Ship identification and characterization in Sentinel-1 SAR images with multi-task deep learning

Clément Dechesne, Sébastien Lefèvre, Rodolphe Vadaine, Guillaume Hajduch, Ronan Fablet

MDPI

Article

Ship identification and characterization in Sentinel-1 SAR images with multi-task deep learning

Clément Dechesne [1], Sébastien Lefèvre [2], Rodolphe Vadaine [3], Guillaume Hajduch [3], Ronan Fablet [1]

[1] IMT Atlantique – Lab-STICC, UMR CNRS 6285, Brest, FR
[2] Univ. Bretagne Sud – IRISA, UMR CNRS 6074, Vannes, FR
[3] Collecte Localisation Satellites, Brest, FR
* Correspondence: sebastien.lefevre@irisa.fr

Version December 12, 2019 submitted to Remote Sens.

Abstract: The monitoring and surveillance of maritime activities are critical issues in both military and civilian fields, including among others fisheries monitoring, maritime traffic surveillance, coastal and at-sea safety operations, tactical situations. In operational contexts, ship detection and identification is traditionally performed by a human observer who identifies all kinds of ships from a visual analysis of remotely-sensed images. Such a task is very time consuming and cannot be conducted at a very large scale, while Sentinel-1 SAR data now provide a regular and worldwide coverage. Meanwhile, with the emergence of GPUs, deep learning methods are now established as state-of-the-art solutions for computer vision, replacing human intervention in many contexts. They have been shown to be adapted for ship detection, most often with very high resolution SAR or optical imagery. In this paper, we go one step further and investigate a deep neural network for the joint classification and characterization of ships from SAR Sentinel-1 data. We benefit from the synergies between AIS (Automatic Identification System) and Sentinel-1 data to build significant training datasets. We design a multi-task neural network architecture composed of one joint convolutional network connected to three task-specific networks, namely for ship detection, classification and length estimation. The experimental assessment showed our network provides promising results, with accurate classification and length performance (classification overall accuracy: 97.25%, mean length error: 4.65 m $\pm$ 8.55 m).

Keywords: Deep neural network, Sentinel-1 SAR images, Ship identification, Ship characterization, Multi-task learning

1. Introduction

Deep learning is considered as one of the major breakthrough related to big data and computer vision [1]. It has become very popular and successful in many fields including remote sensing [2]. Deep learning is a paradigm for representation learning and is based on multiple levels of information. When applied on visual data such as images, it is usually achieved by means of convolutional neural networks. These networks consist of multiple layers (such as convolution, pooling, fully connected and normalization layers) aiming to transform original data (raw input) into higher-level semantic representation. With the composition of enough such elementary operations, very complex functions can be learned. For classification tasks, higher-level representation layers amplify aspects of the input that are important for discrimination and discard irrelevant variations. For humans, it is simple through visual inspection to know what objects are in an image, where they are, and how they interact in a very fast and accurate way, allowing to perform complex tasks. Fast and accurate algorithms for object detection are thus sought to allow computers to perform such tasks, at a much larger scale than

humans can achieve.

Ship detection and classification have been extensively addressed with traditional pattern recognition techniques for optical images. *Zhu et al.*[3] and *Antelo et al.*[4] extracted handcrafted features from images such as shapes, textures and physical properties, while *Chen et al.*[5] and *Wang et al.*[6] exploited Dynamic Bayesian Networks to classify different kinds of ships. Such extracted features are known for their lack of robustness that can raise challenges in practical applications (e.g. they may lead to poor performances when the images are corrupted by blur, distortion, or illumination which are common artifacts in remote sensing). Furthermore, they cannot overcome the issues raised by big data such as image variabilities (i.e. ships of same type may have different shape, color, size, etc.) and data volume. Recently, following the emergence of deep learning, an autoencoder-based deep neural network combined with extreme learning machine was proposed [7] and outperformed some other methods using SPOT-5 spaceborne optical images for ship detection.

Compared with optical remote sensing, satellite SAR imaging appears more suited for maritime traffic surveillance in operational contexts as it is not critically affected by weather conditions and day-night cycles. In this context, open-source Sentinel-1 SAR data are particularly appealing. Almost all coastal zones and shipping routes are covered by Interferometric Wide Swath Mode (IW), while the Extra-Wide Swath Mode (EW) acquires data over open oceans, providing a global coverage for sea-oriented applications. Such images, combined with the Automatic Identification System (AIS), represent a large amount of data that can be employed for training deep learning models [8]. AIS provides meaningful and relevant information about ships (such as position, type, length, rate of turn, speed over ground, etc.). The combination of these two data sources could leverage new applications to the detection and estimation of ship parameters from SAR images, which remains a very challenging task. Indeed, detecting inshore and offshore ships is critical in both military and civilian fields (e.g. for monitoring of fisheries, management of maritime traffics, safety of coast and sea, etc). In operational contexts, the approaches used so far still rely on manual visual interpretations that are time-consuming, possibly error-prone, and definitely irrelevant to scale up to the available data streams. On the contrary, the availability of satellite data such as Sentinel-1 SAR makes possible the exploration of efficient and accurate learning-based schemes.

One may however consider with care AIS data as they involve specific features. AIS is mandatory for large vessels (e.g., >500GT, passenger vessels). As such, it provides representative vessel datasets for international maritime traffic, but may not cover some maritime activities (e.g., small fishing vessels). Though not authorized, ships can easily turn off their AIS and/or spoof their identity. While AIS tracking strategies [9] may be considered to address missing track segments, the evaluation of spoofing behaviour is a complex task. *Iphar et al.* [10] evaluate that amongst ships with AIS, about 6% have no specified type, 3% are only described as "vessels". Besides, respectively 47% and 18% of the vessels may involve uncertain length and beam data. These points should be considered with care in the analysis of AIS datasets, especially when considering learning strategies as addressed in this work.

Among existing methods for ship detection in SAR images, Constant False Alarm Rate (CFAR)-based methods have been widely used [11,12]. The advantage of such methods is their reliability and high efficiency. Using AIS information along with SAR images significantly improves ship detection performance [13]. As the choice of features has an impact on the performance of discrimination, deep neural networks have recently taken the lead thanks to their ability to extract (or learn) features that are richer than hand-crafted (or expert) features. In [14], a framework named Sea-Land Segmentation-based Convolutional Neural Network (SLS-CNN) was proposed for ship detection, combined with the use of saliency computation. A modified Faster R-CNN based on CFAR algorithm for SAR ship detection was proposed in [15] with good detection performance. In [16], texture features

extracted from SAR images are fed into artificial neural networks (TF-ANN) to discriminate ship pixels from sea ones. *Schwegmann et al.*[17] employed highway network for ship detection in SAR images and achieved good results, especially in reducing the false detection rate. These state-of-the-art approaches focused on ship detection in SAR images. In this paper, we aim to go beyond ship detection and investigate higher-level tasks, namely the identification of ship types (a.k.a. classification) and their length estimation, which to our knowledge remain poorly addressed using learning-based frameworks.

The problem of ship length estimation from SAR images has been briefly discussed in [18,19]. In [18], the best shape of a ship is extracted from a SAR image using inertia tensors. The estimated shape allows to obtain the ship length. However, the absence of ground truth does not allow to validate the accuracy of this method. In [19], a three-step method is proposed in order to extract a rectangle that will be the reference model for ship length estimation. The method produces good results (mean absolute error: 30 m $\pm$ 36.6 m). However, the results are presented on a limited dataset (only 127 ships) and their generalization may be questioned.

In this paper, we propose a method based on deep learning for ship identification and characterization with the synergetical use of Sentinel-1 SAR images and AIS data.

2. Material and Methods

The proposed framework combines the creation of a reference groundtruthed dataset using AIS-SAR synergies and the design of a multi-task deep learning model. In this section, we first introduce the proposed multi-task neural network architecture, which jointly addresses ship detection, classification and length estimation. Second, we describe the training framework in terms of the considered training losses and of the implemented optimization scheme. Third, we detail the creation of the considered reference datasets, including how we tackled data augmentation and class imbalance issues, which have been shown to be critical for the learning process.

2.1. Proposed framework

The proposed multi-task framework is based on two stages, with a first common part and then three task-oriented branches for ship detection, classification and length estimation, respectively (see Figure 1). The first part is a convolutional network made of 5 layers. It is followed by the task-oriented branches. All these branches are made of convolutional layers followed by fully connected layers (the number of which depends on the complexity of the task). For the detection task, the output consists in a pixel-wise probability map of the presence of ships. It only requires 1 fully-connected layer after the 4 convolutional layers. For the classification task, we consider 4 or 5 ship classes (Cargo, Tanker, Fishing, Passenger, and optionally Tug). The branch also requires 4 convolutional layers and 2 fully connected layers. The last task is related to the length estimation. This branch is composed of 4 convolutional layers and 5 fully-connected layers.
This architecture is inspired from state-of-the-art architectures [20–22]. The number of layers has been chosen to be similar to the first layers of the VGG network [22]. All the activations of the convolutional layers and fully-connected layers are ReLu [23]. Other activation functions are employed for the output layers: a sigmoid for the detection, a softmax activation for the classification, and a linear activation is employed for the length estimation, further details are presented in Section 2.2. We may emphasize that our model includes a detection component. Though this is not a targeted operational objective in our context, it was shown to improve the performance for the other tasks (See Table 7).

2.2. Training procedure

We describe below the considered training strategy, especially the training losses considered for each task-specific component. The proposed end-to-end learning scheme combines task-specific training losses as follows:

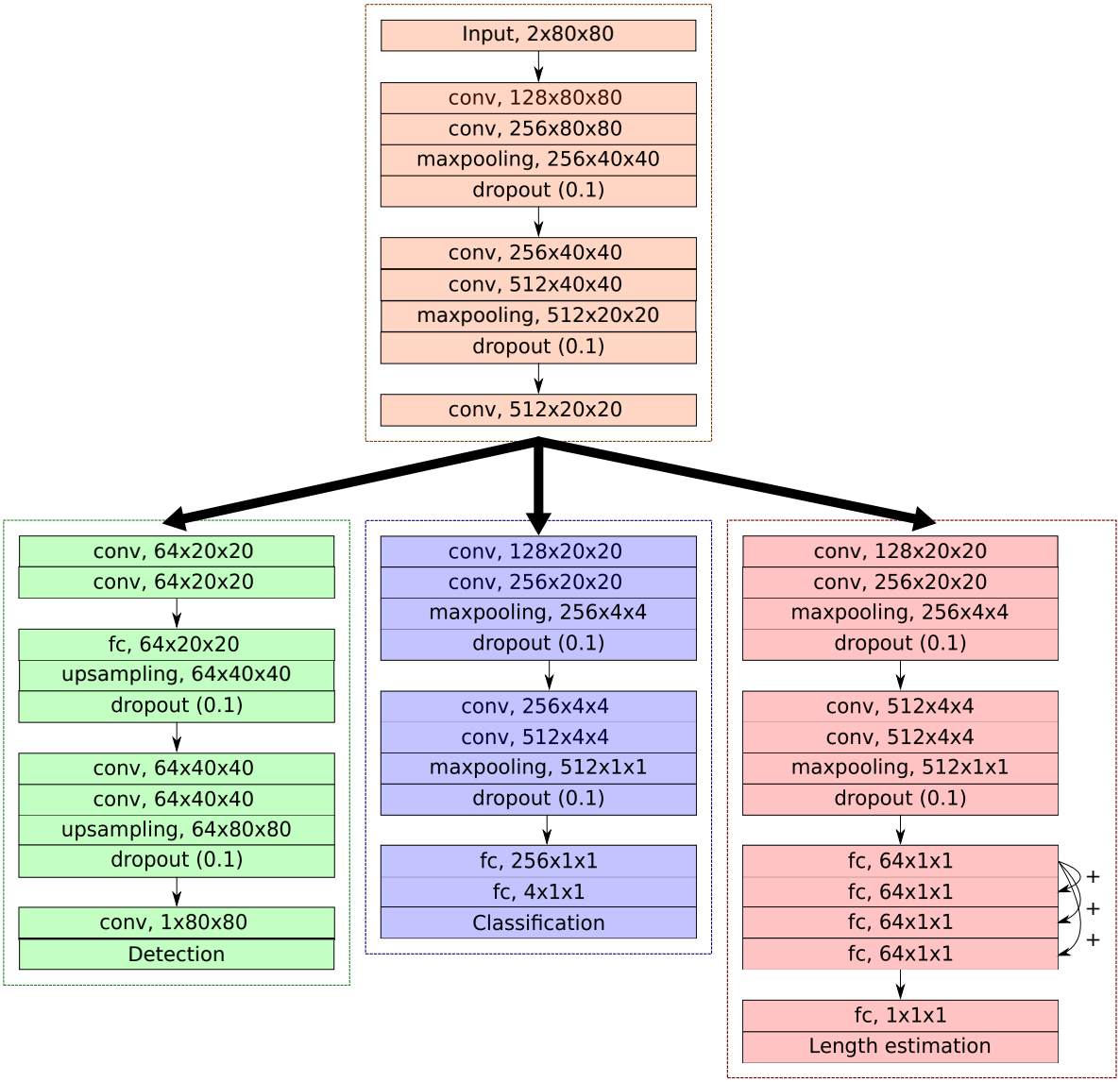


Figure 1. Proposed multi-task architecture for ship detection, classification (4 classes) and length estimation from a Sentinel-1 SAR image.

- **Detection loss**: the detection output is a ship presence probability. We employ a binary cross-entropy loss, which is defined by:

$$L_{det} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k\in I}(y_k \log(p(k)) + (1-y_k)\log(1-p(k))), \quad (1)$$

where N is the number of samples, k is a pixel of the output detection image I, y_k is the ground truth of ship presence (0 or 1), and $p(k)$ is the predicted probability of ship presence. It is a usual loss function for binary classification tasks [24].

- **Classification loss:** The output for the last classification layer is the probability that the input image corresponds to one of the considered ship types. We use here the categorical cross-entropy loss:

$$L_{class} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{n_c}(y_{o,c}\log(p_{o,c})), \quad (2)$$

where N is the number of samples, n_c is the number of classes (here, $n_c = 4$ or $n_c = 5$), $y_{o,c}$ is a binary indicator (0 or 1) if class label c is the correct classification for observation o and $p_{o,c}$ is the predicted probability for the observation o to belong to class c. It is a widely-used loss function for multiclass classification tasks [25,26].

- **Length estimation loss:** in the length estimation network, the 4 fully-connected layers of shape ($64\times1\times1$) are connected to each other (see Figure 1). The idea is to propagate the difference between the first layer and the current layer and is related to residual learning [27]. We use here the mean squared error defined as

$$L_{length} = \frac{1}{N}\sum_{n=1}^{N}(l_{pred} - l_{true})^2, \quad (3)$$

where N is the number of samples, l_{pred} is the predicted length and l_{true} is the true length.

Overall, we define the loss function of the whole network as

$$L = L_{det} + L_{class} + L_{length}. \quad (4)$$

Each specific loss employed to design the loss of the whole network could have been weighted. Nevertheless, we have observed no significant effect of such a weighting scheme. Thus we decided to rely on a simple combination through adding the different task-dedicated losses, giving the same importance to each task. Our network is trained end-to-end using RMSProp optimizer [28]. The weights of the network are updated by using a learning rate of 1e-4 and a learning rate decay over each update of 1e-6 over the 500 iterations. Such parameterization has shown good results for our characterization tasks.

2.3. Creation of reference datasets

With a view to implementing deep learning strategies, we first address the creation of reference datasets from the synergy between AIS data and Sentinel-1 SAR data. AIS transceiver sends data every 2 to 10 seconds. These data mainly consist in a positional accuracy (up to 0.0001 minutes precision), and the course over ground (relative to true north to 0.1°). For a given SAR image, one can interpolate AIS data from the associated acquisition time. Thus it is possible to know the precise location of the ships in the SAR image and the related information (in our case, length and type). The footprint of the ship is obtained by thresholding the SAR image in the area where it is located (the brightest pixel of the image). Since the database is very unbalanced in terms of class distribution, a strategy is also proposed in order to enlarge the training set with translations and rotations, which is a standard procedure for database enlargement (a.k.a. data augmentation). Concurrently to our work, a similar database has been proposed in [29]. We also evaluate our framework with this dataset (see Section 3.5).

In our experiments, we consider a dataset composed of 18,894 raw SAR images of size 400×400 pixels with a 10 m resolution. The polarization of the images are either HH (proportion of horizontally transmitted waves which return horizontally) or VV (proportion of vertically transmitted waves which return vertically). Polarization has a significant effect on SAR backscatter. However, our goal is to allow us to process any Sentinel-1 SAR images. We thus consider any HH and VV polarized image without prior information on the type of polarization. Each image is accompanied with the incidence angle since it impacts the backscatter intensity of the signal. For the proposed architecture, the input is a 2-band image (backscatter intensity and incidence angle). Thus we did not use any pre-trained network since we assume that they can not handle such input data. We rely on Automatic Identification System (AIS) to extract images that contain a ship in their center. AIS also provides us with information about the ship type and length. As stated before, AIS may have been corrupted (e.g. with spoofing), when creating the database, we only consider ships that responds to the two following criteria; *(i)* their type is clearly defined (i.e. they belongs to the retained classes), *(ii)* their length is greater than 0 and smaller than 400 meters (the largest ship in the world). Besides, the SAR images we selected were acquired over European waters, where we expect AIS data to be of higher-quality compared with other maritime areas.

The dataset is strongly imbalanced, amongst the 5 classes (*Tanker*, *Cargo*, *Fishing*, *Passenger* and *Tug*), the *Cargo* is the most represented (10,196 instances), while the *Tug* is the less represented (only 444 instances). The class distribution is detailed in Figure 2 and Table 1. The length distribution shows that *Tanker*, *Cargo*, and *Passenger* ships have similar length distributions. *Fishing* ships have relatively small lengths, while *Tug* ship length are intermediate.

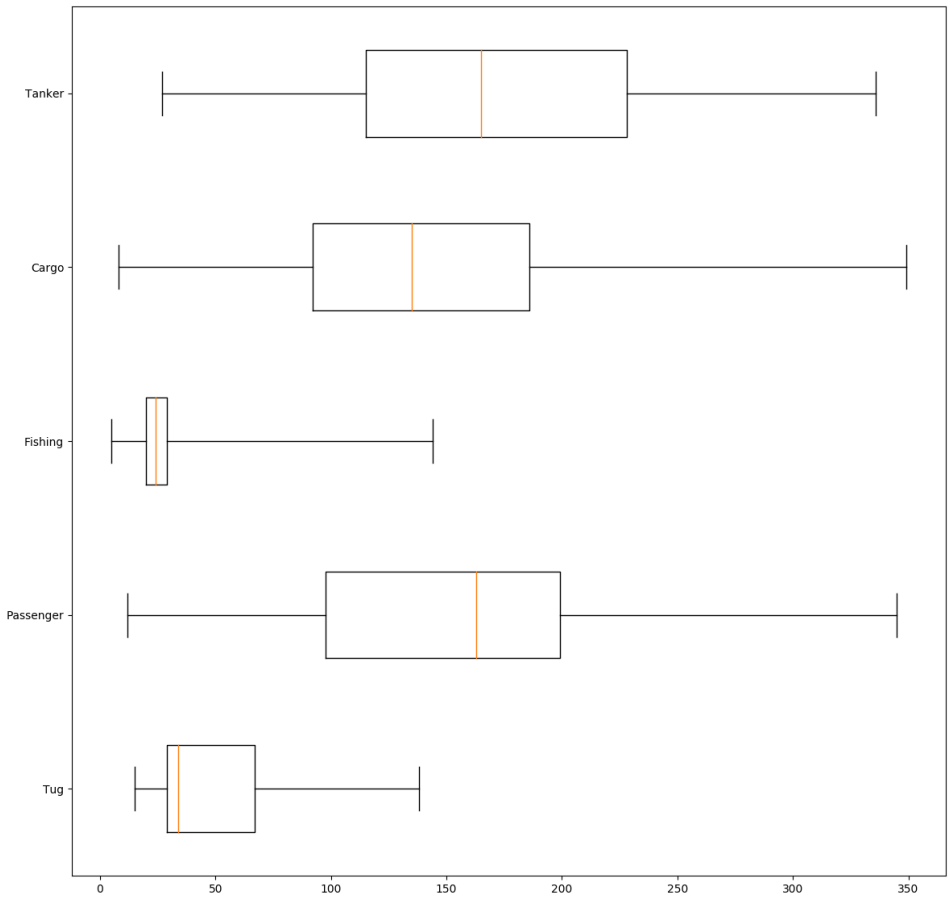
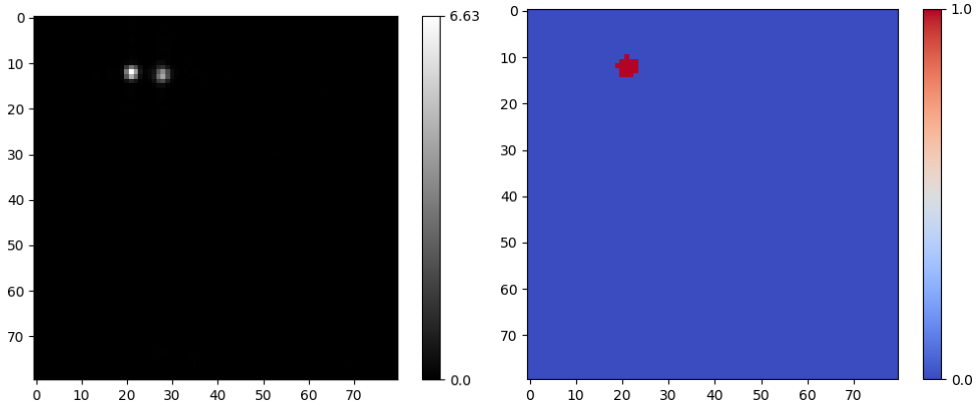


Figure 2. Boxplot of length distribution for each class in our dataset. Length data are given in meters.

	Tanker	Cargo	Fishing	Passenger	Tug
Number of samples	4737	10196	2664	1071	444
Length mean (m)	168.4	146.5	26.6	153.5	47.3
Length standard deviation (m)	64.6	60.8	12.2	68.9	26.0
Number of augmented samples (4 classes)	263	0	2336	3929	-
Number of augmented samples (5 classes)	0	0	1336	2929	3556

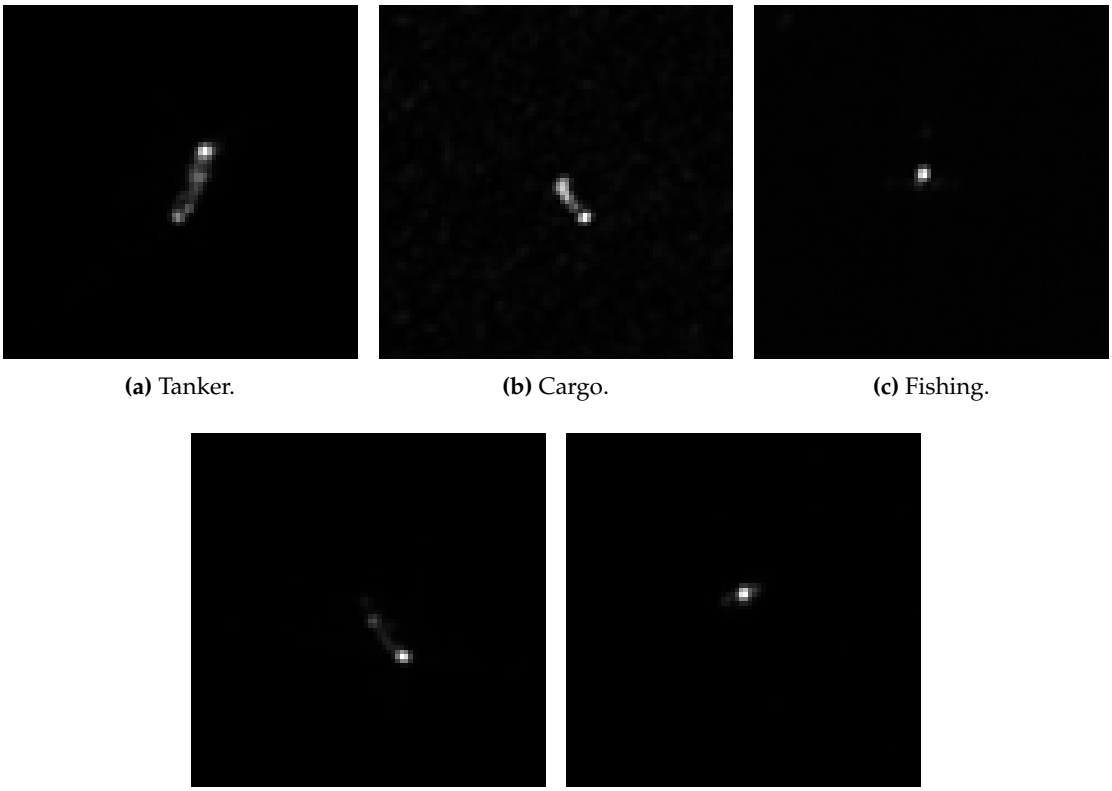
Table 1. Length distribution and number of samples for each class in our dataset.

To account for class imbalance [30], we apply data augmentation with translations and rotations. We first perform a rotation of a random angle centered on the brightest pixel of the SAR image (the center of the ship), and then perform a random translation. The same transformation is applied to the incidence angle image. The images employed to train the networks are of size 80×80 pixels. They contain ships (not necessarily in their center, see Figure 4). The ship footprint groundtruth is generated by thresholding the SAR image since we precisely know the location of the ship (i.e. it is the brightest pixel of the SAR image, see Figure 3). The obtained footprint is not perfect (see Figure 3b) but was shown to be sufficient to train the network. Let us note that a CFAR approach could have been employed in order to extract more precisely the ship footprint [11]. But since our goal is not to detect ships, a coarse ship footprint is sufficient. We considered 2 configurations for the databases; a 4-classes database, employed to compare our baseline to other state-of-the-art approaches (namely MLP and R-CNN), and 5-classes database in order to evaluate how our network responds with more classes. Each database is composed of 20,000 images of 80×80 pixels, with the same amount of samples per class (5,000 per class for the 4-classes database, and 4,000 per class for the 5-classes database). The networks are trained with 16,000 images and the remaining 4,000 are used for validation. Throughout the data augmentation process, we ensure that images can be seen either in the training or validation set, but not in both. Ships with no AIS signal are not considered in our dataset (neither to train or evaluate our model), since our strategy to build the dataset relies on matching AIS signal with SAR imagery. However, once a model has been trained, it can perform in an operational settings to detect ships with no AIS (it is indeed one of our long-term goals).



(a) SAR image. **(b)** Ship footprint.

Figure 3. Example of SAR image (with backscatter intensity) and associated ship footprint (best viewed in color).



(a) Tanker. **(b)** Cargo. **(c)** Fishing.

(d) Passenger. **(e)** Tug.

Figure 4. Examples of SAR image (with backscatter intensity) for each ship type of the collected database.

3. Results

We run all numerical experiments on a PC with a NVIDIA GTX 1080 Ti, an Intel Xeon W-2145 CPU 3.70GHz and 64GB RAM (with a Keras [31] implementation). We evaluate the proposed framework with respect to other popular deep learning-based solutions. We first consider a Multi-Layer Perceptron (MLP) [32] with only one hidden layer with 128 hidden units. The MLP is the most simple network that can be proposed for the desired task and can be a good basis in order to evaluate the performance of our network. We also designed a R-CNN (Regions with CNN features) [33] network in order to extract ship bounding boxes along with classification. Even if the R-CNN-based bounding boxes do not allow to precisely measure the ship length, they can provide a good basis for its estimation. R-CNN is a state-of-the-art algorithm for object detection and classification [33]. Thus it is worth being compared with our proposed model. The R-CNN has a very simple architecture presented in Figure 5. The networks are trained using 16,000 images from the augmented dataset and the remaining 4,000 images are used for validation.

The evaluation of the models is performed using several metrics. The classification task is assessed through the confusion matrix, giving, for each class and overall, several metrics. The Intersection over Union (IoU or Jaccard index) [34] measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets. It has been designed for the evaluation of object detection. The F-score is the harmonic mean of precision and recall, it reaches its best value at 1 and worst at 0. The Kappa coefficient [35] (κ) is generated from a statistical test to evaluate the accuracy of a classification. Kappa essentially evaluates how well the classification performs as compared to just randomly assigning values (i.e. did the classification do better than randomness?) The Kappa coefficient can range from -1 to 1. A value of 0 (respectively -1 or 1) indicates that the classification is no better (respectively worse or better) than a random classification.
For the length estimation task, the mean error (and its standard deviation) are employed. For a ship k, the length error is defined as $e_k = l_{k,pred} - l_{k,true}$, where $l_{k,pred}$ is the predicted length and $l_{k,true}$ is the actual length. The mean error m_{err_length} (respectively the standard deviation $stdev_{err\_length}$), is the mean (respectively the standard deviation) of all the e_k. We further refer mean error to $m_{err\_length} \pm stdev_{err\_length}$.

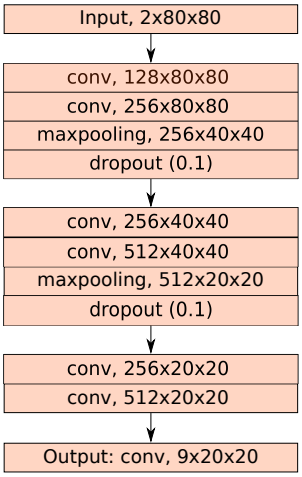


Figure 5. R-CNN architecture considered for ship classification.

3.1. MLP model

For a 80×80 image, the MLP runs at 2,000 frames per second. The whole training takes about one hour. The testing takes less than a minute. It produces very poor results. Indeed, the overall accuracy

for classification is **25%**, which means that the classifier assigns the same class to all the images (see Table 2). The length estimation is also rather inaccurate, the ship length being underestimated with a very large standard deviation (mean error: **-7.5 m** $\pm$ **128 m**).

Confusion matrix					
Prediction / Ground Truth	Tanker	Cargo	Fishing	Passenger	Precision
Tanker	1000	0	0	0	100.0
Cargo	1000	0	0	0	0.0
Fishing	1000	0	0	0	0
Passenger	1000	0	0	0	0.0
Recall	25.0	-	-	-	

Accuracy metrics					
Label	Tanker	Cargo	Fishing	Passenger	Overall
IoU	25.0	0.0	0.0	0.0	6.25
F-Score	40.0	-	-	-	10.00
Accuracy	25.0	75.0	75.0	75.0	25.00
κ	0.0	0.0	0.0	0.0	0.25

Table 2. Confusion matrix and accuracy metrics for the MLP with 4 classes.

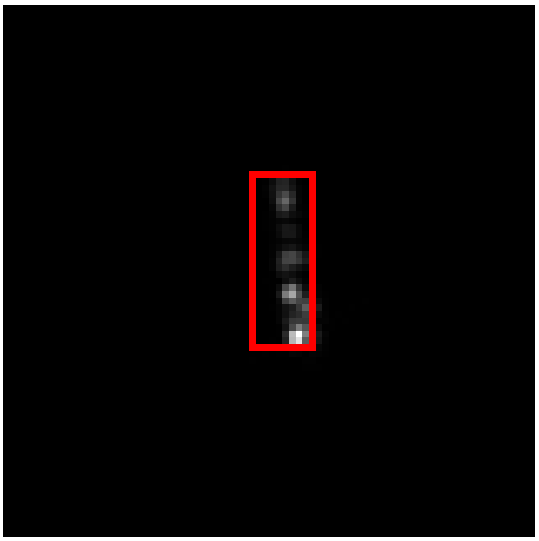
3.2. R-CNN model

For a 80×80 image, the R-CNN runs at 333 frames per second. The whole training takes about 6.5 hours and the testing about a minute. It produces better results than the MLP. The network estimates the 4 corners of the bounding box. As the groundtruth for bounding boxes is obtained from the ship footprint extracted by thresholding the SAR image, it might not be well-defined (see Figure 6c and 6d). In Figure 6c, the bounding box is well centered on the ship, but has a wrong size. In Figure 6d, the bounding box is also not well-sized, and accounts for the brightest part of the ship. We recall that the detection task is not our main objective, but rather regarded as a means to better constrain the training of the models. The R-CNN have a classification overall accuracy of **89.29%**. Several other metrics are presented in Table 3.

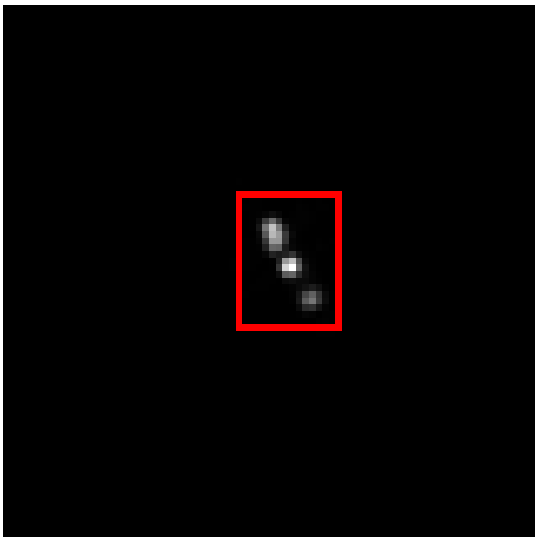
Confusion matrix					
Prediction / Ground Truth	Tanker	Cargo	Fishing	Passenger	Precision
Tanker	845	97	3	33	86.40
Cargo	98	787	24	77	79.82
Fishing	2	9	891	51	93.49
Passenger	8	15	1	961	97.56
Recall	88.67	86.67	96.95	85.65	

Accuracy metrics					
Label	Tanker	Cargo	Fishing	Passenger	Overall
IoU	77.81	71.09	90.83	83.86	80.90
F-Score	87.52	83.10	95.19	91.22	89.26
Accuracy	93.82	91.80	97.69	95.26	89.29
κ	0.83	0.78	0.94	0.88	0.88

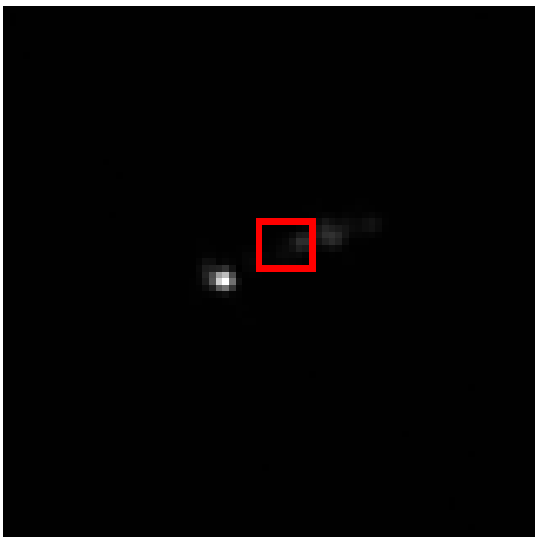
Table 3. Confusion matrix and accuracy metrics for the R-CNN with 4 classes.



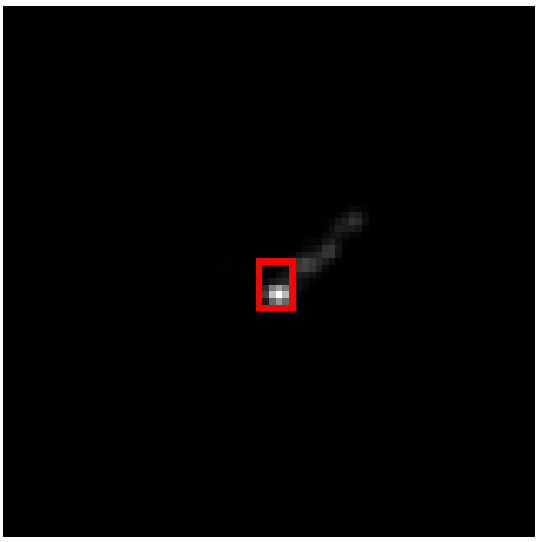
(a) Accurate bounding box superimposed to the SAR image. The ship type (Passenger) is well predicted.



(b) Accurate bounding box superimposed to the SAR image. The ship type (Tanker) is well predicted.



(c) Inaccurate bounding box superimposed to the SAR image. The ship type predicted (Fishing) is not the good one (Cargo).

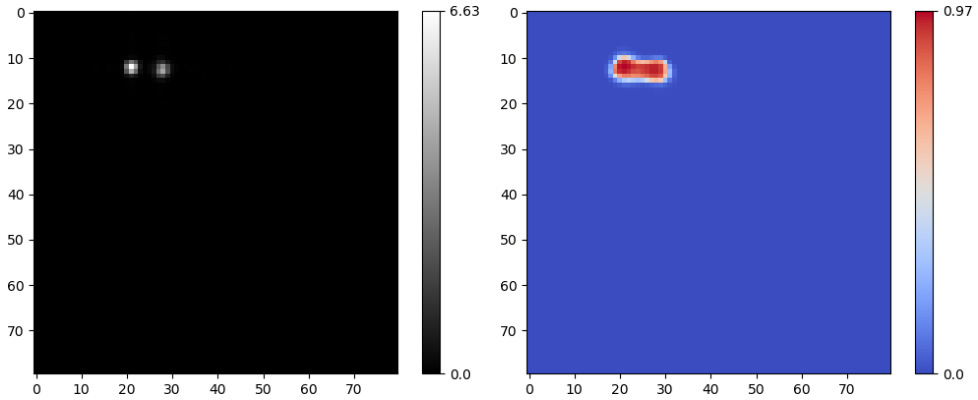


(d) Inaccurate bounding box superimposed to the SAR image. The ship type (Tanker) is well predicted.

Figure 6. Illustration of detection and classification performance of the evaluated R-CNN model: each subpanel depicts a SAR image with superimposed the detected bounding box.

3.3. Our network

For a 80×80 image, our method can run at 250 frames per second. The whole training takes about 9 hours and the testing about a minute. With an overall accuracy and a mean F-score of **97.2%**, the proposed multi-task architecture significantly outperforms the benchmarked MLP and R-CNN models. We report in Table 4 the confusion matrix and additional accuracy metrics. Interestingly, classification performances are relatively homogeneous across ship types (mean accuracy above 92% for all classes). Tankers involve the greater misclassification rate with some confusion with cargo.
Regarding length estimation performance, our framework achieves very promising results. The length is slightly over-estimated (mean error: **4.65 m** ± **8.55 m**), which is very good regarding the spatial resolution of the Sentinel-1 SAR data (10m/pixel). To our knowledge, this is the first demonstration that reasonably-accurate ship length estimates can be derived from SAR images using learning-based schemes, whereas previous attempts using model-driven approaches led to much poorer performance. Overall, the results of the classification and length estimation tasks for all the tested architectures are summarized in Table 6.



(a) SAR image. **(b)** Detection result.

Figure 7. Example of detection output for the considered multi-task architecture: left, SAR image (with backscatter intensity) used as input; right, output of the detection module of the considered architecture.

Confusion matrix					
Prediction / Ground Truth	Tanker	Cargo	Fishing	Passenger	Precision(%)
Tanker	985	11	0	4	98.5
Cargo	65	907	12	16	90.7
Fishing	0	2	998	0	99.8
Passenger	0	0	0	1000	100.0
Recall(%)	93.81	98.59	98.81	98.04	

Accuracy metrics					
Label	Tanker	Cargo	Fishing	Passenger	Overall
IoU(%)	92.49	89.54	98.62	98.04	94.67
F-Score(%)	96.10	94.48	99.30	99.01	97.22
Accuracy(%)	98.00	97.35	99.65	99.50	97.25
κ	0.95	0.93	0.99	0.99	0.97

Table 4. Confusion matrix and accuracy metrics for the proposed network with 4 classes.

We also train our model with 5 classes, and it confirms that our framework performs well. The length is slightly over-estimated (mean error: **1.93 m** ± **8.8 m**) and the classification is also very good (see

Table 5). Here, we still report some light confusion for Tanker and Cargo classes. The accuracy metrics are slightly worse that the 4-class model but still report an overall accuracy and a mean F-score of **97.4%**.

Confusion matrix						
Ground Truth \ Prediction	Tanker	Cargo	Fishing	Passenger	Tug	Precision(%)
Tanker	771	28	0	1	0	96.38
Cargo	60	732	3	3	2	91.50
Fishing	0	1	799	0	0	99.88
Passenger	3	1	0	796	0	99.50
Tug	0	0	0	0	800.0	100.00
Recall(%)	92.45	96.06	99.63	99.50	99.75	

Accuracy metrics						
Label	Tanker	Cargo	Fishing	Passenger	Tug	Overall
IoU(%)	89.34	88.19	99.50	99.00	99.75	95.16
F-Score(%)	94.37	93.73	99.75	99.50	99.88	97.44
Accuracy(%)	97.70	97.55	99.90	99.80	99.95	97.45
κ	0.93	0.92	1.00	0.99	1.00	0.97

Table 5. Confusion matrix and accuracy metrics for the proposed network with 5 classes.

Architecture	Length mean error (m)	Classification overall accuracy (%)
MLP	-7.50 ± 128	25.00
R-CNN	-	88.57
Our network	4.65 ± 8.55	97.25

Table 6. Results of all the tested architectures for the classification (4 classes) and length estimation.

We further analyse the proposed scheme and the relevance of the multi-task setting, compared with task-specific architectures. To this end, we perform an ablation study and train the proposed architecture using (i) length estimation loss only, (ii) classification loss only, (iii) the combination of length estimation and classification losses (i.e., without the detection loss). We report in Table 7 the resulting performances compared to those of the proposed end-to-end learning strategy. Regarding the classification issue, combined losses result in an improvement of about 1.3% (above 25% in terms of relative gain). The improvement is even more significant for length estimation with a relative gain in the mean error of about 36%. Interestingly, we note that the additional use of the detection loss also greatly contributes to the improvement of length estimation performance (mean error 2.85m without using the detection loss during training vs. 1.93m when using jointly detection, classification and length estimation losses). As an illustration of the detection component of the proposed architecture, we illustrate in Figure 7 a detection result. As mentioned above, the thorough evaluation of this detection model is not the main objective of this study. Furthermore, without any precise ship footprint groundtruth, is is impossible to quantitatively evaluate the performance of the network for this specific task. Let us recall that the detection task has been widely addressed in the literature [14–16]. Overall, this complementary evaluation supports the idea that Neural Network architectures for SAR image analysis may share some low-level task-independent layers, whose training can highly benefit from the existence of multi-task datasets.

3.4. Application to a full SAR image

We illustrate here an application of the proposed approach to a real SAR image acquired on April 4, 2017 in Western Brittany, France. We proceed in several steps as follows. First, a CFAR-based ship detector is applied. Then, for each detected ship, we apply the trained deep network model to predict

	Length mean error (m)	Classification overall accuracy (%)
(i)	3.07 ± 9.0	-
(ii)	-	96.10
(iii)	2.85 ± 8.9	97.50
Full network	1.93 ± 8.8	97.45

Table 7. Ablation study, performance of the network for different scenarii: (i) only length estimation, (ii) only classification, (iii) length estimation and classification without detection.

the ship category and its length. For illustration purposes, we report in Figure 8 the detected ships which could be matched to AIS signals.

For the considered SAR image, among the 98 ships detected by the CFAR-based ship detector, 66 ships have their length documented and 69 ships belong to one of the 5 proposed classes after AIS matching. We may point out that the *Tug* class is not represented. We report classification and length estimation performance in Table 8. Ship classification performance is in line with the performance reported above. Regarding length estimation, the mean error **14.56 m** $\pm$ **39.98 m** is larger than that reported for the groudtruthed dataset. Still, this error level is satisfactory given the pixel resolution of 10 m of the SAR image. Let us note that, given the limited samples available, the standard deviation is not fully relevant here. While a special care was undertaken for the creation of our SAR-AIS dataset, this application to a single SAR image exploits the raw AIS data. AIS data may be significanty corrupted, which may partially explain these differences.

Confusion matrix					
Ground Truth \ Prediction	Tanker	Cargo	Fishing	Passenger	Precision(%)
Tanker	13	3	0	2	72.22
Cargo	5	30	0	0	85.71
Fishing	0	0	14	0	100.00
Passenger	0	0	0	1	100.00
Recall	72.22	90.91	100.00	33.33	

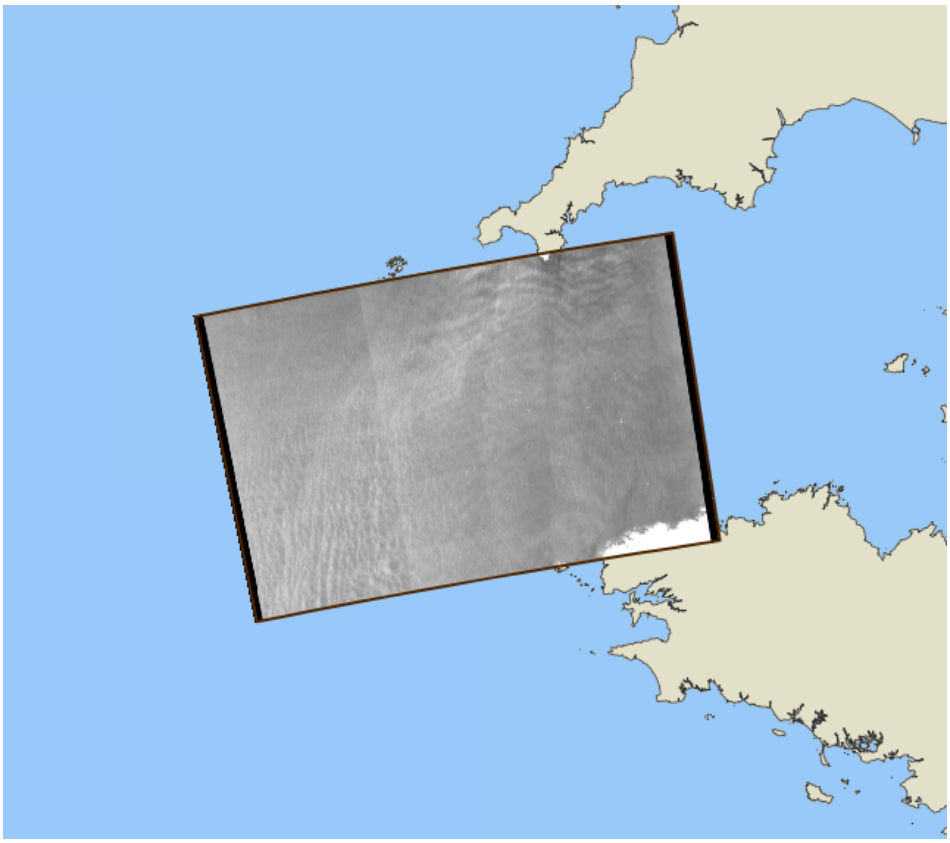
Accuracy metrics					
Label	Tanker	Cargo	Fishing	Passenger	Overall(%)
IoU	56.52	78.95	100.00	33.33	67.20
F-Score	72.22	88.24	100.00	50.00	77.61
Accuracy	85.29	88.24	100.00	97.06	85.29
κ	0.62	0.76	1.00	0.49	0.85

Table 8. Classification scores of the proposed network on small patches extracted from a SAR scene.

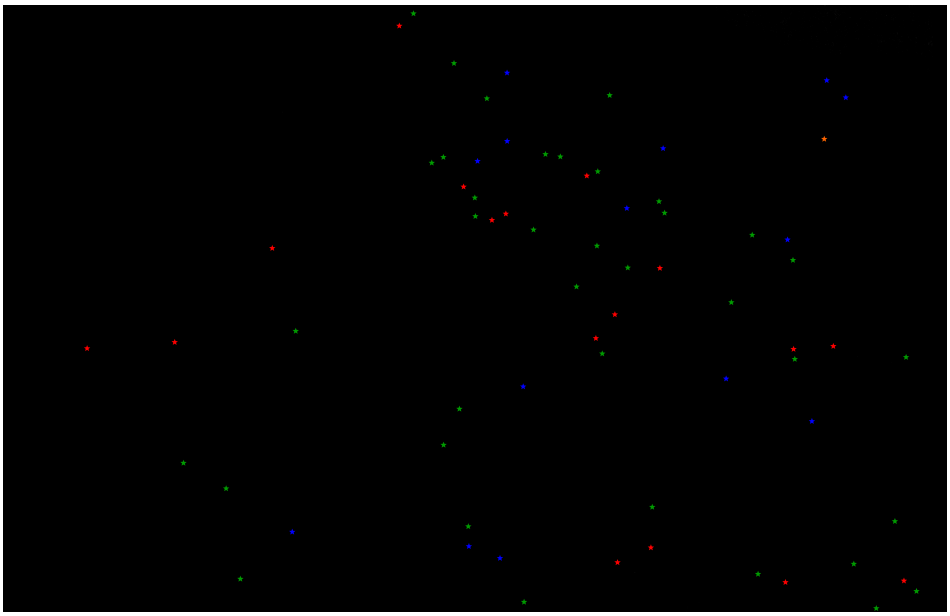
3.5. Application to the OpenSARShip dataset

The OpenSARShip dataset [29] has recently been made available to the community. We report here the results obtained with our framework when applied on this dataset. This dataset comprises SAR data with different polarization characteristics and also includes ship categories. With a view to ease the comparison with the previous results, we focus on SAR images that are in VV polarization and the ship categories considered, which leads to considering the following four categories; tanker, cargo, fishing and passengers. Overall, we considered a dataset of 5,225 ships (80% were employed for training and 20% for testing). In the OpenSARShip dataset, the classes are not equally represented. We report classification and length estimation performance in Table 9. We also evaluate the performance of the model trained on our dataset and applied on OpenSARShip dataset and conversely.

The results show that our model produces good results when trained and tested on the same database. However, the results do not transfer from one dataset to an other. We suggest that this



(a) Image location



(b) SAR image with ships identified from the matching of CFAR-based detection and AIS. ⋆ Tanker, ⋆ Cargo, ⋆ Fishing, ⋆ Passenger (best viewed in color).

Figure 8. SAR image acquired on April 4, 2017 in Western Brittany, France.

may relate to differences in the maritime traffic and environment between Europe (our dataset) and Asia (OpenSARShip dataset). The comparison to previous work on OpenSARShip dataset is not straightforward. For instance, [36] considers only a three-class dataset (*Tanker*, *Cargo* and *Other*). The reported accuracy score (76%) is lower than our 87.7% accuracy score for the considered 5-class dataset. We may also emphasize that [36] does not address ship length estimation.

		Train	
		Ours	OpenSARShip
Test	Ours	97.45	22.18
	OpenSARShip	34.05	87.71

Classification overall accuracy (%)

		Train	
		Ours	OpenSARShip
Test	Ours	1.93 ± 8.8	56.78 ± 314.78
	OpenSARShip	-102.51 ± 123.94	-0.23 ± 11.04

Length mean error (m)

Table 9. Comparison of the results of the network on our database and on the OpenSARShip database.

4. Discussion

The reported results show that a dedicated architecture is necessary for ship classification and length estimation, while state-of-the art architectures failed to achieve satisfying performances. The MLP is sufficient for ship detection on SAR images (from a visual assessment). But this should not be considered as a good result since we only have (positive) examples of ships in our database (no negative samples, so we can not assess the false positives). Thus, the network only learns a thresholding and can not discard a ship from other floating objects (e.g. icebergs). Indeed, iceberg detection and discrimination between iceberg and ship are specific research questions [37,38]. Overall, the performance of the MLP stresses the complexity of the classification and length estimation tasks.
In terms of classification accuracy, the R-CNN performs better than the MLP, with an overall accuracy of **88.57%**. These results support the proposed architecture with three task-specific networks which share a common low-level network. The latter is interpreted as a feature extraction unit which the task-specific networks rely on.
Compared to the state-of-the art architectures (MLP and R-CNN), our model produces better results for ship classification and length estimation from Sentinel-1 SAR images with only few confusions between classes. A multi-task architecture is well adapted for simultaneous ship classification and length estimation. Our model also performs well when a new class is added (e.g. Tug). Furthermore, adding a detection task (even with a coarse ground truth) tends to improve the length estimation. Our experiments also show that the learnt models do not transfer well from a dataset to an other. We suggest that this may relate to differences in the characteristics of the maritime traffic and/or marine environment. Future work should further explore these aspects for the application of the proposed model worldwide.

5. Conclusion

In this paper, a multi-task neural network approach was introduced. It jointly addresses the detection, classification and length estimation of ships in Sentinel-1 SAR images. We exploit synergies between AIS and Sentinel-1 to automatically build reference datasets for training and evaluation purposes, with the ultimate goal of relying solely on SAR imagery to counter lack or corruption of AIS information that correspond to illegal activities. While the polarization type has a significant effect on SAR backscatter, we were able to train a model which jointly processes HH or VV polarisation without prior information on the type of polarisation. Our results support the assumption that HH and VV polarizations share common image features and that differences in backscatter distributions can be handled through an appropriate parameterization of the network.

Regarding the considered architecture, a mutual convolutional branch transforms raw inputs into meaningful information. Such information is fed into three task-specific branches. Experimental evaluation shows improvement over standard MLP or R-CNN. Ship detection cannot be totally assessed, but a visual inspection supports the relevance of this detection stage. Besides, it was shown to significantly contribute to improved performance of the classification and length estimation components. Overall, we report promising performance for ship classification (above 90% of correct classification) and length estimation (relative bias below 10%). Considering a residual architecture appears as a critical feature to reach good length estimation performance, but this would require further investigation.

Future work may further investigate the training and evaluation of the detection stage. The automation of the matching process between AIS data and SAR images has the potential for significantly increasing the size and diversity of the training and evaluation datasets. This may provide new avenues to address generalization and transfer issues between geographic areas pointed out in our results. Furthermore, while SAR imagery less affected by weather conditions than optical imagery, a specific analysis of the impact of weather conditions onto identification performance would also be of interest. Finally, the specificity of the SAR imagery would call for dedicated operations, while our network relies on standard techniques issued from computer vision.

References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436.
2. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine* **2016**, *4*, 22–40.
3. Zhu, C.; Zhou, H.; Wang, R.; Guo, J. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Transactions on Geoscience and Remote Sensing* **2010**, *48*, 3446–3456.
4. Antelo, J.; Ambrosio, G.; Gonzalez, J.; Galindo, C. Ship detection and recognition in high-resolution satellite images. IEEE International Geoscience and Remote Sensing Symposium, 2009, pp. IV–514.
5. Chen, H.; Gao, X. Ship recognition based on improved forwards-backwards algorithm. International Conference on Fuzzy Systems and Knowledge Discovery, 2009, pp. 509–513.
6. Wang, Q.; Gao, X.; Chen, D. Pattern recognition for ship based on Bayesian networks. International Conference on Fuzzy Systems and Knowledge Discovery, 2007, pp. 684–688.
7. Tang, J.; Deng, C.; Huang, G.B.; Zhao, B. Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. *IEEE Transactions on Geoscience and Remote Sensing* **2015**, *53*, 1174–1185.
8. Fablet, R.; Bellec, N.; Chapel, L.; Friguet, C.; Garello, R.; Gloaguen, P.; Hajduch, G.; Lefèvre, S.; Merciol, F.; Morillon, P.; others. Next Step for Big Data Infrastructure and Analytics for the Surveillance of the Maritime Traffic from AIS & Sentinel Satellite Data Streams. BiDS'2017-Conference on Big Data from Space, 2017, pp. 1–4.
9. Nguyen, D.; Vadaine, R.; Hajduch, G.; Garello, R.; Fablet, R. Multi-task Learning for Maritime Traffic Surveillance from AIS Data Streams. *CoRR* **2018**.
10. Iphar, C.; Napoli, A.; Ray, C. Data quality assessment for maritime situation awareness. ISSDQ 2015-The 9th International Symposium on Spatial Data Quality, 2015, Vol. 2.
11. Liao, M.; Wang, C.; Wang, Y.; Jiang, L. Using SAR images to detect ships from sea clutter. *IEEE Geoscience and Remote Sensing Letters* **2008**, *5*, 194–198.
12. An, W.; Xie, C.; Yuan, X. An improved iterative censoring scheme for CFAR ship detection with SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing* **2014**, *52*, 4585–4595.
13. Pelich, R.; Chini, M.; Hostache, R.; Matgen, P.; Lopez-Martinez, C.; Nuevo, M.; Ries, P.; Eiden, G. Large-Scale Automatic Vessel Monitoring Based on Dual-Polarization Sentinel-1 and AIS Data. *Remote Sensing* **2019**, *11*, 1078.
14. Liu, Y.; Zhang, M.h.; Xu, P.; Guo, Z.w. SAR ship detection using sea-land segmentation-based convolutional neural network. International Workshop on Remote Sensing with Intelligent Processing, 2017, pp. 1–4.

15. Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. International Workshop on Remote Sensing with Intelligent Processing, 2017, pp. 1–4.
16. Khesali, E.; Enayati, H.; Modiri, M.; Aref, M.M. Automatic ship detection in Single-Pol SAR Images using texture features in artificial neural networks. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **2015**, *40*, 395.
17. Schwegmann, C.P.; Kleynhans, W.; Salmon, B.P.; Mdakane, L.W.; Meyer, R.G. Very deep learning for ship discrimination in synthetic aperture radar imagery. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2016, pp. 104–107.
18. Bedini, L.; Righi, M.; Salerno, E. Size and Heading of SAR-Detected Ships through the Inertia Tensor. Multidisciplinary Digital Publishing Institute Proceedings, 2018, Vol. 2, p. 97.
19. Stasolla, M.; Greidanus, H. The exploitation of Sentinel-1 images for vessel size estimation. *Remote Sensing Letters* **2016**, *7*, 1219–1228.
20. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, pp. 1097–1105.
22. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* **2014**.
23. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-normalizing neural networks. Advances in Neural Information Processing Systems, 2017, pp. 971–980.
24. Drozdzal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; Pal, C. The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*; Springer, 2016; pp. 179–187.
25. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters* **2017**, *14*, 778–782.
26. Geng, J.; Fan, J.; Wang, H.; Ma, X.; Li, B.; Chen, F. High-resolution SAR image classification via deep convolutional autoencoders. *IEEE Geoscience and Remote Sensing Letters* **2015**, *12*, 2351–2355.
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
28. Tieleman, T.; Hinton, G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. Technical report, 2012.
29. Huang, L.; Liu, B.; Li, B.; Guo, W.; Yu, W.; Zhang, Z.; Yu, W. OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2017**, *11*, 195–208.
30. Kellenberger, B.; Marcos, D.; Tuia, D. Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment* **2018**, *216*, 139–153.
31. Chollet, F.; others. Keras. https://keras.io, 2015.
32. Kruse, R.; Borgelt, C.; Klawonn, F.; Moewes, C.; Steinbrecher, M.; Held, P. Multi-layer perceptrons. In *Computational Intelligence*; Springer, 2013; pp. 47–81.
33. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.
34. Jaccard, P. The distribution of the flora in the alpine zone. 1. *New phytologist* **1912**, *11*, 37–50.
35. Cohen, J. A coefficient of agreement for nominal scales. *Educational and psychological measurement* **1960**, *20*, 37–46.
36. Seungryong, K.; Jeongju, B.; Chan-Su, Y. Satellite image-based ship classification method with sentinel-1 IW mode data. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2019, pp. 1300–1301.
37. Power, D.; Youden, J.; Lane, K.; Randell, C.; Flett, D. Iceberg detection capabilities of RADARSAT synthetic aperture radar. *Canadian Journal of Remote Sensing* **2001**, *27*, 476–486.
38. Bentes, C.; Frost, A.; Velotto, D.; Tings, B. Ship-iceberg discrimination with convolutional neural networks in high resolution SAR images. EUSAR 2016: 11th European Conference on Synthetic Aperture Radar, Proceedings of. VDE, 2016, pp. 1–4.