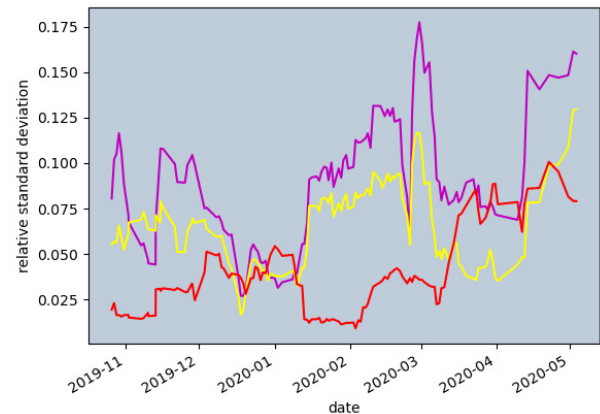

Fig. 6. Relative standard deviation of the distribution of LI, DI, and HDI at each date and the four nearest scenes.

these classes, we compute the relative standard deviation of the ice-type fraction for every scene and its four nearest neighbors (Fig. 6). When interpreting this as a deviation of the classification, we implicitly assume the real ice-type distribution to be stable over five neighboring scenes, which neglects physical changes of the surface. Specifically, we cannot include the OW and TI classes, which are rare and only present in case of strong sea ice dynamics such as leads forming and thereby not able to be analyzed in this way, as we cannot assume these classes to be stable over time.

The chart of the standard errors reveals three time periods with heightened error. In fact, the one stable period around the beginning of January stands out. Here, conditions are optimal, as ice dynamics are minimal and the incidence angle is inside the full performance range. While the early and later periods of increased variance can likely be explained by snow metamorphism, wet snow or increased ice dynamics in melt and freeze seasons, the increased uncertainties from mid-January to early March can be rationalized with the increased incidence angle during this time period [see Fig. 2(a)]. It is also apparent that
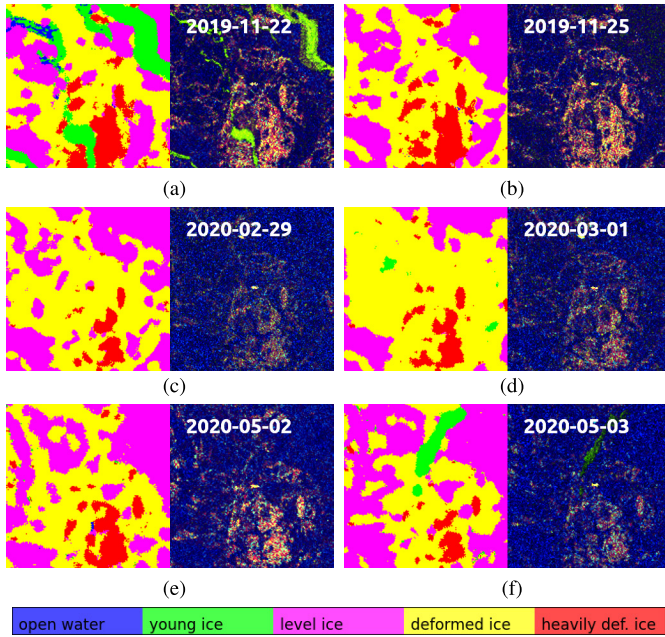
Fig. 7. Pairs of classified scenes from time intervals with low robustness. The left parts show the cropped classified image and the right parts are false-color compositions. (a) November 22, 2019, $\theta_{\mathrm{inc}} = 40.71°$. (b) November 25, 2019, $\theta_{\mathrm{inc}} = 43.33°$. (c) February 29, 2020, $\theta_{\mathrm{inc}} = 57.36°$. (d) March 1, 2020, $\theta_{\mathrm{inc}} = 54.27°$. (e) May 2, 2020, $\theta_{\mathrm{inc}} = 36.90°$. (f) May 3, 2019, $\theta_{\mathrm{inc}} = 50.82°$.
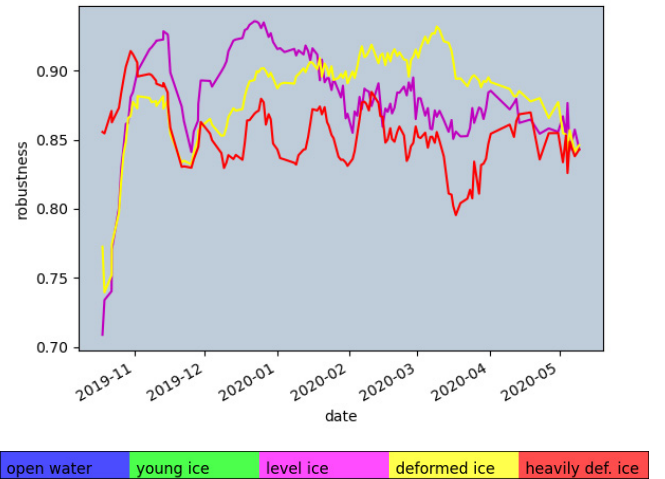

Fig. 8. Chart showing a moving average of the probability $P_r(c)$ of robust classification for three ice classes: LI, DI, and HDI. We compute $P_r(c)$ as the percentage of robustly classified pixels of class c per scene.

this cause of error plays a role in the increased uncertainties observed during the melt season.

To qualitatively gauge the performance of the classifier in these three periods with increased error, we highlight three pairs of scenes (Fig. 7) from those time spans, specifically late November, late February, and finally early May.

The scenes from the end of November are only three days apart yet drastically different due to ice dynamics. In the earlier scene [Fig. 7(a)], one can see some freshly frozen over leads with TI cover, and in the later scene [Fig. 7(b)], the leads have closed up again and all signs of young ice have disappeared. These pictures document the most drastic of these events, where the central floe split; however, most of the strong ice-type deviations in early winter can be attributed to such events and are thus real changes of the surface. The pair of scenes from late February [Fig. 7(c) and (d)] are both taken at very high incidence angles, far above full the performance range of 20°–45°. It is evident from the images that the signal is significantly weaker. At this angle, we see an overestimation of the DI class in the second image. This is especially notable in the top-left quarter of the patch, which was dominated by a LI surface, but is classified as almost entirely DI in the scene from the first of March. At such high incidence angles, the classification seems to become more volatile. Two scenes from early May [Fig. 7(e) and (f)] give insight into how both ice dynamics and incidence angle changes are responsible for high variance in the scenes from early May.

Most of the high variances in ice class distribution change can be attributed either to ice dynamics or to struggles with high incidence angles. The classifier seems robust in the discrimination of classes with a larger area, but the transitional

TABLE VIII
AVERAGE PROBABILITY $P_r(c)$ OF ROBUST CLASSIFICATION FOR
LI, DI, AND HDI ACROSS THE ENTIRE DATASET

	LI	DI	HDI
P_r	88.34%	88.62%	85.34%

TABLE IX
BOUNDS OF ROBUST AND CORRECT CLASSIFICATION FOR
LI, DI, AND HDI OF THE CLASSIFIER

	$P_{rc}(c)$
LI	$84.09\% < P_{rc}(LI) < 95.19\%$
DI	$86.71\% < P_{rc}(DI) < 97.84\%$
HDI	$84.71\% < P_{rc}(HDI) < 99.26\%$

areas between classes are seemingly classified less robustly [see the extent of the LI on the left of the image Fig. 7(e) versus Fig. 7(f)]. This effect is particularly evident at high incidence angles.

In Fig. 8, we can observe the development of robust classification across the analyzed time span. This was smoothed over by a moving average, weighted with a quadratic function, and averaging over five scenes. Note that the dip in the beginning is due to strong ice dynamics. In Fig. 9, we show a comparison of our model with the VGG16-inspired classifier for two months. We also compute an average robust classifications probability over the entire time span. The results are shown in Table VIII.

We now propose a probability $P_{rc}(c)$ of robust and correct classification as the product of the two, i.e.,

$$P_{rc}(c) = P_r(c)P_c(c). \tag{5}$$

With classification probabilities P_c from Table VII and the lower bounds of P_r from Table VIII, we can compute the bounds of probabilities of robust and correct classification P_{rc} for the three solid ice classes. The results are shown in Table IX.
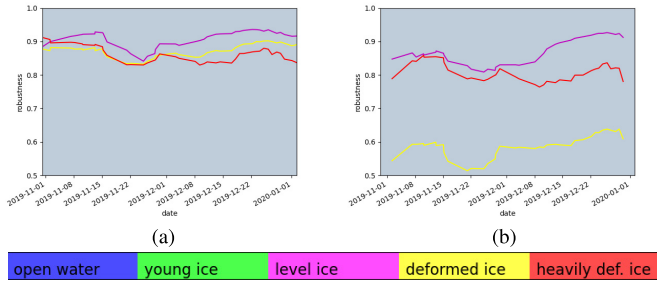
Fig. 9. Charts showing a moving average of the probability $P_r(c)$ of robust classification for three ice classes LI, DI, and HDI, for our classifier and a VGG16-inspired model. (a) Our classifier (Fig. 3). (b) VGG16-inspired CNN (Fig. 12).

V. DISCUSSION

Before turning to the advantages of our approach, we will mention some limitations and challenges. First, let us discuss the data itself—the foundation of any machine learning approach. The training dataset of 44 scenes is, of course, not comprehensive enough to capture all the intricacies of different backscatter from varying ice types, which makes it difficult to classify robustly. Leads freezing over are a good example for one of these regions. Not only is their occurrence sparse in the dataset, but also the dynamics during initial freeze over have a great effect on radar response and are fast relative to the revisit time of the satellite. This makes it difficult to capture enough samples in the training set for the classifier to correctly interpret the entire space of possible radar backscatter. We can observe this struggle in some scenes where the radar response of a frozen lead is so bright, and it becomes very similar to HDI backscatter. This can, for example, occur when frost flowers form atop the lead, leading to high volume scattering. Here, the classifier struggles to differentiate the two classes.

The OW classification also proved to be a challenge for this dataset. Traditionally, the polarization ratio proves very useful in distinguishing this class. We can observe that at high incidence angles, the radar response becomes very similar to that of young smooth ice and the discrimination between the two suffers.

Manual labeling is definitely the greatest source of underlying error and bias. Despite having mitigated the effect of errors with the use of smooth labels, there are some biases arising from manual labels that smooth labeling cannot compensate. This bias is not merely a case of being more likely to mislabel a certain class—this can be kept minimal by only labeling classes that are discernible with certainty—it is rather that the choice of labeled regions is already filtered by a human selection process. For example, there is a tendency not to label a region with a small area as it would make the labeling process very tedious. This translates to the classifier that struggles with smaller regions of one class, often wrongfully mislabeling them to be the same as the surrounding ice class. In addition, when manually selecting polygons, labels at the boundaries between classes are naturally much sparser than labels in the center of ice classes, which leads to increased difficulty of classification in these transitional areas between ice classes. When viewing the classified robustness analysis

data, this effect was most obvious as the boundaries between classes were shifting, while pixels in the center of the same ice regions appear robust. This bias could be eliminated by deriving ground-truth data from *in situ* measurements.

The discrimination of DI and LI relies on nonlocal features and hence suffers most from the abovementioned boundary problem. DI is not always identified by a higher brightness and lower polarization ratio for each individual pixel but also by the density of brighter pixels in the surrounding area. Here, it is especially difficult to define hard boundaries between classes, as the transitional areas between level and DI are not boundaries but a continuum. Hence, it is difficult to define a hard boundary when labeling data manually. Generally, the rule when labeling manually is to only label areas, where one is confident in the label. Therefore, these transitional areas are not only difficult to classify but also sparse in the training dataset, which culminates in misclassification in the transitional areas of DI and LI classes, especially at high incidence angles, where the signal-to-noise ratio suffers (Fig. 7).

The success of the algorithm is self-evident in the discrimination of ice classes at high accuracy in multiple seasons and becomes increasingly apparent in contrast to the VGG16 (see Fig. 9 and Tables IV–VII). Furthermore, the areas of lower robustness can be seen to occur at high incidence angles, well outside of the full performance range of the radar instrument.

Weather effects contribute significantly to snow wetness, metamorphism, and increased ice dynamics. The most notable of these is the seasonal warming and freezing, which leads to decreased robustness in our analysis (Fig. 8). However, our robustness criterion also fails to consider these weather-induced changes and we have less training data available in these time periods as they are at the very beginning and end of the study period. Thus, it is difficult to isolate and make statements about the effect of weather events on the performance of classification.

We have tested this classification approach on Sentinel-1 scenes and obtained the comparable results. We found that the most important parameters to tune, when applying these ideas to different sensors, are the sizes of the contextual windows ("local" and "superlocal" features). On large-scale images, the inclusion of the "global" feature was particularly successful in ice and OW discrimination in the marginal ice zone, where the ice water edge could be detected.

VI. CONCLUSION

We have used accurate geolocation and drift correction to construct a dataset that enabled testing for robustness of SAR ice-type classification quantitatively and we are able to show that our proposed classification method performs accurately and robustly for three surface ice classes: LI, DI, and HDI. OW and TI classes have proven harder to classify. However, it needs to be noted that these classes are also sparser in the dataset and have also been more difficult to identify in some scenes, especially at higher incidence angles. Due to their dynamic nature, we are not able to perform a robustness analysis for these two ice types. We could also

64 x 64 x 4 → 62 x 62 x 16 → 30 x 30 x 32 → 14 x 14 x 32 → 6 x 6 x 32 → 2 x 2 x 64

16 x 16 x 4 → 16 x 16 x 16 → 14 x 14 x 16 → 6 x 6 x 32 → 2 x 2 x 32

5 x 5 x 4 → 5 x 5 x 16 → 4 x 4 x 16 → 3 x 3 x 32

θ, t, x_g, y_g

804 → 64 → 32 → 16 → 8 → 5
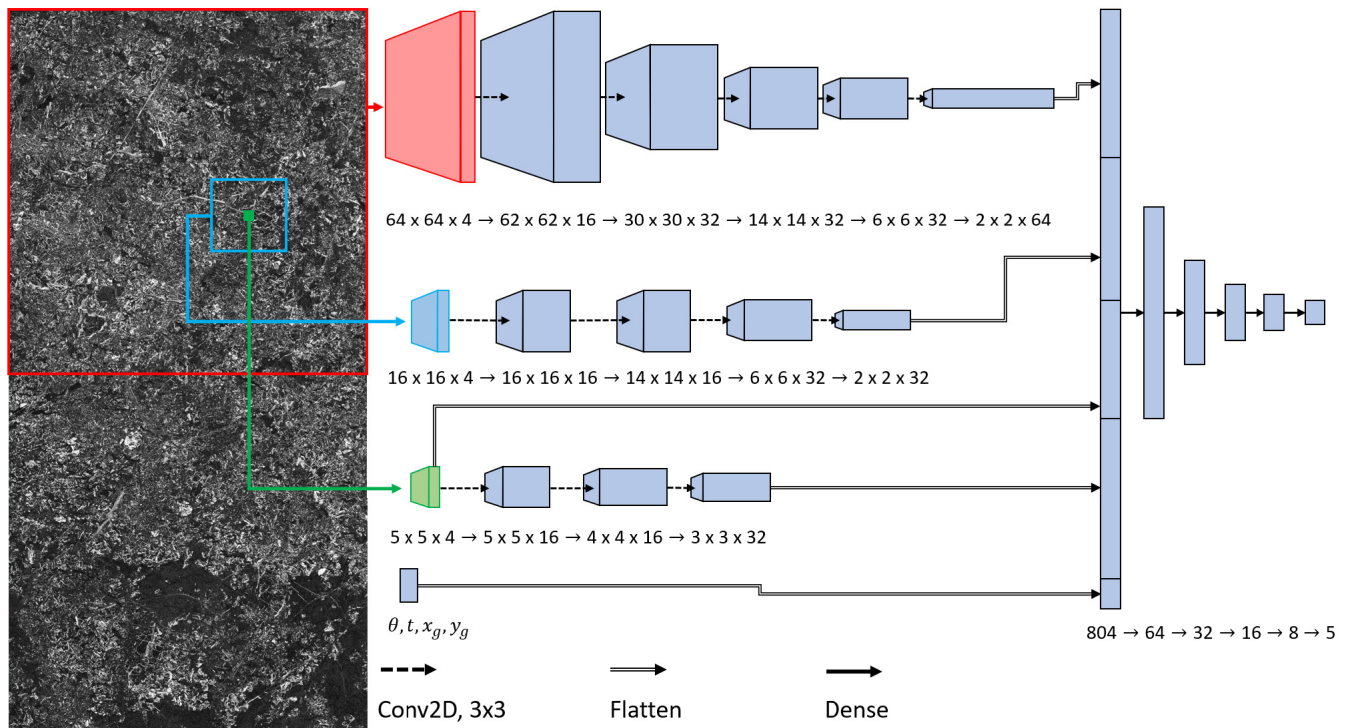
- - - → Conv2D, 3x3   ⟹ Flatten   ⟶ Dense

Fig. 10. Illustration of the CNN architecture used in the proposed classifier. Wherever the spatial dimensions in the convolutional blocks are downsampled (decrease by more than a factor of 0.5), a stride of 2 was used. Not included in the image are the batch normalization layers after the first convolutional layers for each input and the dropout layers used for regularization during training. The parameters x_g and y_g denote the coordinates of the location of the local patch in the global patch, θ is the incidence angle, and t is the acquisition time. Parameter count = 120 421.

64 x 64 x 4 → 62 x 62 x 16 → 30 x 30 x 32 → 14 x 14 x 32 → 6 x 6 x 32 → 2 x 2 x 64

16 x 16 x 4 → 16 x 16 x 16 → 14 x 14 x 16 → 6 x 6 x 32 → 2 x 2 x 32

5 x 5 x 4 → 5 x 5 x 16 → 4 x 4 x 16 → 3 x 3 x 32

θ, t, x_g, y_g

Prediction

- - - → Conv2D, 3x3   ⋯⋯▶ Conv2D, 2x2   ⟹ Flatten   ⟶ Dense
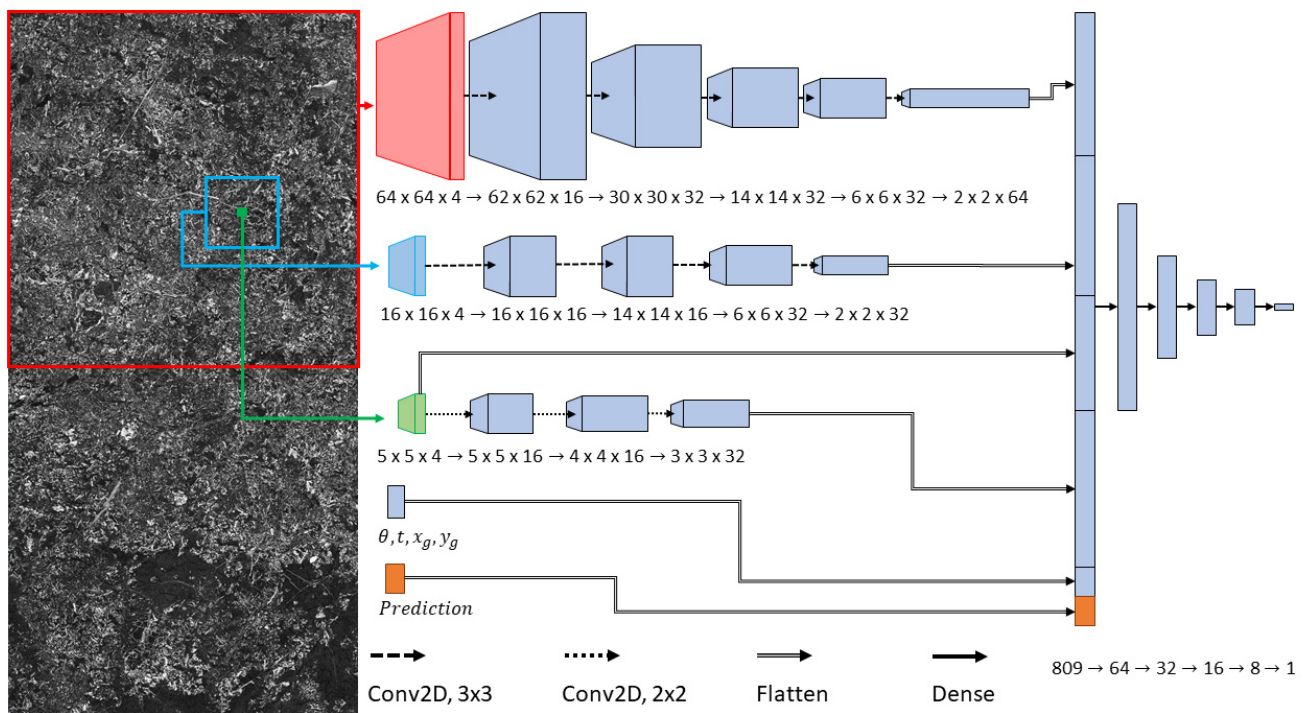
809 → 64 → 32 → 16 → 8 → 1

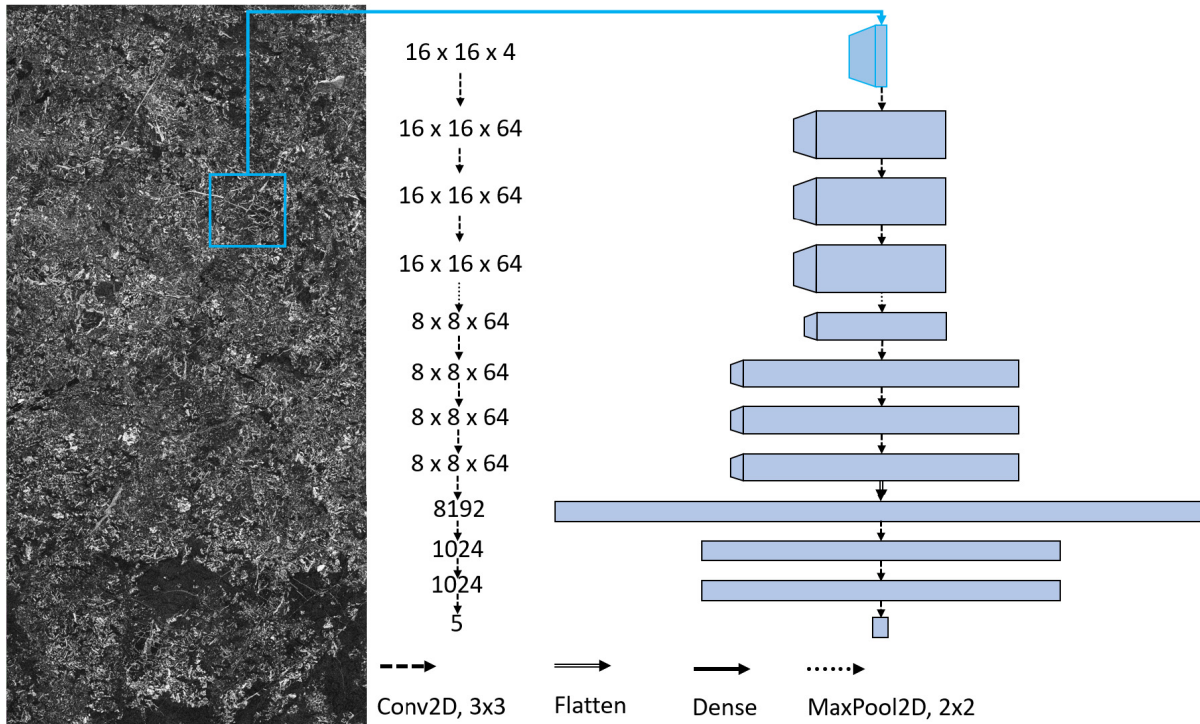Fig. 11. Illustration of the CNN discriminator architecture used in the proposed classifier.

Fig. 12. Illustration of the VGG16-inspired network architecture used as a comparison. Parameter count = 9 889 605.

identify regions of increased classification inaccuracy and lack of robustness that coincide with shortcomings of a manual labeling process. Already, now, our ice-type dataset can provide helpful information for upscaling other MOSAiC *in situ* data to a regional context, such as sea ice physical and chemical properties or ecological samples, which all vary by ice type.

As was mentioned in Section I, this work serves partially as a preliminary study to using these classification methods in analysis with fused measurements, such as airborne laser scanner (ALS) data, e.g., from MOSAiC [21]. Here, we can derive ground-truth labels from the ALS data without any human interaction and thus eliminate the greatest source of bias in the underlying dataset.

APPENDIX
NETWORK ARCHITECTURES
See Figs. 10–12.

ACKNOWLEDGMENT

REFERENCES

[1] M. Fily and D. Rothrock, "Extracting sea ice data from satellite SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-24, no. 6, pp. 849–854, Nov. 1986.

[2] R. Kwok, E. Rignot, B. Holt, and R. Onstott, "Identification of sea ice types in spaceborne synthetic aperture radar data," *J. Geophys. Res., Oceans*, vol. 97, no. 2, pp. 2391–2402, 1992. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/91JC02652

[3] L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 780–795, Mar. 1999.

[4] Y. Hara, R. G. Atkins, R. T. Shin, J. A. Kong, S. H. Yueh, and R. Kwok, "Application of neural networks for sea ice classification in polarimetric SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 740–748, May 1995.

[5] R. Ressel, A. Frost, and S. Lehner, "A neural network-based classification for sea ice types on X-band SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3672–3680, Jul. 2015.

[6] S. Singha, M. Johansson, N. Hughes, S. M. Hvidegaard, and H. Skourup, "Arctic Sea ice characterization using spaceborne fully polarimetric L-, C-, and X-band SAR with validation by airborne measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3715–3734, Jul. 2018.

[7] R. Ressel, S. Singha, S. Lehner, A. Rösel, and G. Spreen, "Investigation into different polarimetric features for sea ice classification using X-band synthetic aperture radar," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 3131–3143, Jul. 2016.

[8] D. Murashkin, G. Spreen, M. Huntemann, and W. Dierking, "Method for detection of leads from Sentinel-1 SAR images," *Ann. Glaciol.*, vol. 59, pp. 1–13, Mar. 2018.

[9] A. M. Johansson *et al.*, "Consistent ice and open water classification combining historical synthetic aperture radar satellite images from ERS-1/2, Envisat ASAR, RADARSAT-2 and Sentinel-1A/B," *Ann. Glaciol.*, vol. 61, pp. 1–11, Jan. 2020.

[10] A. P. Doulgeris, "An automatic v-distribution and Markov random field segmentation algorithm for PalSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1819–1827, Oct. 2015.

[11] H. Boulze, A. Korosov, and J. Brajard, "Classification of sea ice types in Sentinel-1 SAR data using convolutional neural networks," *Remote Sens.*, vol. 12, no. 13, p. 2165, Jul. 2020.

[12] A. S. Komarov and M. Buehner, "Detection of first-year and multi-year sea ice from dual-polarization SAR images under cold conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9109–9123, Nov. 2019.

[13] E. Rignot and M. R. Drinkwater, "Winter sea-ice mapping from multi-parameter synthetic-aperture radar data," *J. Glaciol.*, vol. 40, no. 134, pp. 31–45, 1994.

[14] M. S. Mahmud, V. Nandan, S. E. L. Howell, T. Geldsetzer, and J. Yackel, "Seasonal evolution of L-band SAR backscatter over land-fast Arctic sea ice," *Remote Sens. Environ.*, vol. 251, Dec. 2020, Art. no. 112049. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425720304193

[15] A. M. Johansson, C. Brekke, G. Spreen, and J. A. King, "X-, C-, and L-band SAR signatures of newly formed sea ice in Arctic leads during winter and spring," *Remote Sens. Environ.*, vol. 204, pp. 162–180, Jan. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425717304960

[16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*.

[17] S. Ochilov and D. A. Clausi, "Operational SAR sea-ice image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4397–4408, Nov. 2012.

[18] Y. Hua, D. Marcos, L. Mou, X. X. Zhu, and D. Tuia, "Semantic segmentation of remote sensing images with sparse annotations," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 8006305.

[19] C. L. V. Cooke and K. A. Scott, "Estimating sea ice concentration from SAR: Training convolutional neural networks with passive microwave data," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4735–4747, Jul. 2019.

[20] 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," *Cryosphere*, vol. 14, no. 8, pp. 2629–2645, Aug. 2020.[Online]. Available: https://tc.copernicus.org/articles/14/2629/2020/

[21] M. Nicolaus *et al.*, "Overview of the MOSAiC expedition-snow and sea ice," *Elementa, Sci. Anthropocene*, Dec. 2021.

[22] K. Kortum, S. Singha, G. Spreen, and S. Hendricks, "Automating sea ice characterisation from X-band SAR with co-located airborne laser scanner data obtained during the mosaic expedition," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2021, pp. 2624–2627.

[23] T. Geldsetzer and J. J. Yackel, "Sea ice type and open water discrimination using dual co-polarized C-band SAR," *Can. J. Remote Sens.*, vol. 35, no. 1, pp. 73–84, 2009.

[24] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[25] M. S. Mahmud, T. Geldsetzer, S. E. L. Howell, J. J. Yackel, V. Nandan, and R. K. Scharien, "Incidence angle dependence of HH-polarized C- and L-band wintertime backscatter over Arctic sea ice," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6686–6698, Nov. 2018.

[26] J. Lohse, A. P. Doulgeris, and W. Dierking, "Incident angle dependence of Sentinel-1 texture features for sea ice classification," *Remote Sens.*, vol. 13, no. 4, p. 552, Feb. 2021.

[27] N. Asadi, K. A. Scott, A. S. Komarov, M. Buehner, and D. A. Clausi, "Evaluation of a neural network with uncertainty for detection of ice and water in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 247–259, Jan. 2021.

[28] M. Abadi. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. [Online]. Available: https://www.tensorflow.org/

[29] L. Beyer. (2015). *Pydensecrf*. [Online]. Available: https://github.com/lucasb-eyer/pydensecrf

[30] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFs with Gaussian edge potentials," 2012, *arXiv:1210.5644*.

[31] S. Hendricks. (2019). *ICE Drift—Transformation of GPS Positions Into a Translating and Rotating Coordinate Reference System*. [Online]. Available: https://gitlab.awi.de/floenavi-crs/icedrift

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[33] U. Nixdorf *et al.* (Sep. 2021). *MOSAiC Extended Acknowledgement*, doi: 10.5281/zenodo.5541624.

Karl Kortum (Graduate Student Member, IEEE) received the master's degree in physics from the University of Hamburg, Hamburg, Germany, in 2019. He is currently pursuing the Ph.D. degree with the German Aerospace Center (DLR), Bremen, Germany, and the University of Bremen, Bremen.

His research focuses on the development of artificial intelligence-based algorithms to automate sea ice property retrieval from synthetic aperture radar and additional fused measurements.

Suman Singha (Member, IEEE) received the M.Tech. degree in remote sensing from IIT Roorkee, Roorkee, India, in 2009, and the M.Sc. degree in remote sensing and the Ph.D. degree in microwave remote sensing from the University of Hull, Hull, U.K., in 2010 and 2014, respectively.

He was a Visiting Scientist with the European Maritime Safety Agency (EMSA), Lisbon, Portugal, in 2012. Since 2013, he has been a Research Scientist with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Bremen, Germany. Since 2020, he has been an Adjunct Assistant Professor with the Department of Geography, University of Calgary, Calgary, AB, Canada. His research interest includes the application of artificial intelligence and machine learning approaches to classify traditional and polarimetric synthetic aperture radar (SAR) images with an emphasis on sea ice properties retrieval and oil spill detection.

Gunnar Spreen (Member, IEEE) received the M.Sc. degree in physics (Diplomphysiker) and the Ph.D. degree in oceanography from the University of Hamburg, Hamburg, Germany, in 2004 and 2008, respectively.

He works on satellite remote sensing of Polar Regions with a focus on monitoring changes of sea ice (extent, mass, and dynamics) and on understanding underlying climate processes. Satellite measurements are validated using ground-based and airborne field observations, recently in particular from the 2019/2020 Multidisciplinary Drifting Observatory for the Study of the Arctic Climate (MOSAiC) expedition, for which he serves on the project board and coordinates the remote sensing activities. He was a Research Scientist with the Norwegian Polar Institute, Tromsø, Norway, and a Post-Doctoral Scholar with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. He is currently the Head of the Research Group "Remote Sensing of Polar Regions," Institute of Environmental Physics, University of Bremen, Bremen, Germany. His work addresses the development of new retrievals for sea ice parameters, such as leads and ice type from synthetic aperture radar (SAR) and microwave radiometer data as well as snow on sea ice.