

Robust Multi-Seasonal Ice Classification from High-Resolution X-Band SAR

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

Abstract—Automated solutions for sea ice type classification from synthetic aperture (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 which 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 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% .

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 by pixel, whilst 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 look up 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], [6], [7], [8] or unsupervised segmentation, with manual segment selection [9], [10] to convolutional neural network 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. 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 downwards 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 to autonomous classification, 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 convolutional neural network (CNN) is especially helpful, because the neural network can learn to relate all the information in the window to only the centre pixel one 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 SAR have historically been preferred for sea ice classification. Not only is there greater coverage, with large satellite missions such as Sentinel-1 and Radarsat, longer wavelengths also offer bigger penetration depths [15] and make it easier to discriminate between ice classes from backscatter and texture features alone. 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 pixel-wise at high accuracy by using context windows at various zoom levels. In a post-processing 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 post processing step with sparse labels in image processing [18].

Karl Kortum is with the Maritime Safety and Security Laboratory, Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 28359 Bremen, Germany, and also with the Institute of Environmental Physics, University of Bremen, 28359 Bremen, Germany (e-mail: karl.kortum@dlr.de).

Suman Singha is with the Maritime Safety and Security Laboratory, Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 28359 Bremen, Germany, and also with the Department of Geography, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: suman.singha@dlr.de).

Gunnar Spreen is with the Institute of Environmental Physics, University of Bremen, 28359 Bremen, Germany (e-mail: gunnar.spreen@uni-bremen.de).

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 labelled 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 labelled inconsistently in different scenes. This problem is discussed in [11], where a 50x50 pixel patch of Sentinel 1 imagery is classified using a CNN. At such patch size the effects of the rough labelling are still manageable, but when classifying smaller patches, it becomes increasingly impactful. Classification at high resolution thus relies on manually labelled data. Here, however, due to the tediousness of the labelling 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 demonstrate 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 most 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, 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, it also needs to be predicted, due to the delay of the ordering and the capturing of a scene. Over the course of 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.

Whilst 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 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 paper in further research with high resolution airborne measurements [22].

II. DATA

The training and test data sets used in this paper 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, whilst performance on the validation data set is used for to stop the training in time to prevent overfitting. The data used for robustness analysis is 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 polarisations, respectively. The images have a row and column spacing of 3.5 m, and are typically around 16000 pixels by 4000 pixels in size. This corresponds to 56 km by 14 km.

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

TABLE I: Table showing class definitions and labels.

Labelling 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 labelled, 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 colour coding used can be found there as well.

All scenes in the robustness analysis data set entirely contain the immediate area around Polarstern (a 3 km by 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 in the beginning of 2020. In figure 2a, we see that in this time the SAR images were consistently taken outside of full performance range, which is between 20° to 45° for StripMap images. The SAR measurements for

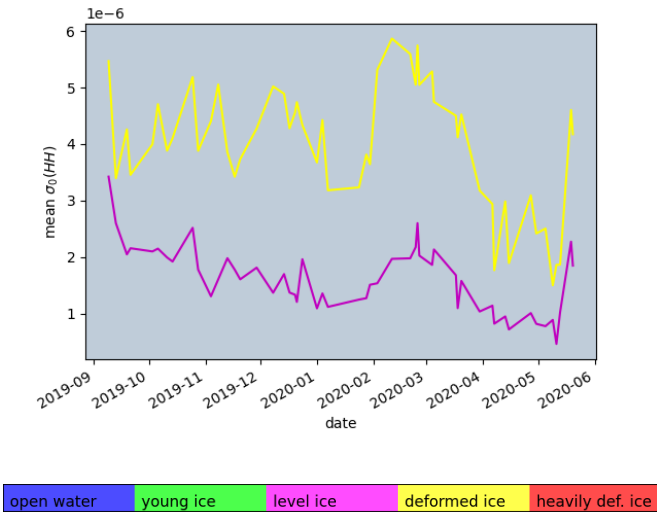


Fig. 1: Chart showing the mean sigma nought calibrated backscatter for the HH band, for level ice (LI) and deformed ice (DI) from October through May. The data is extracted from 52 manually labelled TerraSAR-X scenes.

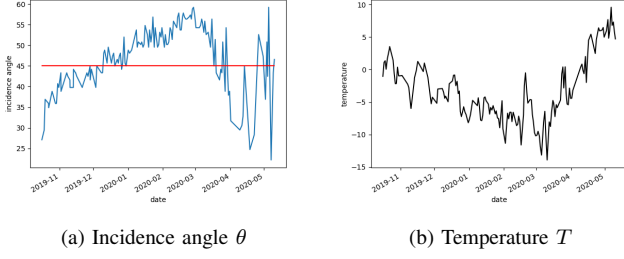


Fig. 2: a) Average incidence angle the scenes were acquired at. The red line at 45° indicates the limit of the full performance range of the satellite sensor. Figure b) shows the mean ambient temperature measured on the Polarstern vessel over the time period.

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. 2b). 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 to the general approach and then the individual parts are described in detail. The core of the classifier is a convolutional neural network (CNN). For added robustness a discriminator and a dense conditional random field (DCRF) are used as further processing. The algorithm assigns one of five classes (tab. I) to a given 5×5 pixel patch of a SAR imagery.

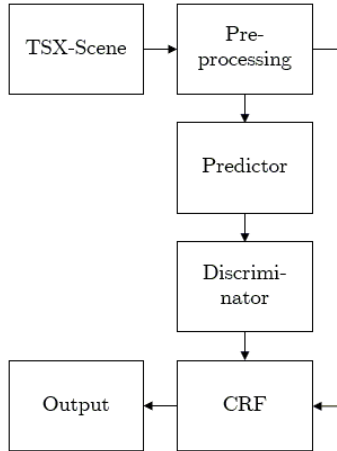


Fig. 3: Flowchart showing the pipeline for the proposed ice classification algorithm.

Figure 3 depicts the classification pipeline for the algorithm used. After pre-processing, features of varying scope and resolution (zoom levels) are supplied along with each 5×5 image slice that is to be classified (tab. II). A CNN is fed these features and makes an initial prediction for that patch (see figure 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 dense conditional random field smooths over the labels by relating spatial context of the labelled data and the underlying image. This also fills the missing values left by the discrimination step.

1) *Pre-Processing*: In the initial step of data pre-processing, the original dual polarised SAR scene is calibrated to slant range (β_0) and a false colour composition of the data is constructed. The composition consists of four channels: HH, VV, HH-VV, HH/VV. The difference and ratio are common for manual ice charting and visualisation of SAR scenes, as they promote contrast across ice types and open water. Additionally 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 good contrast. Whilst 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 Convolutional Neural Network (CNN) (see figure 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, that is taken from the SAR scene rescaled by a factor of five. 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 is more likely to occur with well-defined edges, like the edge of a multi year 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 it 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 is 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 the general brightness of the scene, interfaces between ice masses or (not important for this data set) 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 (eg. melt

onset) or low radar response (eg. 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 centre 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 64x64 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 data set 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 $\approx 44 \times 4 = 176$. Herein lies a substantial risk that the algorithm overfits to the training data. It might memorise 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 normalisation 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 [25] and [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 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 parameters of our network are largely based on heuristics and our own 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 maxpooling layers - in tests it seems the network lost a little information upon in the maxpool layer that was still useful for classification. We apply LeakyRelu as an activation function, which introduces some necessary non-linearity and does not suffer from the problem of vanishing gradients. The latter property is especially useful for deeper networks. We used strong regularisation with multiple dropout layers with a dropout rate of 0.3, as our data is still quite sparse in contrast to the scope of possible backscatter signatures from sea ice and thus overfitting is a concern. Additionally we opted for small spatial dimensions after convolutions (before flattening) and lower number of neurons to force the network to parametrise the input features - which lead to better extrapolation to unseen data. Batch normalisation in the early layers slightly sped up

the convergence of our network. The adam optimizer was used to update weights during training.

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	$x_{global}, y_{global}, \theta_{inc}, t$

TABLE II: Table showing 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, HH/VV). The coordinates x_{global}, y_{global} are the normalised pixel coordinates of the region to be classified in the rescaled global patch.

The network uses a categorical crossentropy loss appended by an additional term from the FESTA loss [18], specifically the distance of the softmax outputs. The additional term encourages separation of labels independent of correct classification, which helped with the convergence of our classifier.

Additionally, 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 labelling 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 the manual labelling is most error prone. In our case distinguishing deformed and level ice benefitted most of this treatment, because it is partially non-local 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 normalised.

	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

TABLE III: Table shows the label smoothing parameters used. $U_{[a,b]}$ denotes a random sample from a uniform distribution from the interval $[a, b]$. The label vectors were normalised after random sampling.

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

3) *Discriminator*: The discriminator model has near identical structure to the classifier (fig. 11), except for the additional

input layer containing a proposed label and the output being one dimensional. It's task is to check if the proposed label is correct or not. This binary classification is fundamentally easier than predicting one of five classes and can correct for some systematic errors the classifier makes. The discriminator is trained on randomly mislabelled data as ground truth for mislabelled patches, which performed better than a discriminator trained specifically on the correctly labelled and mislabelled 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 open water and thin ice misclassifications.

4) *Conditional Random Field*: The pixels deemed to be wrongly classified by the discriminator are removed from the classified product. Then a dense conditional random field is applied, that has a bilateral kernel next to the unary potential. This fills the missing values as well as clearing up some noise-like mislabels, like single pixels classified differently than all their surrounding pixels. The application of the dense conditional random field (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 penalised by the energy function. In this case, the features consist of a vector of colour intensities $\vec{I} = (I_r, I_g, I_b)$ across the RGB channels of a colour composite image as well as the position of the pixel \vec{P} . The RGB channels in the colour 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 modelled by the softmax output of our classifier. The bilateral term consists of weighted differences in position and colour. 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 $\|\dots\|$ 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 the introduction, we have a large number of scenes to test the robustness of our classifier across 8 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 which 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 by 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. Rapid change in ice type is constrained to the open water and thin ice classes. Change due to shifting of the floe is easily spotted by looking at the individual images and thus can be taken into account 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 ships GPS information, we can correct for the drift of the ice using a coordinate transformation to local ship coordinates [31] and identify 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 neighbours 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

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

The regions of ice we use to test this are the pixels in the stabilised images, such as seen in figure (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 stabilisation 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 the robustness.

IV. RESULTS

The classifier was trained on 44 scenes and tested on 8, which do not contribute to training data. The 8 scenes that make up the test set are randomly selected scene from each month of October through May. We will be looking at the performance of the classifier across the two datasets. As a comparison we also use a simple VGG16 [32] inspired architecture as an alternative classifier (see figure 12 in the appendix for details), that classifies based on only the superlocal 16 x 16 input data. The classifiers' performances are shown in tables IV, V for the training set and tables VI, VII for the test set.

	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 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 open water (OW), thin ice (TI), level ice (LI), deformed ice (DI) and heavily deformed ice (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%

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 open water (OW), thin ice (TI), level ice (LI), deformed ice (DI) and heavily deformed ice (HDI).

The two classes on which our and the VGG16 style classifier perform worst are open water and thin ice. The classification across the other three ice classes are more accurate, particularly for our classifier. Heavily deformed ice stands out as being especially easy to classify accurately. Open water and young ice examples are scarcest in the training and test scenes. Despite balancing the amount of samples per class upon training, this means the data set is a lot less diverse for these samples. In other words the number of regions with open water or thin ice is significantly lower than that of the other

	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 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 open water (OW), thin ice (TI), level ice (LI), deformed ice (DI) and heavily deformed ice (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%

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 open water (OW), thin ice (TI), level ice (LI), deformed ice (DI) and heavily deformed ice (HDI).

classes. This can be seen (albeit at a smaller scale) in figure 5.

To illustrate the entire process of assessing our classifier, we give an example with two consecutive scenes from the 13th and the 14th of December 2019 in figure 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 (II). This slice is then labelled using the classifier and checked by the discriminator. The results of this step are shown in figure 4a and 4b. 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 figure 4c and 4d. 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 figure 4e and 4f.

By executing the procedure illustrated in figure 4 for all scenes from October through May, we have the data necessary to perform quantitative robustness analysis. Figure 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 deformed ice. Spikes of open water and thin ice are generally tied to some ice dynamics. To gain some additional insight into the variance of these classes, we compute the relative standard deviation of the ice type fraction for every scene and its four nearest neighbours (fig. 6). When interpreting this as a deviation of the classification, we implicitly assume the real ice type distribution to be stable over five neighbouring 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 analysed 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 full performance range. Whilst 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 rationalised with the increased incidence angle during this time period (see fig.2a). It is also apparent that this cause of error plays a role in the increased uncertainties observed during 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 (7a) one can see some freshly frozen over leads with thin ice cover and in the later scene (7b), the leads have closed up again and all signs of young ice have disappeared.

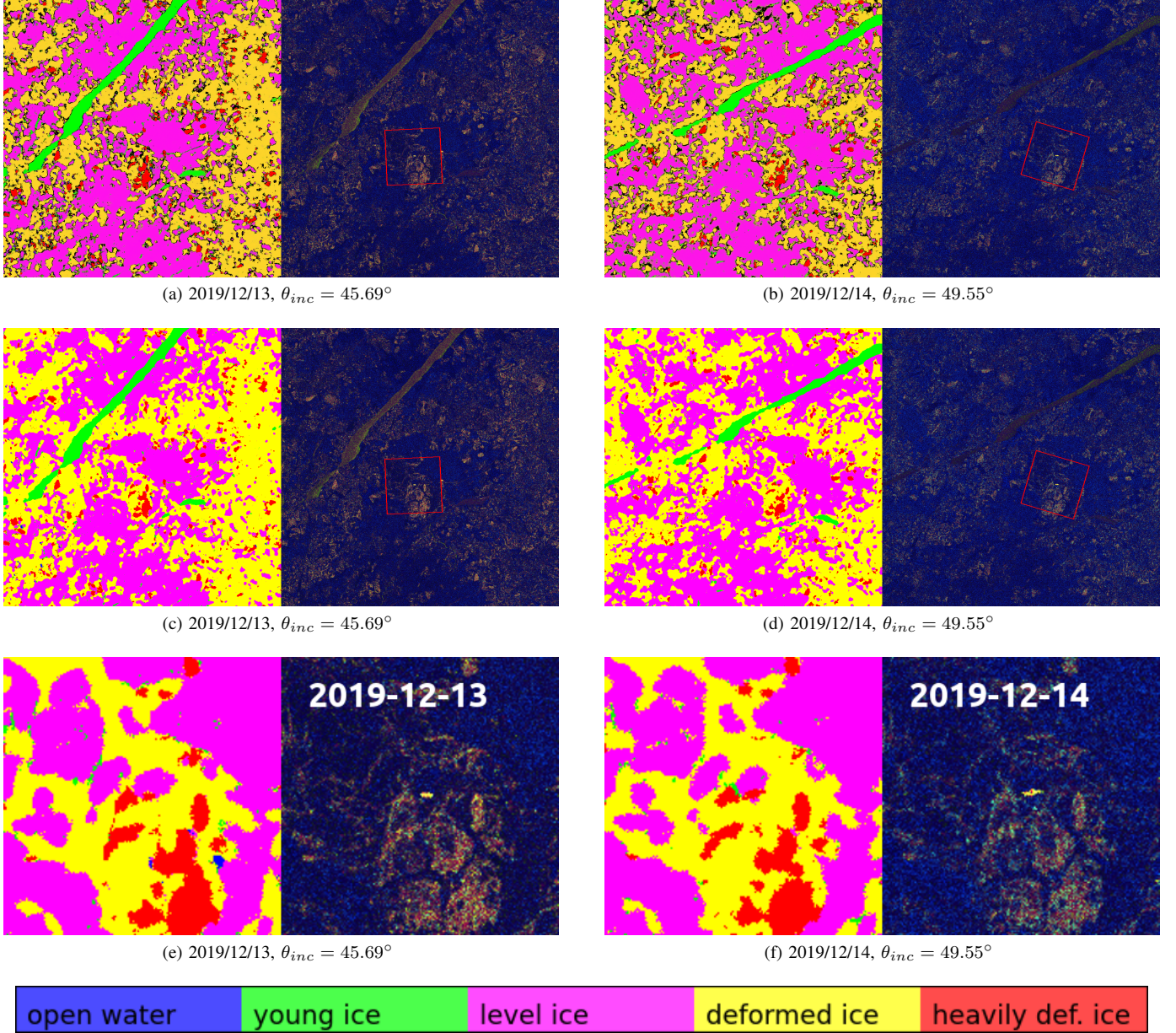


Fig. 4: Illustration of robustness analysis for two consecutive days in December 2019. Pictures (a, b) show the subscenes after classification and discrimination. In (c, d) the dense conditional random field has been applied and (e, f) show the scenes cropped and rotated to the region of interest. The false colour compositions are using (VV, HH-VV, HH/VV) in the (R, G, B) channels respectively. The top two rows additionally show the area used for robustness analysis, that is cropped to in the final row.

These pictures document the most drastic of these events, were 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 (7c, 7d) are both taken at very high incidence angles, far above full the performance range of 20–45 degrees. It is evident from the images, that the signal is significantly weaker. At this angle we see an overestimation of the deformed ice class in the second image. This is especially notable in the top left quarter of the patch, which was dominated by a level ice surface, but is classified as almost entirely deformed ice

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 (7e, 7f) 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 larger area, but the transitional areas between classes are seemingly classified less robustly (see the extend of the level ice on the left of the image

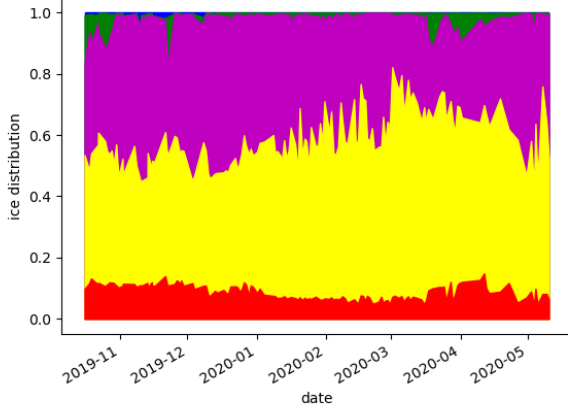


Fig. 5: Chart showing the predicted ice distribution in the 3 km by 3 km area surrounding the Polarstern vessel from October through May.

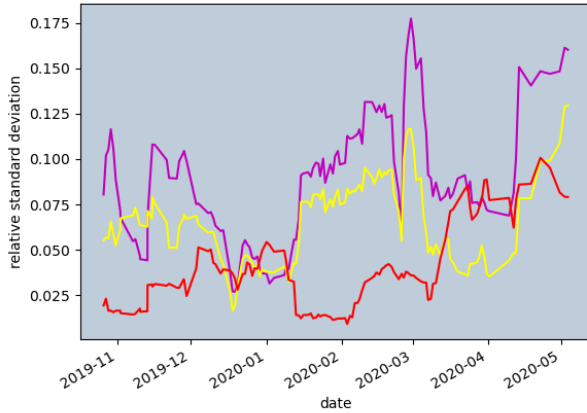


Fig. 6: Relative standard deviation of the distribution of level, deformed and heavily deformed ice at each date and the four nearest scenes.

7e versus 7f)). This effect is particularly evident at high incidence angles.

In figure 8, we can observe the development of robust classification across the analysed 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 figure 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. 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.

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

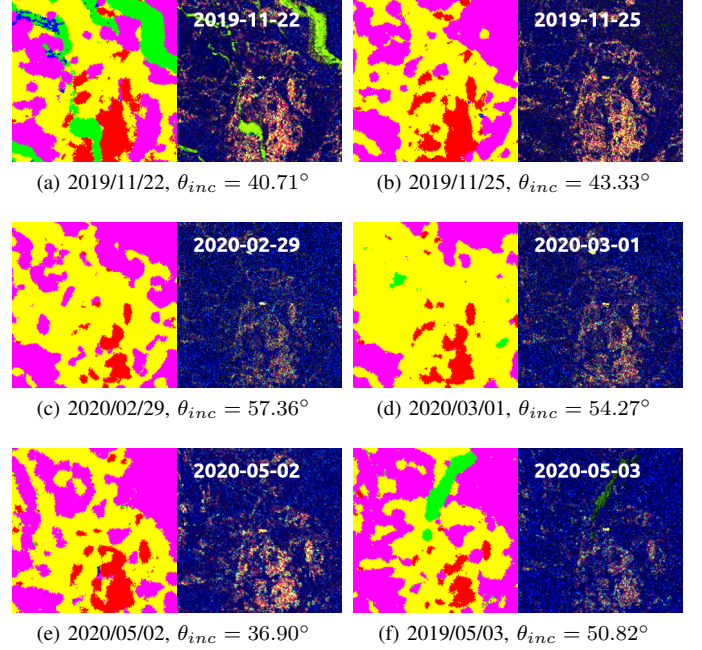


Fig. 7: Pairs of classified scenes from time intervals with low robustness. The left parts show the cropped classified image and the right parts are false colour compositions.

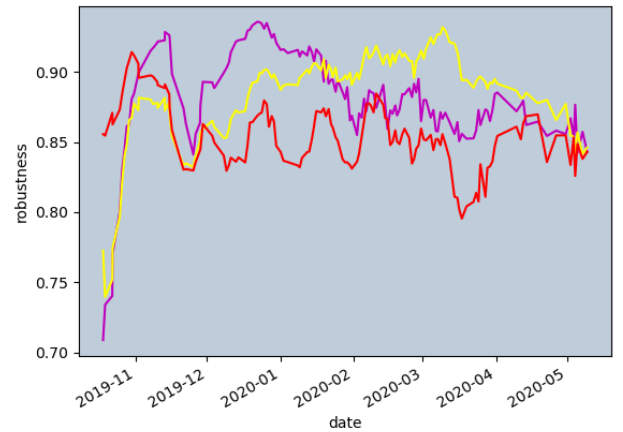


Fig. 8: Chart showing a moving average of the probability $P_r(c)$ of robust classification for three ice classes: level ice, deformed ice, heavily deformed ice. We compute $P_r(c)$ as the percentage of robustly classified pixels of class c per scene.

	LI	DI	HDI
P_r	88.34%	88.62%	85.34%

TABLE VIII: Table showing the average probability $P_r(c)$ of robust classification for LI, DI and HDI across the entire dataset

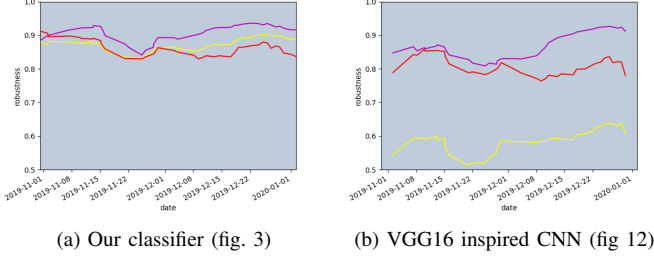


Fig. 9: Charts showing a moving average of the probability $P_r(c)$ of robust classification for three ice classes LI, DI, HDI, for our classifier and a VGG16 inspired model.

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

TABLE IX: Table showing the bounds of robust and correct classification for LI, DI and HDI of the classifier.

With classification probabilities P_c from table VII and the lower bounds of P_r from VIII, we can compute the bounds of probabilities of robust and correct classification P_{rc} for the three solid ice classes. Results are shown in table IX.

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 data set 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 data set, the dynamics during initial freeze-over have a great effects 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, it becomes very similar to heavily deformed ice 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 open water classification also proved to be a challenge for this data set. Traditionally the polarisation 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 labelling 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 labelling 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 labelling classes which are discernible with certainty - it is rather that the choice of labelled regions is already filtered by a human selection process. For example there is a tendency not to label a region with small area as it would make the labelling process very tedious. This translates to the classifier which struggles with smaller regions of one class, often wrongfully mislabelling them to be the same as the surrounding ice class. Additionally 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, whilst pixels in the center of same-ice regions appear robust. This bias could be eliminated by deriving ground truth data from in-situ measurements.

The discrimination of deformed and level ice relies on non local features and hence suffers most from the above mentioned boundary problem. Deformed ice is not always identified by a higher brightness and lower polarisation 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 deformed ice are not boundaries but a continuum. Hence it is difficult to define a hard boundary when labelling data manually. Generally the rule when labelling 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 data set, which culminates in misclassification in the transitional areas of deformed and level ice 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, tab. 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 take into account 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 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 open water 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: level, deformed and heavily deformed ice. Open water and thin ice 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 robustness analysis for these two ice types. We could also identify regions of increased classification inaccuracy and lack of robustness, that coincide with shortcomings of a manual labelling process. Already now our ice type dataset can provide helpful information for upscaling other MOSAiC in-situ data to a regional context, like sea ice physical and chemical properties or ecological samples, which all vary by ice type.

As was mentioned in the introduction, 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 data set.

ACKNOWLEDGMENT

This study was funded by Deutsche Forschungsgemeinschaft (DFG) under project name 'MOSAICmicrowaveRS' (SI 2564/1-1 and SP 1128/8-1).

Data used in this manuscript was produced as part of the international Multidisciplinary drifting Observatory for the Study of the Arctic Climate (MOSAIC) with the tag MOSAiC20192020 and Project_ID: AWI_PS122_00..

We thank all persons involved in the expedition of the Research Vessel Polarstern during MOSAiC in 2019-2020 as listed in Nixdorf et al. (2021) [33]. TerraSAR-X images used in this study were acquired using the TerraSAR-X AO OCE3562_4 (PI: Suman Singha).

REFERENCES

- [1] M. Fily and D. A. Rothrock, "Extracting sea ice data from satellite SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-24, no. 6, pp. 849–854, 1986.
- [2] R. Kwok, E. Rignot, B. Holt, and R. Onstott, "Identification of sea ice types in spaceborne synthetic aperture radar data," *Journal of Geophysical Research: Oceans*, vol. 97, no. C2, 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 Transactions on Geoscience and Remote Sensing*, vol. 37, pp. 780 – 795, 04 1999.
- [4] Y. Hara, R. Atkins, R. Shin, J. A. Kong, S. Yueh, and R. Kwok, "Application of neural networks for sea ice classification in polarimetric SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 3, pp. 740–748, 1995.
- [5] R. Ressel, A. Frost, and S. Lehner, "A neural network-based classification for sea ice types on X-band SAR images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, pp. 1–9, 07 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 Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 3715–3734, 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 7, pp. 3131–3143, 2016.
- [8] D. Murashkin, G. Spreen, M. Huntemann, and W. Dierking, "Method for detection of leads from Sentinel-1 SAR images," *Annals of Glaciology*, vol. 59, pp. 1–13, 03 2018.
- [9] A. M. Johansson, E. Malnes, S. Gerland, A. Cristea, A. Doulgeris, D. Divine, O. Pavlova, and T. R. Lauknes, "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," *Annals of Glaciology*, vol. 61, pp. 1–11, 01 2020.
- [10] A. P. Doulgeris, "An automatic Γ -distribution and markov random field segmentation algorithm for PolSAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 1819–1827, 2015.
- [11] H. Boulze, A. Korosov, and J. Brajard, "Classification of sea ice types in Sentinel-1 SAR data using convolutional neural networks," *Remote Sensing*, vol. 12, 07 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 Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9109–9123, 2019.
- [13] E. Rignot and M. R. Drinkwater, "Winter sea-ice mapping from multi-parameter synthetic-aperture radar data," *Journal of Glaciology*, vol. 40, no. 134, p. 31–45, 1994.
- [14] M. S. Mahmud, V. Nandan, S. E. Howell, T. Geldsetzer, and J. Yackel, "Seasonal evolution of L-band SAR backscatter over landfast Arctic sea ice," *Remote Sensing of Environment*, vol. 251, p. 112049, 2020. [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 Sensing of Environment*, vol. 204, pp. 162–180, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425717304960>
- [16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," *CoRR*, arXiv, 12 2014.
- [17] S. Ochilov and D. A. Clausi, "Operational SAR sea-ice image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4397–4408, 2012.
- [18] Y. Hua, D. Marcos, L. Mou, X. X. Zhu, and D. Tuia, "Semantic segmentation of remote sensing images with sparse annotations," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2021.
- [19] C. L. V. Cooke and K. A. Scott, "Estimating sea ice concentration from SAR: Training convolutional neural networks with passive microwave data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4735–4747, 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," *The Cryosphere*, vol. 14, no. 8, pp. 2629–2645, 2020. [Online]. Available: <https://tc.copernicus.org/articles/14/2629/2020/>
- [21] M. Nicolaus, D. K. Perovich, M. A. G. G. Spreen, L. von Albedyll, M. Angelopoulos, P. Anhaus, S. Arndt, H. J. Belter, V. Bessonov, G. Birnbaum, J. Brauchle, R. Calmer, E. Cardellach, B. Cheng, D. Clemens-Sewall, R. Dadic, E. Damm, G. de Boer, O. Demir, D. V. Divine, A. A. Fong, S. Fons, M. Frey, N. Fuchs, C. Gabarró, S. Gerland, H. F. Goessling, R. Gradinger, J. Haapala, C. Haas, J. Hamilton, H.-R. Hannula, S. Hendricks, A. Herber, C. Heuzé, M. Hoppmann, K. V. Høyland, M. Huntemann, J. K. Hutchings, B. Hwang, P. Itkin, H.-W. J. M. Jaggi, A. Jutila, L. Kaleschke, C. Katlein, N. Kolabutin, D. Krampe, S. S. Kristensen, T. Krumpen, N. Kurtz, A. Lampert, B. A. Lange, B. L. R. Lei, F. Linhardt, G. E. Liston, B. Loose, A. R. Macfarlane, M. Mahmud, I. O. Matero, A. M. S. Maus, R. Naderpour, V. Nandan, A. Niubom, M. Oggier, N. Oppelt, F. Pätzold, C. Perron, T. Petrovsky, R. Pirazzini, C. Polashenski, B. Rabe, I. A. Raphael,

- J. Regnery, M. Rex, R. Ricker, K. Riemann-Campe, J. R. A. Rinke, E. Salganik, R. K. Scharien, M. Schiller, M. Schneebeli, M. Semmling, E. Shimanchuk, M. D. Shupe, M. M. Smith, V. Smolyanitsky, V. Sokolov, T. Stanton, J. Stroeve, L. Thielke, A. Timofeeva, R. T. Tonboe, A. Tavri, M. Tsamados, D. N. Wagner, D. Watkins, M. Webster, and M. Wendisch, "Overview of the mosaic expedition - snow and sea ice," *Elementa: Science of the Anthropocene*, 12 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 *Geoscience and Remote Sensing Symposium, 2021 IEEE International*, 07 2021.
- [23] T. Geldsetzer and J. J. Yackel, "Sea ice type and open water discrimination using dual co-polarized C-band SAR," *Canadian Journal of Remote Sensing*, vol. 35, no. 1, pp. 73–84, 2009. [Online]. Available: <https://doi.org/10.5589/m08-075>
- [24] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, 07 2019. [Online]. Available: <https://doi.org/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 Transactions on Geoscience and Remote Sensing*, vol. 56, pp. 6686–6698, 2018.
- [26] J. Lohse, A. Doulgeris, and W. Dierking, "Incident angle dependence of sentinel-1 texture features for sea ice classification," *Remote Sensing*, vol. 13, 02 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 Transactions on Geoscience and Remote Sensing*, vol. 59, no. 1, pp. 247–259, 2021.
- [28] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: <https://www.tensorflow.org/>
- [29] L. Beyer, "Pydensecrf," 2015. [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. [Online]. Available: <https://arxiv.org/abs/1210.5644>
- [31] S. Hendricks, "Ice drift - transformation of GPS positions into a translating and rotating coordinate reference system," 2019. [Online]. Available: <https://gitlab.awi.de/floenavi-crs/icedrift>
- [32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [33] U. Nixdorf, K. Dethloff, M. Rex, M. Shupe, A. Sommerfeld, D. K. Perovich, M. Nicolaus, C. Heuzé, B. Rabe, B. Loose, E. Damm, R. Grading, A. Fong, W. Maslowski, A. Rinke, R. Kwok, G. Spreen, M. Wendisch, A. Herber, M. Hirsekorn, V. Mohaupt, S. Frickenhaus, A. Immerz, K. Weiss-Tuider, B. König, D. Mengedocht, J. Regnery, P. Gerchow, D. Ransby, T. Krumpfen, A. Morgenstern, C. Haas, T. Kanzow, F. R. Rack, V. Saitzev, V. Sokolov, A. Makarov, S. Schwarze, T. Wunderlich, K. Wurr, and A. Boetius, "Mosaic extended acknowledgement," 09 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.5541624>



Karl Kortum (Student Member, IEEE) received his Masters degree in physics in 2019 from the University of Hamburg, Germany and is currently pursuing his PhD at the German Aerospace Center (DLR), Bremen, Germany and the University of Bremen, Germany.

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 (M'08) received his M.Sc. degree in physics (Diplomphysiker) in 2004 and the Ph.D. degree in oceanography in 2008, both from the University of Hamburg, Hamburg, Germany.

He works on satellite remote sensing of Polar Regions with 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 MOSAiC expedition. For which he serves

on the project board and coordinates the remote sensing activities.

Gunnar Spreen is the head of the research group "Remote Sensing of Polar Regions" at the University of Bremen, Institute of Environmental Physics, Bremen, Germany. Before, he was a research scientist at the Norwegian Polar Institute, Tromsø, Norway and a postdoctoral scholar at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA.

Current work addresses the development of new retrievals for sea ice parameters like leads and ice type from SAR and microwave radiometer data as well as snow on sea ice.

APPENDIX NETWORK ARCHITECTURES

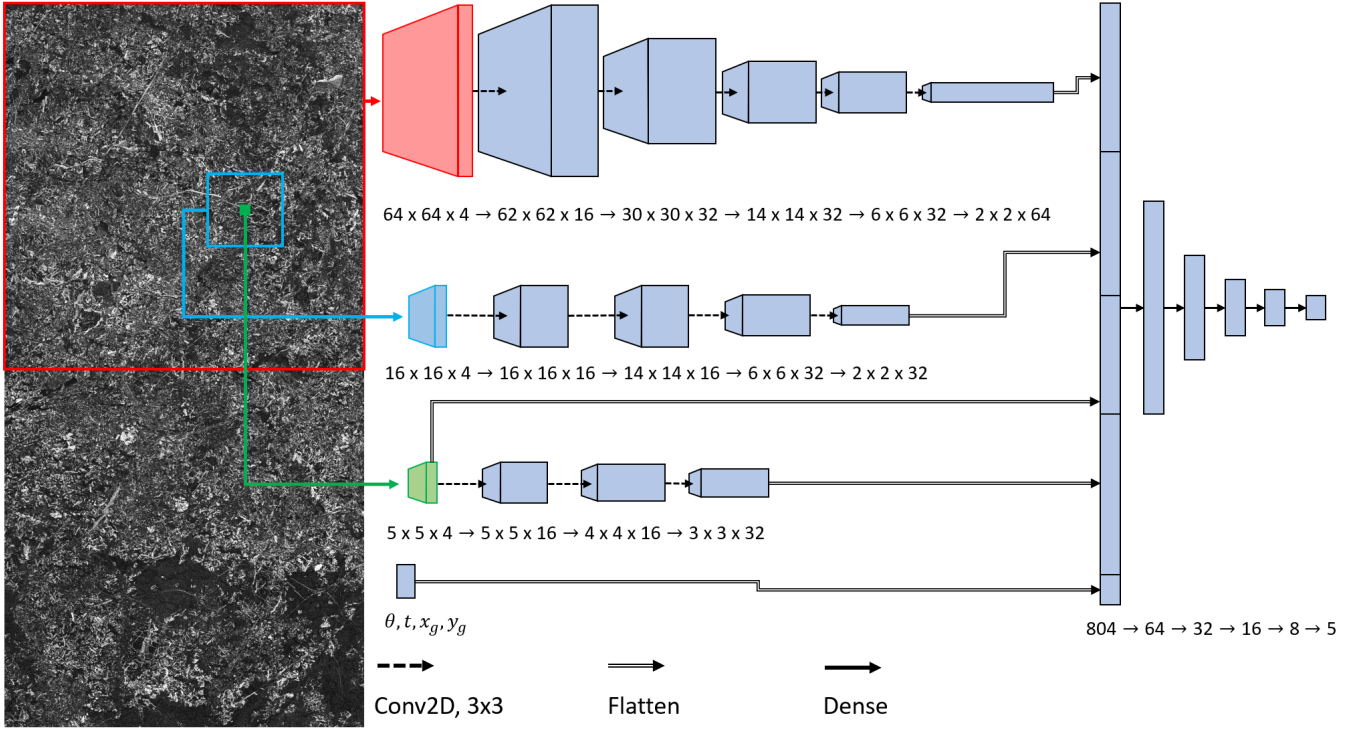


Fig. 10: Illustration of the CNN architecture used in the proposed classifier. Wherever the spatial dimensions in the convolutional blocks are downsampled (decrease by more than a factor of .5), a stride of 2 was used. Not included in the Image are the batch normalisation layers after the first convolutional layers for each input and the dropout layers used for regularisation 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 the acquisition time. Parameter count = 120421.

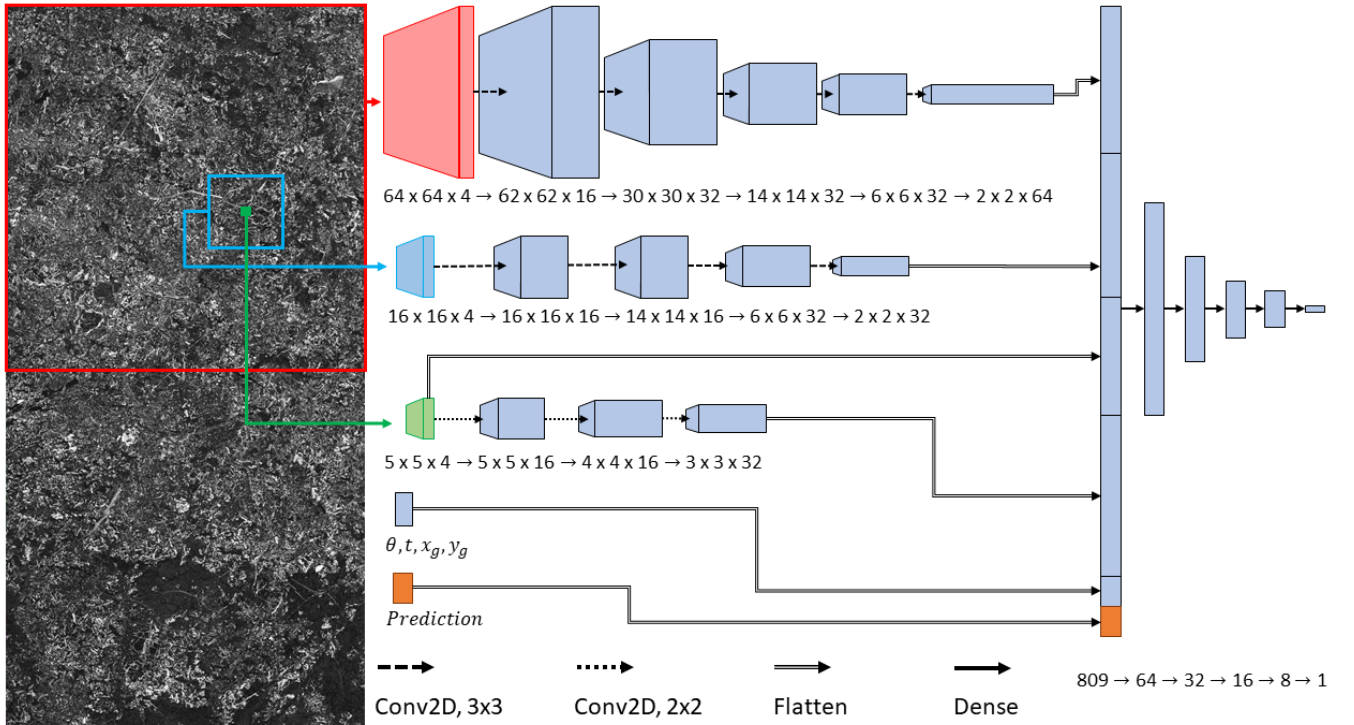


Fig. 11: Illustration of the CNN discriminator architecture used in the proposed classifier.

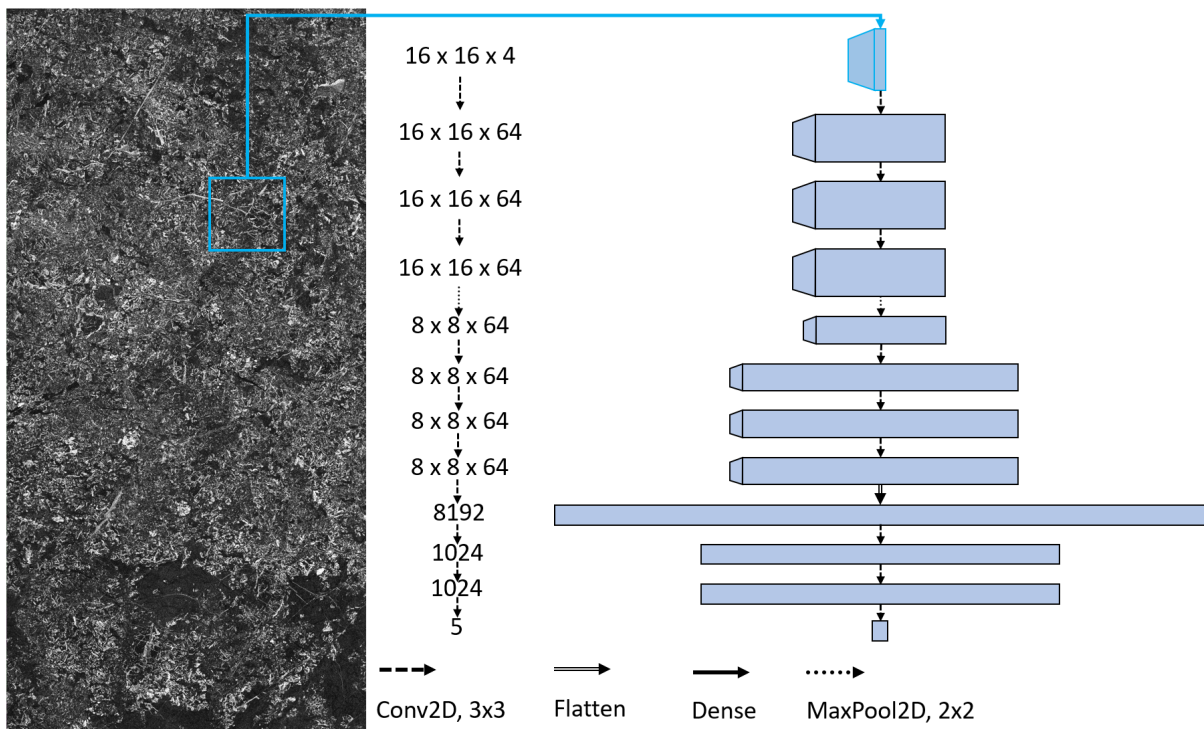


Fig. 12: Illustration of the VGG16 inspired network architecture used as comparison. Parameter count = 9889605.