MULTI-TASK DEEP LEARNING FROM SENTINEL-1 SAR: SHIP DETECTION, CLASSIFICATION AND LENGTH ESTIMATION

C. Dechesne [1], *S. Lefèvre* [2], *R. Vadaine*[3], *G. Hajduch*[3], *R. Fablet* [1]

[1] IMT Atlantique – Lab-STICC, UMR CNRS 6285, Brest, FR
[2] Université Bretagne Sud – IRISA, Vannes, FR [3] Collecte Localisation Satellites, Brest, FR

ABSTRACT

The detection of inshore and offshore ships is an important issue in both military and civilian fields. It helps monitoring fisheries, managing maritime traffics, ensuring safety of coast and sea, etc. In operational contexts, ship detection is traditionally performed by a human observer who identifies all kind of ships from visual analysis on remote sensing images. Such a task is very time consuming and cannot be conducted at a very large scale, while Sentinel-1 SAR data now provides regular, worldwide coverage. Meanwhile, with the emergence of GPUs, deep learning methods are now established as a state-of-the-art solution for computer vision, replacing human intervention in many contexts. They have been shown to be adapted for ship detection and recognition, most often with very high resolution SAR or optical imagery. In this paper, we go one step further and propose a deep neural network for the detection, classification and length estimation of ships from SAR Sentinel-1 data. We benefit from synergies between AIS (Automatic Identification System) and Sentinel-1 data to build significant training datasets. We then design a multi-task neural-network architecture composed of one joint convolutional network connected to 3 networks dedicated to the different tasks: ship detection, classification and length estimation. Experimental assessment showed our network provides satisfactory results, with accurate classification and length estimation.

***Index Terms*—** Deep neural network, Sentinel-1 SAR images, Ship detection, Classification, Length estimation, Multi-task learning

1. INTRODUCTION

Deep learning is considered as one of the major breakthrough related to big data and computer vision [8]. It has become very popular and successful in many fields including remote sensing [14]. Deep learning are representation-learning methods providing multiple levels of representation. When applied on visual data such as images, it is usually achieved by means of convolutional neural networks, that consists of multiple layers (such as convolution, pooling, fully connected and normalization layers) that aim to transform original data (raw input) into higher-semantics representation. With the composition of enough such operations, very complex functions can be learned. For classification tasks, higher 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, accurate, algorithms for object detection are thus sought to allow computers to perform such tasks, at a much large scale than humans can achieve.

Sentinel-1 SAR images are well adapted for ship detection. Almost all coastal zones and shipping routes are covered by Interferometric Wide Swath Mode (IW), while 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 deep learning models. AIS provides meaningful and relevant information about ships (such as position, type, length, rate of turn, speed over ground, etc.). Combining these two data sources could ease accurate detection and estimation of ship parameters from SAR images, which remains a very challenging task. Indeed, detecting inshore and offshore ships has an important significance 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 not able to cope with big data issues. On the contrary, the availability of satellite data such as Sentinel-1 SAR makes possible the efficient and accurate ship detection.

Among existing methods for ship detection from SAR images, constant false alarm rate (CFAR) based methods have been widely used to detect ships in the sea [9, 1]. The advantage of CFAR-based methods is their reliability and high efficiency. As the choice of features has an impact on the performance of discrimination, deep neural networks took the lead thanks to their ability to extract (or learn) features that are richer than hand-crafted features. In [10], 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 [4] with good detection performances. In [6], a method categorizing ship targets from SAR images using texture features in artificial neural networks (TF-ANN) was proposed. The TF-ANN method selects an appropriate texture feature for SAR images and uses the feature as the input of neural network to extract ship pixels from sea ones. [12] employed highway network for ship detection in SAR images and achieved good results, especially in detecting correctly false positive. These state-of-the-art approaches focused on ship detection in SAR images. In this paper, we address not only to detect ships from SAR images, but also the recognition of ship types and length estimation, which to our knowledge has not been dealt with before.

2. PROPOSED APPROACH

2.1. Creation of groundtruthed datasets

With a view to implementing deep learning strategies, we first address the creation of groundtruthed datasets from the synergy between AIS data and Sentinel-1 SAR data. AIS data are in interpolated in order to know the ship location when the SAR image have been captured. Thus it is possible to know the precise location of the ship in the SAR image and its 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.

2.2. 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 Fig. 1). The first part is a convolutional network made of 5 layers. It is a mutual network that is used for all the 3 tasks. It is followed by the task-oriented branches. For the detection task, the output consists in a pixel-wise probability of presence of ship. It is composed of 4 convolutional layers and 1 fully connected layer. For the classification task, we consider 4 ship classes (Cargo, Tanker, Fishing and Passenger). The branch is composed of 4 convolutional layers and 2 fully connected layers. The last task is related to the length estimation. The related branch is composed of 4 convolutional layers and 5 fully connected layers.

Such settings are commonly employed in deep learning methods [11]. All the activations of the convolutional layers and fully connected layers are ReLu [7]. 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.

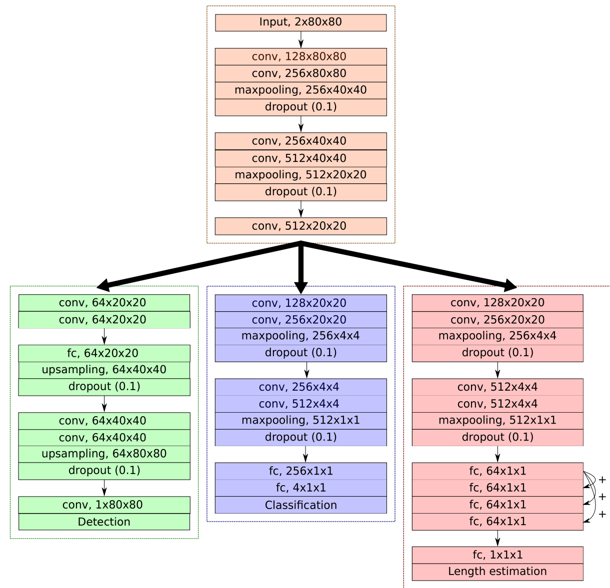


Fig. 1: Proposed multi-task architecture for ship detection, classification and length estimation from Sentinel-1 SAR.

2.3. Loss functions

2.3.1. Detection

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

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

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.

2.3.2. Classification

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, defined by:

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

where N is the number of samples, n_c is the number of classes (here, $n_c = 4$), $y_{o,c}$ 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 c.

2.3.3. Length

In the length estimation network, the 4 fully-connected layers of shape ($64\times1\times1$) are connected to each other (see Fig. 2). The idea is to propagate the difference between the first layer

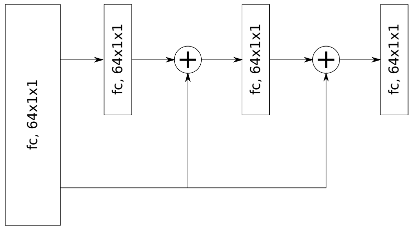


Fig. 2: Difference propagation flowchart in the fully-connected layers.

and the current layer and is related to residual learning [3]. We use here the mean squared error defined as

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

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

2.3.4. End-to-end training

We define the loss function of the whole network as

$$L = L_{det} + L_{class} + L_{length}. \tag{4}$$

Our network is trained end-to-end using RMSProp optimizer [13]. 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.

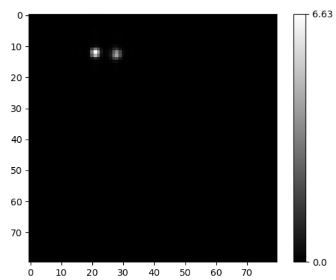
3. DATA

In our experiments, we consider a dataset composed of $18,894$ 400×400 SAR images having a 10 m resolution. Each image is accompanied with the incidence angle since it impacts the backscatter intensity of the signal. We rely on Automatic Identification System (AIS) to extract images that contain a ship in their center. Furthermore, AIS also provides us with information about the ship type and length. The dataset is very unbalanced ($10,430$ Tanker and only $1,071$ Passenger), thus requiring dedicated strategy [5]. Here we simply decided to enlarge our dataset by performing translations and rotations in order to have $20,000$ balanced images. The images employed to train our network are 80×80 images containing ships (not necessarily in the center). The ship footprint ground truth 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). The obtained footprint is not perfect (see Fig. 3b) but is sufficient in order to train the network. Let us note that a CFAR approach could have been employed in order to extract more precisely the ship footprint [9].

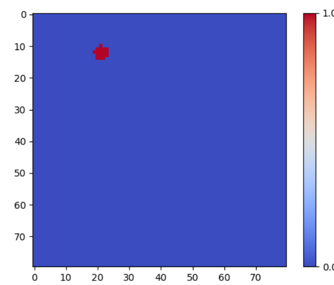
4. RESULTS

We train and test our network on a PC with a single NVIDIA GTX 1080 Ti, an Intel Xeon W-2145 CPU 3.70GHz and 64GB RAM (Keras [2] implementation). For a 80×80 image, our method can run at 55 frames per seconds.

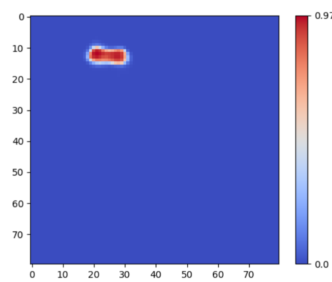
The network is trained using $16,000$ images from the augmented dataset and the remaining $4,000$ images are used for validation. Accurate evaluation of ship detection is difficult, so we conduct a visual inspection to confirm that the detection is well performed by our network (see Fig. 3). Let us note that the detection task has been widely addressed in the literature [10, 4, 6] and is not our main purpose here.



(a) SAR image.



(b) Ship footprint ground truth.



(c) Detection result.

Fig. 3: SAR image (with backscatter intensity), the generated ground truth and result of detection from the network.

To our knowledge, the length estimation is a task that has never been investigated yet using learning-based schemes. Our framework performs well with very promising results. The length is slightly under-estimated: -2.4 m $\pm$ 9.5 m, which is very good regarding the spatial resolution of the Sentinel-1 SAR data. Indeed, having only the ship footprint and the spatial resolution of the image is not sufficient and often leads to an over-estimation of the length. The classification task is of high importance. Table 1 gives the confusion matrix, and several accuracy metrics are also presented in Table 2. The confusion matrix shows some light confusions for passenger ships, decreasing slightly the precision for this class. Some fishing ships are classified as passenger ships impacting the recall for this class. For the tanker and cargo ships, the classification is very accurate. The accuracy metrics confirm these satisfactory results with an overall accuracy and a mean F-score of 95.4%.

5. CONCLUSION

In this paper, a multi-task deep neural network approach was introduced to address joint detection, classification

Label	Tanker	Cargo	Fishing	Passenger	Recall
Tanker	978	7	6	9	97.8
Cargo	8	946	7	39	94.6
Fishing	1	15	934	50	93.4
Passenger	5	13	24	958	95.8
Precision	98.6	96.4	96.2	90.7	

Table 1: Confusion matrix of ship classification.

Label	Tanker	Cargo	Fishing	Passenger	Overall
IoU	96.5	91.4	90.1	87.3	91.3
F-score	98.2	95.5	94.8	93.2	95.4
Accuracy	99.1	97.8	97.4	96.5	95.4
κ	0.98	0.94	0.93	0.91	0.95

Table 2: Accuracy metrics of ship classification.

and length estimation for ships in Sentinel-1 SAR images. We exploit AIS-Sentinel-1 synergies to automatically build groundtruthed training and evaluation datasets. Regarding the considered architecture, a mutual convolutional branch transforms raw inputs into meaningful information. Such information is fed into specific branches for each of the three considered tasks. Ship detection cannot be totally assessed, but a visual inspection still shows our network achieved good performances. As expected, we reach state-of-the-art performance for the detection task but jointly deliver relevant performance for ship classification (above 90% of correct classification) and length estimation (relative bias and standard deviation below 10%). We may point out that the considered residual architecture for length estimation seems to be a critical feature to reach good estimation performance, but should be further investigated in order to confirm its relevance.

Further improvements will be investigated. Using false positive in the dataset would allow to evaluate the relevance of our detection network. We also consider to increase the number of classes and see if our network is robust to more complex scenarios.

REFERENCES

[1] Wentao An, Chunhua Xie, and Xinzhe Yuan. An improved iterative censoring scheme for cfar ship detection with sar imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(8):4585–4595, 2014.

[2] François Chollet et al. Keras: Deep learning library for theano and tensorflow. *URL: https://keras. io/k*, 7(8), 2015.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[4] Miao Kang, Xiangguang Leng, Zhao Lin, and Kefeng Ji. A modified faster r-cnn based on cfar algorithm for sar ship detection. In *Remote Sensing with Intelligent Processing (RSIP), 2017 International Workshop on*, pages 1–4. IEEE, 2017.

[5] Benjamin Kellenberger, Diego Marcos, and Devis Tuia. Detecting mammals in uav images: Best practices to address a substantially imbalanced dataset with deep learning. *Remote Sensing of Environment*, 216:139–153, 2018.

[6] E Khesali, H Enayati, M Modiri, and M Mohseni Aref. 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*, 40(1):395, 2015.

[7] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. In *Advances in Neural Information Processing Systems*, pages 971–980, 2017.

[8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[9] Mingsheng Liao, Changcheng Wang, Yong Wang, and Liming Jiang. Using sar images to detect ships from sea clutter. *IEEE Geoscience and Remote Sensing Letters*, 5 (2):194–198, 2008.

[10] Yang Liu, Miao-hui Zhang, Peng Xu, and Zheng-wei Guo. Sar ship detection using sea-land segmentation-based convolutional neural network. In *Remote Sensing with Intelligent Processing (RSIP), 2017 International Workshop on*, pages 1–4. IEEE, 2017.

[11] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[12] Colin P Schwegmann, Waldo Kleynhans, Brian P Salmon, Lizwe W Mdakane, and Rory GV Meyer. Very deep learning for ship discrimination in synthetic aperture radar imagery. In *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pages 104–107. IEEE, 2016.

[13] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.

[14] Liangpei Zhang, Lefei Zhang, and Bo Du. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):22–40, 2016.