SAR Ship–Iceberg Discrimination in Arctic Conditions Using Deep Learning

Peder Heiselberg *[ID], Kristian A. Sørensen [ID], Henning Heiselberg [ID] and Ole B. Andersen [ID]

Geodesy and Earth Observation, National Space Institute of Denmark, Technical University of Denmark, 2800 Kongens Lyngby, Denmark; kaaso@space.dtu.dk (K.A.S.); hh@space.dtu.dk (H.H.); oa@space.dtu.dk (O.B.A.)
* Correspondence: ph@space.dtu.dk; Tel.: +45-21389182

Abstract: Maritime surveillance of the Arctic region is of growing importance as shipping, fishing and tourism are increasing due to the sea ice retreat caused by global warming. Ships that do not identify themselves with a transponder system, so-called dark ships, pose a security risk. They can be detected by SAR satellites, which can monitor the vast Arctic region through clouds, day and night, with the caveat that the abundant icebergs in the Arctic cause false alarms. We collect and analyze 200 Sentinel-1 horizontally polarized SAR scenes from areas with high maritime traffic and from the Arctic region with a high density of icebergs. Ships and icebergs are detected using a continuous wavelet transform, which is optimized by correlating ships to known AIS positions. Globally, we are able to assign 72% of the AIS signals to a SAR ship and 32% of the SAR ships to an AIS signal. The ships are used to construct an annotated dataset of more than 9000 ships and ten times as many icebergs. The dataset is used for training several convolutional neural networks, and we propose a new network which achieves state of the art performance compared to previous ship–iceberg discrimination networks, reaching 93% validation accuracy. Furthermore, we collect a smaller test dataset consisting of 424 ships from 100 Arctic scenes which are correlated to AIS positions. This dataset constitutes an operational Arctic test scenario. We find these ships harder to classify with a lower test accuracy of 83%, because some of the ships sail near icebergs and ice floes, which confuses the classification algorithms.

Keywords: Synthetic Aperture Radar (SAR); convolutional neural network (CNN); deep learning; Automatic Identification System (AIS); ship detection; ship–iceberg discrimination; dark ships

1. Introduction

The Arctic landscape is rapidly changing in a way that can disrupt the global maritime traffic. Global warming has led to routes that historically have been covered by sea ice and navigable only part of the year now seeing increased traffic. In 2017, a Russian tanker sailed through the Arctic Ocean for the first time without the assistance of icebreakers [1]. With the Northwest Passage being the most direct shipping route between the Atlantic and Pacific Oceans [2], the Arctic requires improved monitoring of both ships and icebergs for safety and surveillance.

Larger ships must identify themselves by ship transponder systems, such as the Automatic Identification System (AIS). However, AIS receivers are especially sparse in the Arctic and open seas where messages can experience temporal gaps, be days old and in areas with high traffic signals are frequently lost in data collisions [3,4]. AIS transponders may also be turned off by accident or deliberately, and recently, nearly 100 warships were found to have faked their own AIS signal by spoofing [5]. Dark ships are such non-cooperative vessels that do not transmit AIS signals. These ships pose a risk for marine traffic safety and may be involved in criminal activities such as piracy, smuggling, oil spills, trespassing, and Illegal, Unreported and Unregulated (IUU) fishing. Dark ships

can be detected in Synthetic Aperture Radar (SAR) satellite imagery independent of AIS transmission, cloud coverage and time of day. However, the Arctic waters do not only contain ships but also abundant icebergs, and correct discrimination is therefore vital for identifying ships, including dark ships in the Arctic.

Ship detection in SAR images has been extensively studied [6–13], whereas the literature on discriminating ships from icebergs in SAR images is very sparse and relies on small datasets. Refs. [14,15] applied a CFAR detector to SAR images and extracted small ship and iceberg images centered on the detection. Ref. [14] surveyed 19 large supply vessels and 20 icebergs and achieved an accuracy of 97% using a polarimetric area ratio threshold between the HV and HH polarization of the detections. Ref. [15] visually labeled 76 icebergs and 125 ships from SAR detections and obtained a 93.5% discrimination accuracy with a maximum likelihood Gaussian classifier that relied on extracted features from the SAR images of the ships and iceberg such as the area and polarimetric ratios. Ref. [16] used more advanced supervised learning methods leading to an accuracy of 95%; however, this was for optical satellite images. In [17], a convolutional neural network (CNN) was trained on a small TerraSAR-X dataset with 277 ships and 68 icebergs (300 of each after augmentation), yielding a precision of 98%. The Statoil/C-CORE Iceberg Classifier Challenge [18] introduced a dataset of 1604 ship and iceberg images from Sentinel-1 which was employed for training a large number of CNNs (see e.g., [19–21]). However, as the preprocessing and normalization of the dataset is undocumented, we can not use or compare results to other real scenario data. In addition, the winners of the competition exploited an artificial grouping of incident angles of icebergs in order to overfit and achieve the highest accuracy.

The limited study of ship and iceberg discrimination in SAR images is due to the difficulty of constructing large datasets that is further hindered by the limited availability of SAR images with the same polarisation. The Sentinel-1 satellites provide freely available SAR imagery [22], and large ship datasets already exist, e.g., in [23], but these contain only ships and only in VV+VH polarization. Sentinel-1 images of the Arctic are predominantly recorded in HH+HV polarization, and transfer learning across polarizations is not possible, as the images are fundamentally different. Few ships sail in the Arctic, and therefore, it is difficult to construct a large dataset of ships of HH+HV polarized Sentinel-1 SAR images.

In this article, unlike previous studies, we aim at providing an accuracy estimate for ship and iceberg discrimination including ships sailing in the Arctic waters near icebergs and ice floes. To achieve this goal, we first describe the methods, provide a general framework for ship detection in SAR images using AIS and include an analysis of dark ships. Then, we create a large ship and iceberg dataset with the same HH+HV polarizations using 200 Sentinel-1 SAR scenes. Finally, we train and compare several CNNs, including proposing a new CNN model, and perform a test in an operational scenario of ships sailing in Arctic waters.

2. Data Acquisition

In this study, we analyze and combine two types of data. SAR images and AIS data. The Sentinel-1 satellites provide a wide selection of freely available SAR products. We analyze the dual polarization level-1 Ground Range Detected High resolution (GRDH) scenes in the Interferometric Wide Swath mode (IW). These scenes provide a high resolution, of 20×22 m with a 10 m pixel spacing, which is needed in order to detect smaller ships. The Sentinel-1 SAR operates with two dual polarizations with each scene containing a co-polarized and cross-polarized channel. The Arctic region is almost exclusively acquired in HH+HV, while the rest of the world is acquired in VV+VH. Consequently, icebergs are abundant in the HH+HV polarization, while ships are abundant in the VV+VH polarization. The scarcity of ships sailing at Arctic latitudes did not allow us to build a large annotated dataset of ships in the Arctic exclusively. However, in the first years of the Sentinel-1 mission, the satellites occasionally acquired HH+HV polarized scenes globally. Because of this, we have acquired 100 SAR scenes in the HH+HV polarization of the English Channel (see Figure 1), the Gibraltar Straight, the north sea, South Africa and other areas with dense

maritime traffic. For icebergs, we selected 100 scenes near the Disco Bay in Greenland (see Figure 2). It is home to the Ilulissat Icefjord, which is one of the fastest flowing glaciers in the world and thus a major iceberg producer. The scenes of Greenland were sampled in the spring to autumn period, where icebergs are prevalent and no sea ice is present. All scenes were retrieved via the ASF DAAC [24].
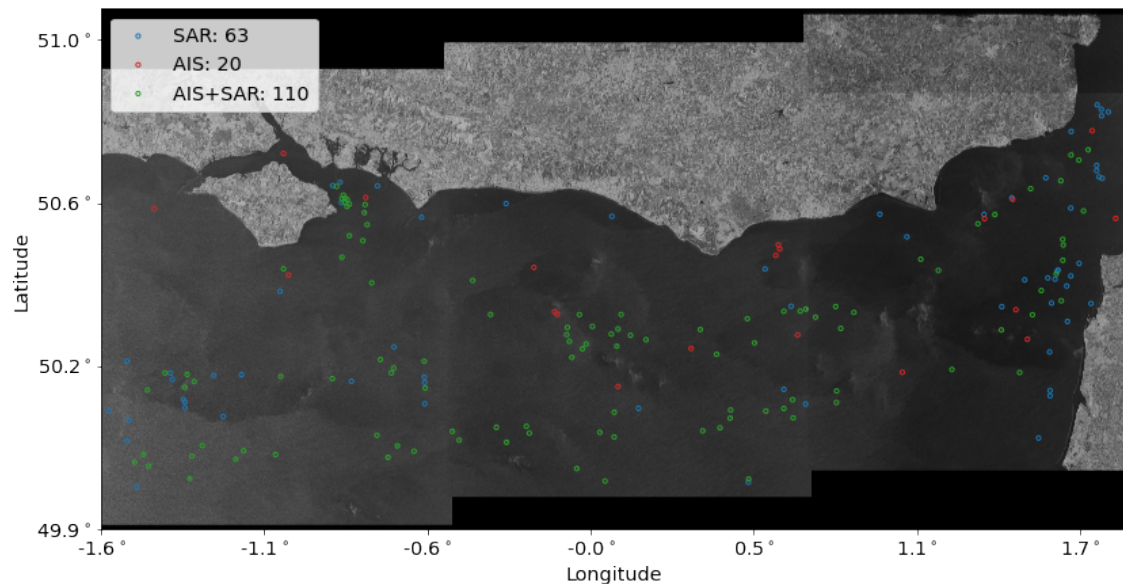

Figure 1. Sentinel-1 SAR image of the English Channel, where 85% of AIS signals are assigned to a SAR detection, while 64% of detections were assigned to an AIS signal. A majority of the ships with AIS are tankers and cargo ships that follow the main shipping route through the English Channel.
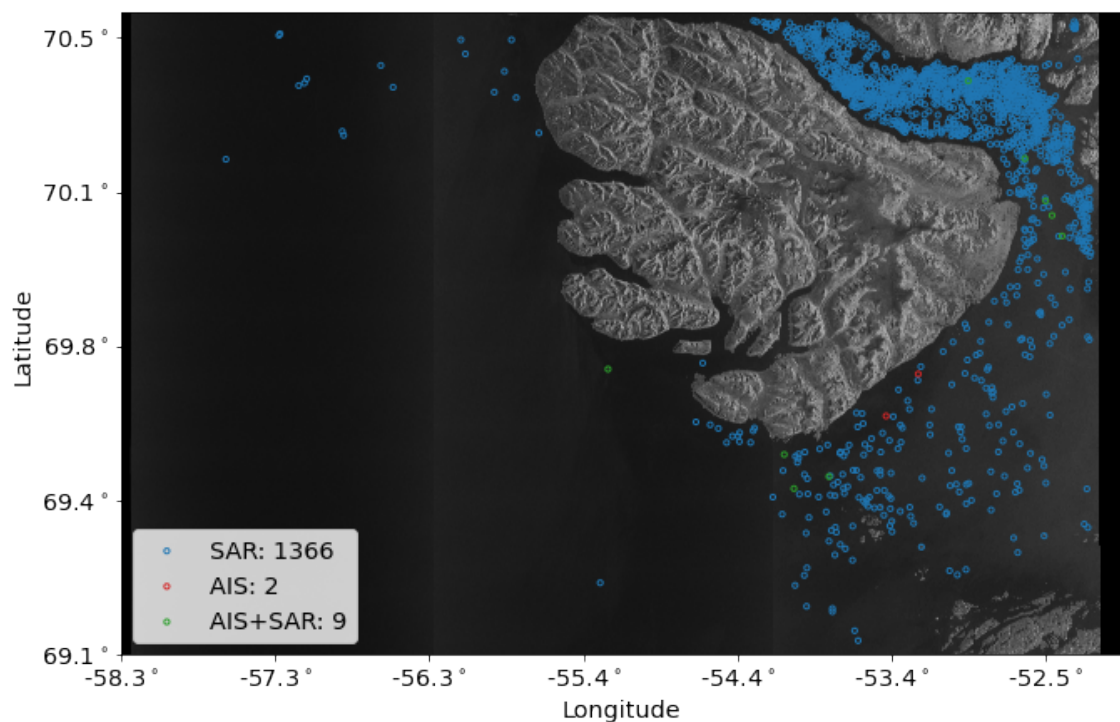
Figure 2. Sentinel-1 SAR image of the Disco Bay of Greenland, where 82% of AIS signals can be assigned (AIS+SAR) to a SAR detection. The Illulisat glacier is located to the east of this scene. The ocean currents of this area are north bound, which causes icebergs to bunch together in the narrow straight north of the Disco island.

AIS data were acquired corresponding to the spatial extent (footprint) of the SAR scenes and filtered in a temporal interval from two hours before and after the sensing time. The interval is chosen to guarantee ample amounts of AIS data on each side of the SAR scenes' recording time. AIS data points contain a Maritime Mobile Service Identity (MMSI) number, which is an identifier of the ship, a timestamp, latitude and longitude coordinates, as well as other information about the ship. We aggregate data with the same MMSI number to form a track of ship positions. The raw AIS data and tracks are shown in Figure 3.
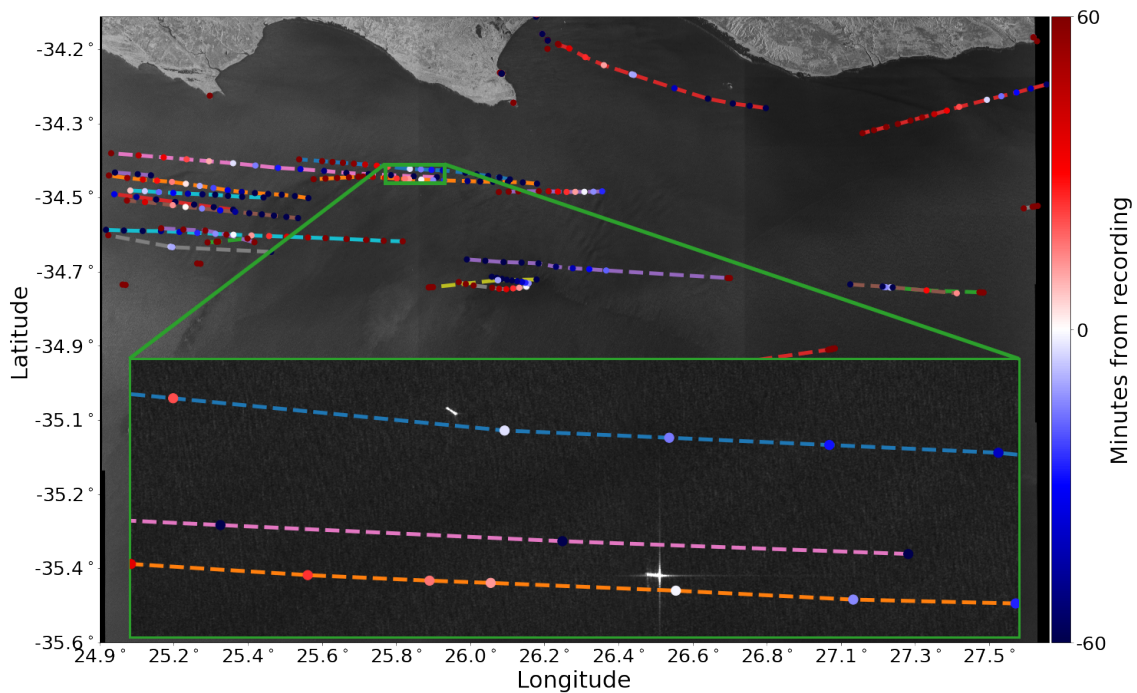
Figure 3. Sentinel-1 SAR image south of Port Elizabeth, South Africa. The image is overlaid with raw AIS data ranging in color from blue (before) to red (after) the recording time T_{SAR} of the SAR image. Each unique ship track is shown with a colored dashed line. Two vessels can be seen in the bottom insert displaced in the azimuth direction, which is corrected for by Equation (3).

3. Methods for Ship and Iceberg Detection

In this section, we present the processing steps applied to the SAR scenes and AIS data in order to create a dataset of ships and icebergs.

3.1. Land Masking and Geocoding

The Sentinel-1 SAR scenes include a metadata file which contains information such as recording times and information about the latitude, longitude, slant range time, and incidence angle of the satellite for all pixel in the scene. By cubic interpolation, it is then possible to translate from AIS latitude and longitude to pixel coordinates and vice versa. Through this geocoding, we used the GSHHG database [25] for land-masking with an additional 200 pixel padding corresponding to 2 km. This generous padding was necessary because of the poorly chartered rocky outcrops that are abundant in Greenland and breakwaters, which are common near harbors. As there are sufficient ships to build the dataset, it is not critical that we miss some near the coast due to the padding.

3.2. Detection Algorithm

Wavelet detectors are particularly robust to noise and allow for detection at multiple scales with low false alarm rates [6]. We therefore apply a two-dimensional continuous wavelet transform (CWT)-based detector implemented as described in [26]. For the wavelet mother function ψ, we choose the Mexican hat $\psi(\mathbf{r}) = (2 - \mathbf{r}^2)exp(-\mathbf{r}^2/2)$. This wavelet is

the double gradient of a Gaussian and as such automatically removes any constant and linearly increasing background. The scene pixel coordinates are $\mathbf{r} = (x, y)$ and CWT is the pixel intensity $s(\mathbf{r})$ folded with the wavelet scaled by a factor $a = 1, 1.5, 2, 2.5, \ldots 6$, and translated by a distance $\mathbf{b}$

$$CWT(a, \mathbf{b}) = \frac{1}{a} \int_{-\infty}^{\infty} s(\mathbf{r}) \psi((\mathbf{r} - \mathbf{b})/a) d\mathbf{r} \tag{1}$$

CWT will have a local maximum for values of $(a, \mathbf{b})$, where the wavelet matches a ship signal. As described in [26], *CWT* maxima can then be connected across a range of scale values a to form a ridge. The maximum *CWT* of the ridge is then located at the scale corresponding to the spatial size of the ship or iceberg. The lowest scale $a = 1$ is a measure of the noise level, which lets us define the signal to noise ratio as

$$SNR(a, \mathbf{b}) = CWT(a, \mathbf{b})/CWT(1, \mathbf{b}^*), \tag{2}$$

where $CWT(1, \mathbf{b}^*)$ is chosen as the 95% quantile of CWT at the noise level ($a = 1$) within the local wavelet window surrounding the peak at $\mathbf{b}$. The SNR is a measure of both the target's signal strength and size. Furthermore, we can measure the ridge length $L = 1, 2, \ldots$ as the number of *CWT* maxima in a ridge. By setting a threshold for L and SNR, we can detect ships and icebergs while removing many false alarms from noise in single pixels and clutter. An analysis of parameter selection is provided in Section 3.4. A special category of false alarms is sea turbines, which often are closely bunched in farms. These are easily removed in the few scenes they appear by requiring a mean distance of more than 600 pixels to the nearest 40 detections. This threshold is quick and effective, and due to the few scenes, we could verify the correct removal by eye. For a more detailed study of ship detection around sea turbine farms, we suggest using the maps available from the European Marine Observation and Data Network (EMODnet) [27].

3.3. AIS-SAR Data Temporal and Spatial Association

Ships move only a few pixels during the around 15 s recording time of the SAR scene. Consequently, the center time T_{SAR} is only a few seconds off on average. All AIS tracks were interpolated with a cubic spline as in [28,29] yielding a single AIS_{SAR} coordinate at T_{SAR}. The ship speed over ground (*SOG*) and course over ground (*COG*) are provided with the AIS messages, but as these are sometimes erroneous [3,4], we derived the *SOG* and *COG* directly from the interpolation. An azimuth offset appears due to the range velocity component of the ships (see Figure 3). This Doppler shift offset is (see e.g., [8]):

$$\Delta d_{azimuth} = \frac{SOG \cdot \cos(COG^*)}{v_{sat}} \cdot R_{sr} \cdot \sin(\theta), \tag{3}$$

where $v_{sat} = 7.4$ km/s is the satellite velocity, R_{sr} the slant range distance and θ is the angle of incidence. The COG^* is the COG relative to the satellite direction, which depends on the satellite inclination angle and the latitude of the AIS signal (see [30] for details). AIS coordinates outside the scene and inside the land-mask after azimuth correction were discarded. The AIS_{SAR} coordinates were then assigned to SAR detections based on the Euclidian distance between pixel coordinates. If multiple AIS_{SAR} datapoints were assigned to the same SAR detection or vice versa, only the pair with the lowest distance was paired. The distance between the detected ships and assigned AIS signals showed a Gaussian distribution with outliers. These outliers were removed by requiring that this distance should not exceed 3σ of the Gaussian part of the distance distribution corresponding to 30 pixels or 300 m in both azimuth and range.
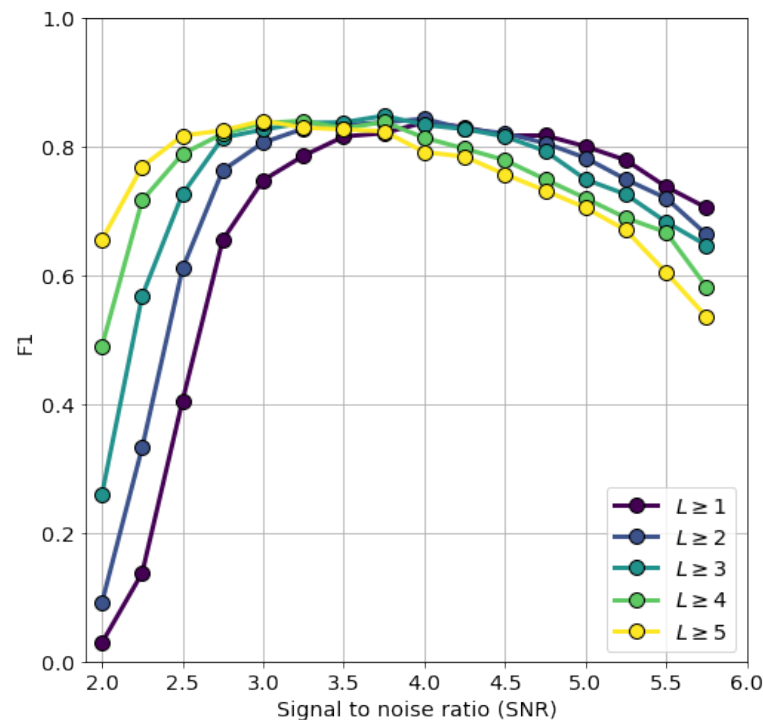
3.4. Hyperparameter Selection

The marine traffic in the English Channel consists of many large tankers and cargo vessels that commonly have frequent AIS messaging. However, there are also many smaller

fishing and pleasure ships with irregular or no AIS messaging. The English Channel is thus an ideal setting for hyperparameter tuning due to the variety of vessels and messaging. A standard method is to optimize the measure

$$F1 = \frac{2 \cdot R \cdot P}{R + P}, \tag{4}$$

which is the harmonic mean between precision $P = TP/(TP + FP)$ and recall $R = TP/(TP + FN)$. We can measure true positives (TP) as the number of detections assigned to an AIS signal, false negatives (FN) as the number of unassigned AIS signals, and false positives (FP) as the number of unassigned detections. However, this choice of false positives contains both true positives (dark ships) and false positives (noise). The F1 score is shown in Figure 4 as a function of SNR and L thresholds. As the true precision is expected to be higher, the optimal hyperparameters would thus lie to the left of the F1 curve maxima, and we have therefore chosen the thresholds $SNR > 2.5$ and $L \geq 3$ based on Figures 4 and 5. Two scenes based on this choice of parameters are shown in Figures 1 and 2. By visual inspection of Figure 1, we estimate that most of the false positives are in fact dark ships.


Figure 4. F1 score by SNR and L thresholds. Optimal detection hyperparameters can be chosen by maximizing the F1.

The cross-polarized channel HV was found to be best suited for ship detection and more robust to clutter for incidence angles below $< 50°$ by [9–11], while [7] applies detection to each polarization with the resulting detections being the union of detections. Refs. [12,13] also argue for fusing the channels which we find more suitable for faster processing of the large number of scenes. By the same means as above, we optimized the linear combination of the polarizations $r \cdot HH + (1 - r) \cdot HV$ with $r \in [0, 1]$, finding an optimal value of $r = 0.2$. The total number of detections and AIS assignments (ground truths) for the 200 scenes can be seen in Figure 6.
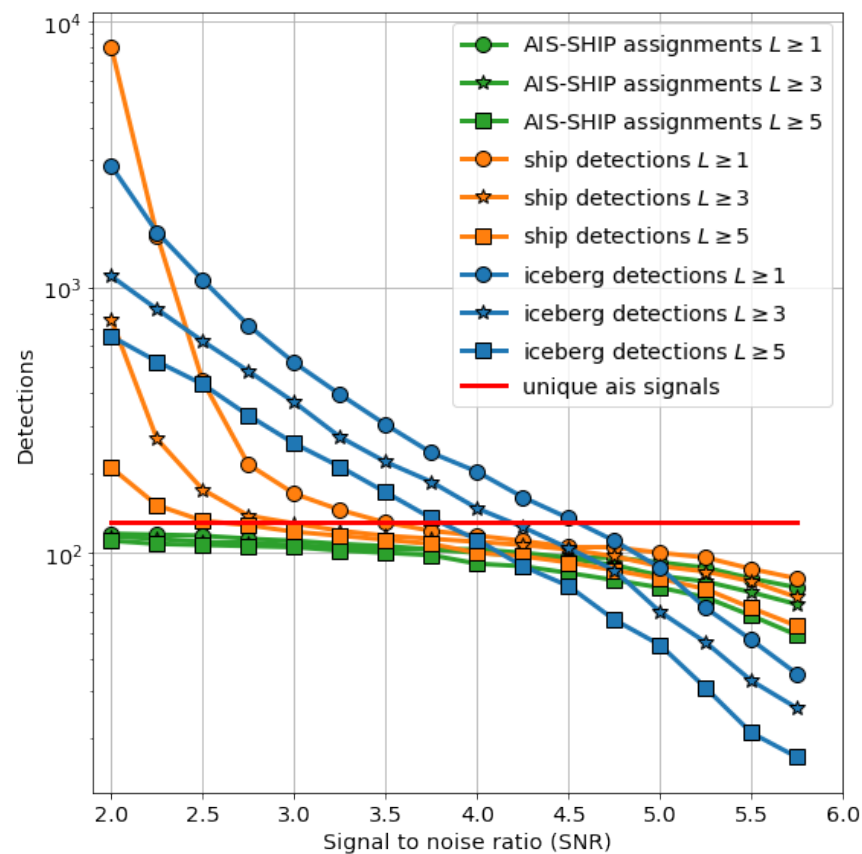
Figure 5. Assignments of AIS signals to SAR detections by SNR and L thresholds. Ship detections, AIS-Ship assignments and Unique AIS signals are from the scene shown in Figure 1, while iceberg detections are from the scene shown in Figure 2.
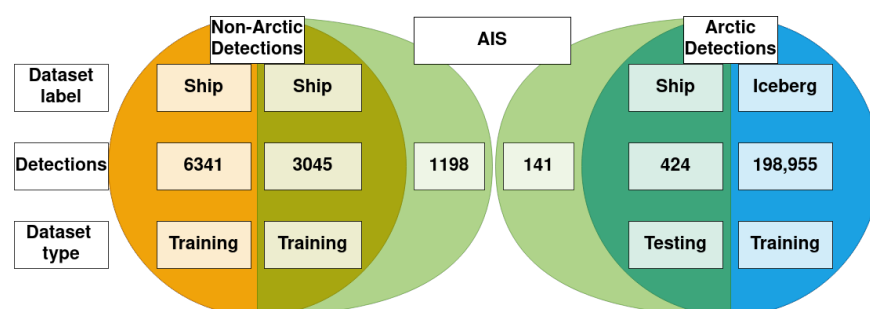
Figure 6. Overview of SAR detections and AIS assignment of the 200 SAR scenes. Non-Arctic detections (orange) with label "ship" regardless of AIS assignment (green overlap). Arctic detections (blue) with label "iceberg" unless an AIS assignment was made. Orange and green overlap respectively indicates the number of dark ships, AIS-SHIP assignments and unassigned AIS signals for the 100 non-Arctic SAR scenes. Likewise for blue Arctic scenes. Dataset type specifies whether the data were used for training and validation or testing.

4. Methods for Ship–Iceberg Discrimination

We first describe the construction of an annotated dataset from the ship and iceberg detections followed by the division into training and testing datasets. Then, we describe the CNN models, parameters, training and select and discuss the best performing model.

4.1. Dataset

A total of 200 Sentinel-1 SAR scenes with HH+HV polarization were analyzed, and the resulting detections were matched to an AIS database. Detections from non-Arctic

scenes or with an AIS assignment were labeled as a ship; otherwise, they were labeled as an iceberg. This yielded 9810 ships and more than ten times as many icebergs. A dataset of images was then constructed by cropping $75 \times 75 \times 3$ pixels around the SAR detections location. The three channels consisted of HH, HV and $(HH + HV)/2$ as in [21], which is also compatible with common image classifiers. We then annotated 1 for ship and 0 for iceberg and split the now annotated dataset into training and testing sets. The training set consisted of all the 9386 ships from the 100 non-Arctic scenes and an equal amount of icebergs selected at random, while the testing set comprised the 424 ships from the 100 Arctic scenes. This Arctic test dataset represents an operational setting with the same latitudes, weather conditions, background, etc. for both ship and icebergs. A full overview of the number of detections, labeling and train/test split is presented in Figure 6, and samples of the ships are shown in Figure 7.
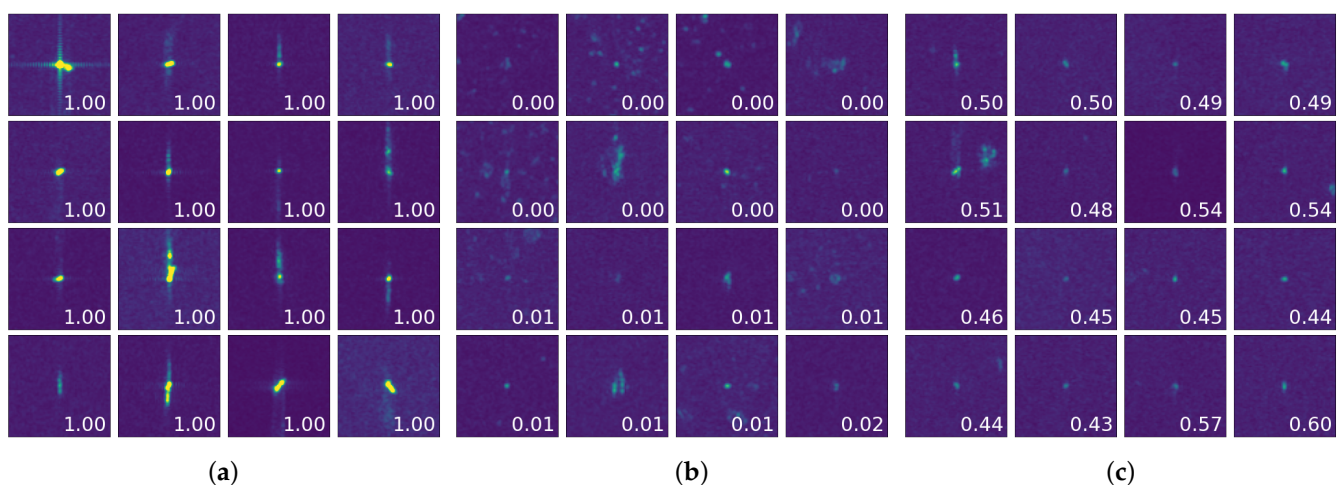
Figure 7. Samples of ships from the test dataset. White number indicates model output probability for the sample, i.e., 0 for iceberg and 1 for ship. All of the shown samples are labeled as ships. The 16 most correct, wrong and uncertain predictions are shown in respectively (**a**–**c**).

4.2. Convolutional Neural Networks

Machine learning is data driven rather than model driven. In this study, we do not intend to analyze very deep CNNs, complex architectures or fine-tune parameters and instead aim at a general investigation of ship–iceberg discrimination. We therefore construct a new CNN, the IceNet, and compare it to two models used in previous ship–iceberg discrimination studies and two deeper models with commonly used architectures. The IceNet has four layers consisting each of two blocks of an Inception module (see [31]), ReLU activation and batch normalization, which is then followed by maxpooling and dropout. We use 64 features per layer, which are split in the Inception module into 1/8 1 by 1 convolutions, 1/2 3 by 3 convolutions, 1/4 5 by 5 convolutions, 1/8 3 by 3 maxpooling and subsequently concatenated. Following the two blocks, a maxpool with a 2 by 2 kernel and a stride of 2 is used to downsample the image, and lastly, a 20% dropout is applied. The output of the four layers is then passed through an adaptive average pool that averages the features into 64, which is subsequently reduced to 1 by a linear layer. From two previous ship–iceberg discrimination studies, we adopted a two-layer model [17] (Two Layer) and a deeper four-layer network [21] (Four layer). We compare these to two more complex general models belonging to a family of models that were developed for difficult classification tasks and previously have been used for SAR ship detection [32,33]: a ResNet18 [34] and an Inception v1 (GoogleNet) [31] both imported from torchvision [35] without pre-training. An overview of the models and number of parameters is listed in Table 1. For a better comparison of classification, we use an Adam optimizer (learning rate 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$), a Binary Cross Entropy loss function, and a batch size of 24 for all models.

Table 1. Lowest obtained validation loss and accuracy for each model across the 5 folds.

Model	Mean Loss	Mean Accuracy	Trainable Parameters
Two Layer	0.35 ± 0.02	$82.7 \pm 0.9\%$	16.417
IceNet	0.12 ± 0.00	$93.4 \pm 0.7\%$	155.777
Four Layer	0.14 ± 0.00	$93.0 \pm 0.6\%$	561.217
ResNet18	0.20 ± 0.01	$88.8 \pm 0.7\%$	5600.929
GoogleNet	0.14 ± 0.01	$92.7 \pm 0.7\%$	11,177.025

We used a 5-fold cross-validation, i.e., training five models using 80/20% training/validation split, such that the combined five models had been trained and validated on the entire training dataset to form an ensemble. After a minimum of 10 epochs of training, we applied an early stopping after 15 epochs when the validation loss did not improve. Computations were carried out using PyTorch [36] on an Intel Core i7-10710U CPU with base frequency of 1.10 GHz. The accuracy is an average over predictions calculated for each ship and iceberg as

$$Accuracy = \frac{1}{N} \sum_{i=1}^{N} 1 - |\hat{y}_i - y_i|, \tag{5}$$

where y is the ground truth, i.e., 1 for ship and 0 for iceberg, while $\hat{y}$ is the model prediction, a probability between 0 and 1.

The training and validation results for each of the five models are shown in Figure 8 and Table 1 with the IceNet achieving the lowest validation loss and highest accuracy.
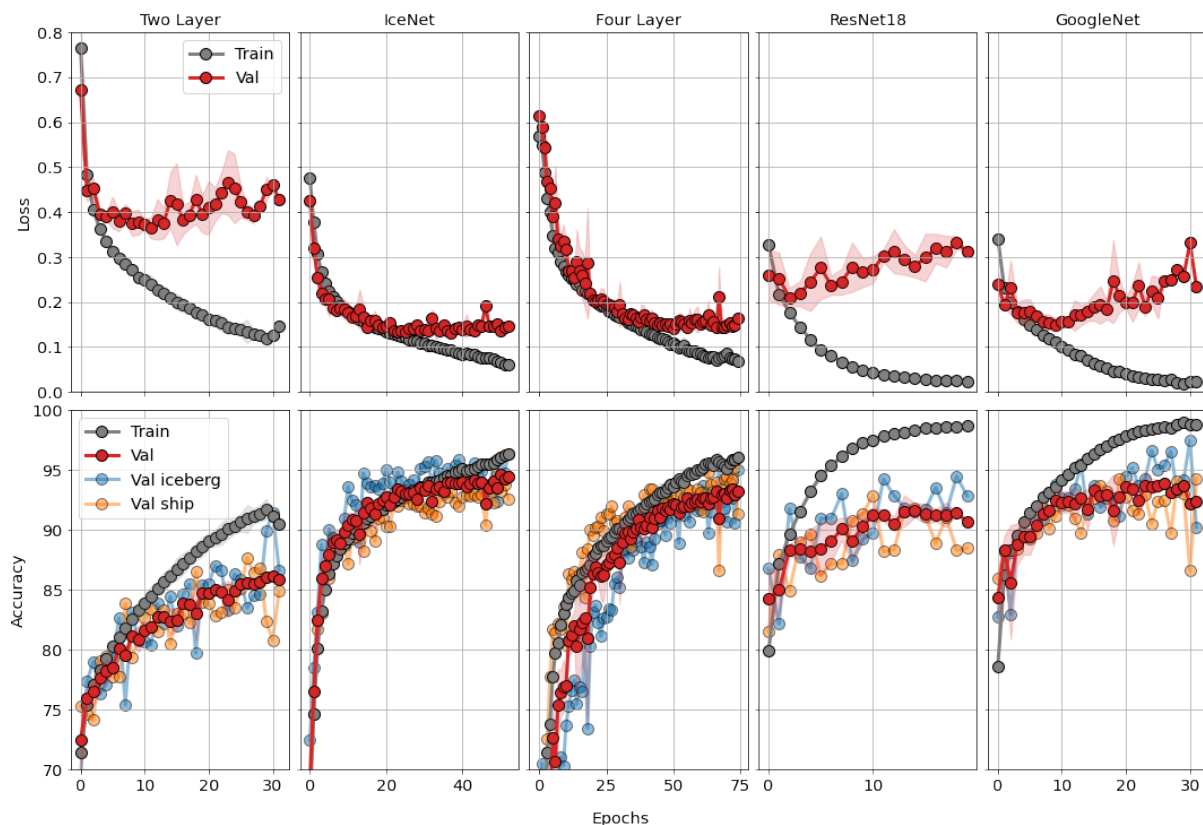

Figure 8. (**Top**) Mean training (gray) and validation (red) loss per epoch for Two Layer; IceNet; Four layer; ResNet18 and GoogleNet models across folds. (**Bottom**) Mean training and validation accuracy per epoch of each model across folds. Colored area represents one standard deviation. Note the varying training epochs due to early stopping on validation loss.

CNNs such as the ResNet and GoogleNet were created for huge datasets of images with three distinct channels and several hundreds of classes. In relation, the binary clas-

sification of this study is simple, which is why these models tend to overfit fast, as they are not regularized for the case. However, the GoogleNet is still capable of achieving a low validation loss, which we believe is due to the maxpooling operation of the Inception module used in the model. In the creation of the IceNet, we noticed that replacing the maxpooling downsampling operation with strided convolutions as in [37] led to a significant drop in performance of the network. The necessity of the maxpooling layer may be connected to the importance of the SAR intensities, which was almost entirely the deciding factor in the ship–iceberg discrimination of [14]. By stacking Inception modules and using maxpooling for downsampling, we allow for the SAR intensities to more freely flow through the IceNet model. Figure 8 show lower training than validation loss, which is a well-known overfitting symptom. The Four-Layer network achieved similar low validation loss as the GoogleNet by utilizing the regularization from the several dropout layers that reduce overfitting [38]. IceNet is therefore regularized both by dropout layers and the use of batch normalization [39]. Choosing the right amount of regularization was ultimately done by trial and error. The Two-Layer and Four-Layer models were both capable of achieving relatively high accuracy despite having many fewer parameters than the ResNet and GoogleNet. Since large ships are distinctly different to icebergs, the difficulty in ship–iceberg discrimination comes from smaller ships that are more similar to icebergs in shape and intensity near the resolution limit of the SAR instrument. Our development of the IceNet therefore focused on having few parameters which both reduce the training time and memory usage of the model. In line with the simplicity of the discrimination, we opted for the straightforward CNN architecture of stacked convolutional layers with pooling operations instead of investigating very deep networks or complicated architectures.

The validation accuracies are higher for AIS assigned ships than dark ships (see Figure 9) because the AIS transmitting ships generally are larger than dark ships.
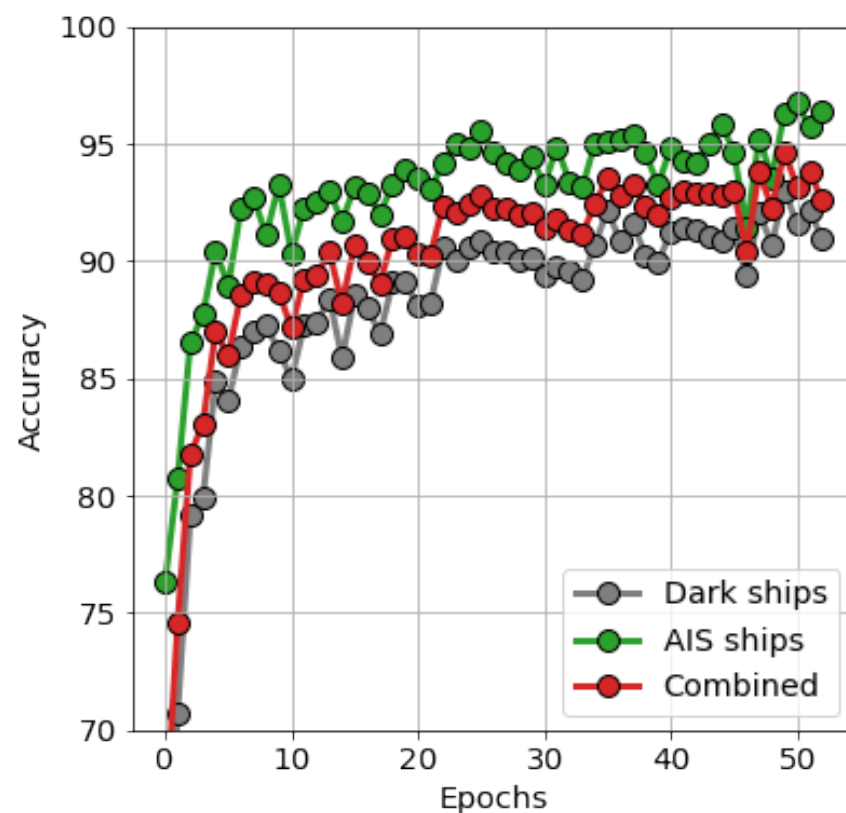
**Figure 9.** IceNet validation accuracies for dark ships and AIS ships.

The smaller dark ships impact the accuracy negatively when training. However, as icebergs are also smaller on average than AIS transmitting ships, including the dark ships in

training may help the model to discriminate smaller ships from icebergs. For this purpose, we train an IceNet model, as this was the best-performing model using only the 3045 AIS assigned ships and an equal amount of icebergs. This ensemble reaches the lowest loss of 0.08 and highest accuracy of 95% due to the omittance of harder to discriminate dark ships.

5. Results

The IceNet ensemble was subsequently used for testing on the Arctic test dataset consisting of the 424 AIS assigned ships and 20,000 icebergs, reaching an accuracy of 83% for ships and 97% for icebergs.

Figure 10 presents the correlation between the size of the ship or iceberg and the accuracy and PPV; here, the signal to noise ratio (SNR, see Equation (2)) is a measure of both the ship signal strength and size. The ships in the test dataset are on average slightly smaller than the AIS transmitting ships of the training dataset, while dark ships are more similar to icebergs.
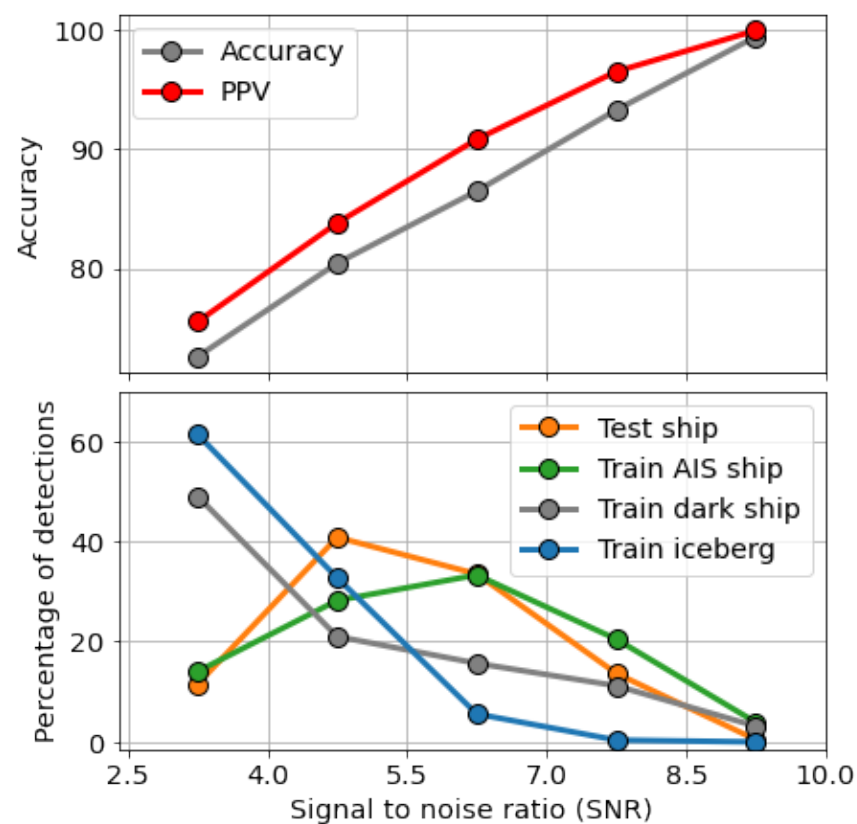
Figure 10. Test ship accuracy vs. signal to noise ratio (SNR, see Equation (2)) for the IceNet model. (**Top**) Test ship accuracy and PPV for all samples in the SNR bin (vertical lines). (**Bottom**) Percentage of the dataset in the SNR bin.

To further evaluate the model accuracy, we exploit the positive predictive value (PPV). The output probability of the model is a number between 0 and 1, where 1 is a ship and 0 is an iceberg. Usually probabilities above and below the threshold of 0.5 are then classified as ships and icebergs, respectively. The PPV is presented in Figure 11 and Table 2 with the accuracy for both the models trained using the full dataset including dark ships and the dataset excluding dark ships, i.e., with only AIS assigned ships. The ensemble trained on the dataset including dark ships achieved a slightly higher accuracy and PPV for ships than the model excluding dark ships.
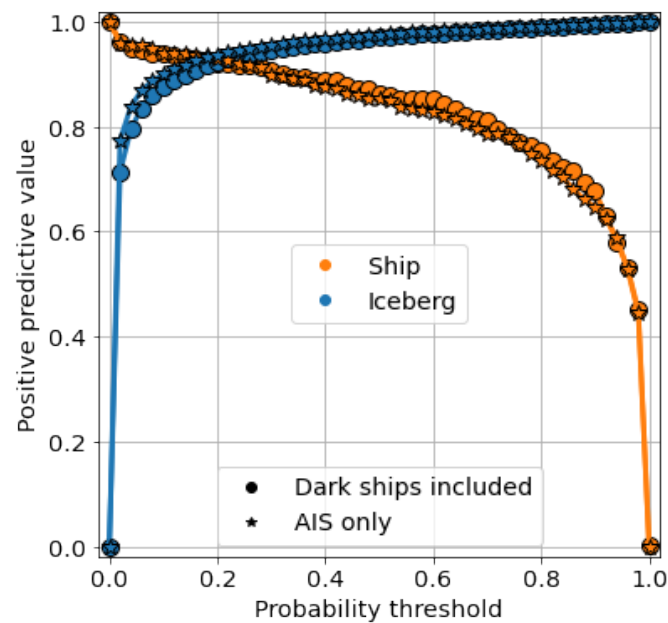
Figure 11. Positive predictive value (PPV) for the IceNet model ensembles trained on the full dataset including dark ships (Normal), only ships with AIS transmission (AIS).

Table 2. Test accuracies and PPV for the IceNet model ensembles trained on the dataset including dark ships or only ships with AIS signal.

	Dark Ships Included	AIS Only
Ship accuracy	83%	82%
Iceberg accuracy	95%	96%
Ship PPV	86%	85%
Iceberg PPV	97%	97%

6. Discussion

The discussion is divided into two sections. First, we analyze the ship detections and the creation of the dataset. Then, we discuss the CNN models and ship–iceberg discrimination.

6.1. Detection of Ships and Icebergs, and AIS Correlation

In total, 200 SAR scenes were processed using a CWT-based ship detector. We could optimize the detector by selecting hyperparameters that resulted in a high precision and recall by maximizing the F1 measure (see Figure 4). This was done by an analysis of a scene from the English Channel, as shown in Figure 1, and correlating the detections to AIS signals.

There was a range of hyperparameters with almost optimal F1 score (Equation (4)). Lowering SNR or L leads to more assigned ships (true positives) at the cost of more unassigned detections (false positives), which increased the recall but decreased the precision, resulting in the broad maxima of Figure 4. For the scene presented in Figure 1, we respectively found a precision and recall of 64% and 85%. If parameters had been chosen corresponding to the optimal F1 value, we could obtain results comparable to those of [29], who analyzed a similar scene of the English Channel and found 83% precision and 76% recall. However, by a closer look at our unassigned detections, we are convinced that many of them are in fact dark ships, and the precision we obtained would significantly increase if dark ships could be annotated. By allowing the detection of these smaller dark ships, we also avoid biasing the dataset toward larger cargo and tanker ships.

The precision and recall were found to vary considerably from scene to scene. Across the 100 non-Arctic scenes, we found an average precision of only 32% and a recall of

72% (numbers are provided in Figure 6). Generally, there is lower AIS coverage in the North Sea and other open oceans, while in the Bay of Guinea and the east coast of Africa, ships are known to turn off their transponders due to pirates, leading to more dark ships and thus lower precision. In contrast, larger vessels such as tanker and cargo ships have well-functioning AIS transponders and are easily detected at high SNR, resulting in the high global recall. This is also evident from the analysis of the 100 Arctic scenes, which resulted in a recall of 75%. Initially, we tested several detection algorithms, e.g., the CFAR-type algorithms of [40], but found slightly better results with the CWT algorithm. It automatically removes a sloping background and as such does not require pre-processing of the SAR scenes. We also found the CWT algorithm to be robust to sea clutter caused by heavy weather.

The interpolated AIS signals were assigned to ship detection using a nearest neighbor assignment algorithm. Correcting for the Doppler shift led to a more precise AIS and detection match spatially, which was especially important for the Arctic scenes where ships sail close to icebergs (see Figure 2). Some of these ships may in a few cases have caused the assignment algorithm to assign an iceberg detection to the AIS signal, but the number of false assignments were reduced by selecting a low assignment threshold.

In the creation of the LS-SSDD dataset [23], a number of SAR experts were used to label ships by referencing SAR scenes and AIS data. We considered such a task unfeasible, as this study analyzed 200 SAR scenes compared to the 15 of the LS-SSDD. Our resulting database of 9810 ships includes not only assigned ships but also the unassigned non-Arctic detections, which we believe are mostly dark ships. However, there may be some false alarms, but they are kept to a minimum by the land mask, as well as sea turbine removal.

6.2. Ship and Iceberg Discrimination

All trained models were capable of achieving high performance for each of the folds, indicating that the training/validation splits resolved the data well, and even the very simple two-layer model with the fewest parameters achieved a reasonable accuracy, indicating that the features discriminating ships from icebergs can be considered relatively low level. The GoogleNet with the highest number of parameters was capable of achieving almost identical loss as the Four-Layer model while the ResNet18 did worse, which could be due to the double descent phenomenon [41], and as such, deeper models should be explored in future work. We did not find augmentation to improve training, as also found in [21].

The IceNet model obtained the lowest validation loss, and we therefore used it for testing on the test dataset achieving high accuracy and PPV, which are listed in Table 2. The iceberg accuracy was lower than the validation result shown in Table 1 which indicates that the model is slightly biased toward icebergs. Inversely, the ship classification accuracy was much lower than the validation, which stems from discrepancies between the two datasets. Several of the 16 Arctic ship images with the lowest accuracy, shown in Figure 7c, contain structures aside from the ship, which complicates the matter, as the model was not trained on such images. Icebergs and ice floes are often more tightly packed than ships (see Figures 1 and 2 for a comparison), and the model may have learned to associate images containing several structures to the iceberg class and naturally but wrongly classifies the images of Figure 7b as icebergs. Improving upon the models by achieving lower validation accuracy is therefore unlikely to improve the test accuracy much compared to including examples such as the ones of Figure 7b to the dataset.

Bentes et al. [17] obtained an impressive PPV of 100% for ships and 95% for icebergs with the simple Two-Layer model. A major part of their improvement was due to the more than 10 times higher pixel resolution of 3 m in their TerraSar-X images. In addition, their dataset contained only 277 ships and 68 icebergs, which was augmented to 300 ships and 300 icebergs before training; subsequently the same ship rotated or flipped was used both for training and validation. Nonetheless, better resolution SAR images will improve classification but have to be outweighed against coverage, recurrence time, and the price and availability of such satellites. A manually labeled dataset of 125 ships and 76 icebergs

was used in [15], achieving an almost equal accuracy of 93.5% to that of our validation accuracy using a maximum likelihood Gaussian classifier. Smaller ships, such as the ones in Figure 7b,c, were found to be more difficult to classify due the limited resolution by [16,17,42]. We also find this correlation between the ship size and accuracy as seen in Figure 10. Ref. [14] analyzed a dataset of 19 large supply vessels and 20 icebergs by using the size ratio of the detections polarizations, HV/HH to achieve a 97% accuracy, which is comparable to ours for large SNR. However, it should also be noted that the ships used in [14] were almost all comparatively larger than the icebergs. In general, it is easy to discriminate large ships from icebergs even by the naked eye, as these have unique characteristics such as sidelobes, wake, doppler shift and shape (see Figure 7a). By removing smaller dark ships from the training data, we were also able to achieve a higher validation accuracy. However, since the ships in the Arctic test dataset are on average slightly smaller than the training set (see Figure 10), the model has not been trained sufficiently to discriminate smaller ships from icebergs and achieves a slightly lower test accuracy.

7. Conclusions

A total of 200 Sentinel-1 SAR scenes were analyzed and ships and icebergs were detected using a continuous wavelet transform-based detection algorithm, using AIS signals as ground truth to optimize parameters and reduce false alarms. Globally, we were able to assign 72% of the AIS signals to a SAR detection and 32% of the SAR detections to an AIS signal. The precision and recall fluctuated across the 200 scenes, which was mainly due to the AIS coverage and type of ships. Areas with good coverage and traditional shipping routes with large tankers and supply vessels had highest precision and recall, whereas the precision dropped significantly in areas with sparse AIS coverage and smaller vessels.

A large annotated dataset of small SAR images was constructed from the ship and iceberg detections. Here, we include smaller ships and dark ships which do not transmit AIS signals, but which we are confident are ships. Several CNN models were trained to discriminate ships from icebergs that proved to be highly effective by achieving up to 93% validation accuracy. Subsequently, the best model was evaluated on a test dataset consisting of AIS assigned ship detections from the SAR scenes over the Disco Bay. The model achieved an accuracy of 83% in this operational test scenario of ships sailing in the Arctic. The lower accuracy could be attributed to ships sailing close to icebergs and ice floes in the test dataset and it containing both ships and icebergs, which confuses the discrimination algorithm. It is difficult to apply binary classification to discriminate ships from icebergs in such testing environments, and future investigations should therefore go beyond binary classification. Methods such as semantic segmentation or region proposal could solve this problem, and dependencies on ship size and velocity could be explored.

The accurate detection and classification of ships and icebergs is crucial for false alarm reduction in Arctic surveillance, where, e.g., the Royal Danish Arctic Command with limited resources has to survey the enormous Arctic territories all year in a difficult climate.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SAR	Synthetic Aperture Radar
AIS	Automatic Identification System
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
SNR	Signal to Noise Ratio
CFAR	Constant False Alarm Rate
ESA	European Space Agency

References

1. Barkham, P. Russian Tanker Sails through Arctic without Icebreaker for First Time. *The Guardian*, 24 August 2017.
2. Chen, J.L.; Kang, S.C.; Guo, J.M.; Xu, M.; Zhang, Z.M. Variation of sea ice and perspectives of the Northwest Passage in the Arctic Ocean. *Adv. Clim. Chang. Res.* **2021**, *12*, 447–455. [CrossRef]
3. Ball, H. *Satellite AIS for Dummies*; Wiley: Hoboken, NJ, USA, 2013.
4. Høye, G.K.; Eriksen, T.; Meland, B.J.; Narheim, B.T. Space-based AIS for global maritime traffic monitoring. *Acta Astronaut.* **2008**, *62*, 240–245. [CrossRef]
5. BBC. Warship Positions Faked Including UK Aircraft Carrier. *BBC News*, 2 August 2021.
6. Crisp, D.J. *The State-of-the-Art in Ship Detection in Synthetic Aperture Radar Imagery*; Technical Report; Defence Science And Technology Organisation Salisbury (Australia). 2004. Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1000.6717&rep=rep1&type=pdf (accessed on 1 June 2021).
7. Greidanus, H.; Alvarez, M.; Santamaria, C.; Thoorens, F.X.; Kourti, N.; Argentieri, P. The SUMO Ship Detector Algorithm for Satellite Radar Images. *Remote Sens.* **2017**, *9*, 246. [CrossRef]
8. Palubinskas, G.; Reinartz, P.; Brusch, S.; Lehner, S. Joint Use of Optical and SAR Data for Ship Detection. 2009. pp. 1–6. Available online: https://elib.dlr.de (accessed on 1 June 2021).
9. Yeremy, M.; Campbell, J.; Mattar, K.; Potter, T. Ocean Surveillance with Polarimetric SAR. *Can. J. Remote Sens.* **2001**, *27*, 328–344. [CrossRef]
10. Touzi, R.; Charbonneau, F.J.; Hawkins, R.K.; Vachon, P.W. Ship detection and characterization using polarimetric SAR. *Can. J. Remote Sens.* **2004**, *30*, 552–559. [CrossRef]
11. Touzi, R.; Charbonneau, F.; Hawkins, R.; Murnaghan, K.; Kavoun, X. Ship-sea contrast optimization when using polarimetric SARs. In Proceedings of the IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), Sydney, Ausralia, 9–13 July 2001; Volume 1, pp. 426–428. [CrossRef]
12. Sciotti, M.; Pastina, D.; Lombardo, P. Polarimetric detectors of extended targets for ship detection in SAR images. In Proceedings of the IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), Sydney, Ausralia, 9–13 July 2001; Volume 7, pp. 3132–3134. [CrossRef]
13. Sciotti, M.; Pastina, D.; Lombardo, P. Exploiting the polarimetric information for the detection of ship targets in non-homogeneous SAR images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 24–28 June 2002; Volume 3, pp. 1911–1913. [CrossRef]
14. Howell, C.; Youden, J.; Lane, K.; Power, D.; Randell, C.; Flett, D. Iceberg and ship discrimination with ENVISAT multipolarization ASAR. In Proceedings of the IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AL, USA, 20–24 September 2004; Volume 1, p. 116. [CrossRef]
15. Howell, C.; Mills, J.; Power, D.; Youden, J.; Dodge, K.; Randell, C.; Churchill, S.; Flett, D. A multivariate approach to iceberg and ship classification in HH/HV ASAR data. In Proceedings of the 2006 IEEE International Symposium on Geoscience and Remote Sensing, Denver, CO, USA, 31 July–4 August 2006; pp. 3583–3586. [CrossRef]
16. Heiselberg, P.; Heiselberg, H. Ship-Iceberg Discrimination in Sentinel-2 Multispectral Imagery by Supervised Classification. *Remote Sens.* **2017**, *9*, 1156. [CrossRef]
17. Bentes, C.; Frost, A.; Velotto, D.; Tings, B. Ship-Iceberg Discrimination with Convolutional Neural Networks in High Resolution SAR Images. In Proceedings of the EUSAR 2016: 11th European Conference on Synthetic Aperture Radar, Hamburg, Germany, 6–9 June 2016; pp. 1–4.
18. Statoil/C-CORE. Statoil/C-CORE Iceberg Classifier Challenge. 2018. Available online: https://www.kaggle.com/c/statoil-iceberg-classifier-challenge (accessed on 1 April 2021).
19. Jang, H.; Kim, S.; Lam, T. *Kaggle Competitions: Author Identification & Statoil/C-CORE Iceberg Classifier Challenge*; Technical Report; School of Informatics, Computing, and Engineering, Indiana University: Bloomington, IN, USA, 2017.
20. Yang, X.; Ding, J. A Computational Framework for Iceberg and Ship Discrimination: Case Study on Kaggle Competition. *IEEE Access* **2020**, *8*, 82320–82327. [CrossRef]

21.   Heiselberg, H. Ship-iceberg classification in SAR and multispectral satellite images with neural networks. *Remote Sens.* **2020**, *12*, 2353. [CrossRef]
22.   ESA Copernicus Program, Sentinel Scientific Data Hub. Available online: https://scihub.copernicus.eu/ (accessed on 1 October 2021).
23.   Zhang, T.; Zhang, X.; Ke, X.; Zhan, X.; Shi, J.; Wei, S.; Pan, D.; Li, J.; Su, H.; Zhou, Y.; et al. LS-SSDD-v1.0: A Deep Learning Dataset Dedicated to Small Ship Detection from Large-Scale Sentinel-1 SAR Images. *Remote Sens.* **2020**, *12*, 2997. [CrossRef]
24.   Alaska Satellite Facility (ASF) Distributed Active Archive Center. Available online: https://asf.alaska.edu/ (accessed on 1 October 2021).
25.   Wessel, P.; Smith, W.H. A global, self-consistent, hierarchical, high-resolution shoreline database. *J. Geophys. Res. Solid Earth* **1996**, *101*, 8741–8743. [CrossRef]
26.   Du, P.; Kibbe, W.A.; Lin, S.M. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* **2006**, *22*, 2059–2065. [CrossRef] [PubMed]
27.   European Marine 140 Observation and Data Network (EMODnet). Available online: https://www.emodnet-humanactivities.eu/view-data.php (accessed on 14 March 2022).
28.   Sang, L.Z.; Yan, X.P.; Mao, Z.; Ma, F. Restoring method of vessel track based on AIS information. In Proceedings of the 2012 11th International Symposium on Distributed Computing and Applications to Business, Engineering & Science. Guilin, China, 19–22 October 2012; pp. 336–340.
29.   Rodger, M.; Guida, R. Classification-Aided SAR and AIS Data Fusion for Space-Based Maritime Surveillance. *Remote Sens.* **2020**, *13*, 104. [CrossRef]
30.   Heiselberg, H. Aircraft and Ship Velocity Determination in Sentinel-2 Multispectral Images. *Sensors* **2019**, *19*, 2873. [CrossRef] [PubMed]
31.   Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.E.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. *arXiv* **2014**, arXiv:1409.4842.
32.   Li, Y.; Ding, Z.; Zhang, C.; Wang, Y.; Chen, J. SAR ship detection based on resnet and transfer learning. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2019), Yokohama, Japan, 28 July–2 August 2019; pp. 1188–1191.
33.   Kang, K.M.; Kim, D.J. Ship velocity estimation from ship wakes detected using convolutional neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4379–4388. [CrossRef]
34.   He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
35.   Marcel, S.; Rodriguez, Y. Torchvision the machine-vision package of torch. In Proceedings of the 18th ACM international conference on Multimedia, Firenze, Italy, 25–29 October 2010.
36.   Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*; Curran Associates, Inc.: Red Hook, NY, USA, 2019. Available online: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf (accessed on 1 April 2021).
37.   Springenberg, J.T.; Dosovitskiy, A.; Brox, T.; Riedmiller, M. Striving for simplicity: The all convolutional net. *arXiv* **2014**, arXiv:1412.6806.
38.   Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
39.   Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 448–456.
40.   Richards, M.A. *Fundamentals of Radar Signal Processing*; McGraw-Hill Education: New York, NY, USA, 2014.
41.   Nakkiran, P.; Kaplun, G.; Bansal, Y.; Yang, T.; Barak, B.; Sutskever, I. Deep double descent: Where bigger models and more data hurt. *J. Stat. Mech. Theory Exp.* **2021**, *2021*, 124003. [CrossRef]
42.   Heiselberg, H. Ship-iceberg detection & classification in sentinel-1 SAR images. *TransNav* **2020**, *14*, 235–241. [CrossRef]