Ship Detection From Raw SAR Echoes Using Convolutional Neural Networks

Kevin de Sousa, Georgios Pilikos, Mario Azcueta, Nicolas Floury

Ship Detection From Raw SAR Echoes Using Convolutional Neural Networks

Kevin De Sousa, Georgios Pilikos, Mario Azcueta, and Nicolas Floury

***Abstract*—Synthetic aperture radar (SAR) is an indispensable tool for marine monitoring. Conventional data processing involves data down-linking and on-ground operations for image focusing, analysis, and ship detection. These steps take significant amount of time, resulting in potentially critical delays. In this work, we propose a ship detection algorithm that operates directly on raw SAR echoes, based on convolutional neural networks. To evaluate our approach, we performed experiments using raw data simulations and real raw SAR data from Sentinel-1 stripmap mode scenes. Preliminary results on this set show the capability of detecting multiple ships from raw data with similar accuracy as using single-look-complex images as input. Simultaneously, running time is reduced significantly, by-passing the image focusing step. This illustrates the great potential of deep learning, moving toward more intelligent SAR systems.**

***Index Terms*—Deep learning, raw data, ship detection, synthetic aperture radar (SAR).**

I. Introduction

In RECENT years, synthetic aperture radar (SAR) has emerged as a powerful remote sensing technology with the ability to operate in all-weather day and night conditions. SAR missions are offering increasingly enhanced coverage capacity and thus, have become a valuable tool for maritime surveillance applications. They offer the capability to detect both small and large ships while not being dependent on the utilization of the automatic identification system.

SAR transmits microwave signals to an area of interest and measures the reflected echoes. Their amplitude and phase are utilized to create high-resolution images of the observed scene.

From these images, analysts can inspect and identify ships or other targets of interest. Nevertheless, with the growing amount of SAR data, automatic approaches are becoming a necessity. Traditionally, automatic ship detection is performed on the ground due to the heavy computational load required. Usually, the raw data are compressed, transmitted to the ground, decompressed, focused to an image and then ship detection algorithms are used. The delay resulting from these data processing steps impedes real-time ship detection.

At the same time, deep learning algorithms have seen a surge in popularity within the SAR community [1] (e.g., classification [2] and on-board data compression [3]). In particular, for ship detection, some early work with deep highway networks [4] and convolutional neural networks [5] showed the potential of deep learning for this application.

The research community has over the years built on these early efforts, moving from more traditional techniques [6], [7], [8], [9], [10] to deep learning based techniques. Several general trends can be observed: two-stage detectors [11], [12], [13], [14], [15], [16], [17], [18], one-stage detectors [19], [20], [21], [22], [23], [24], [25], and anchor free detectors based on different architectures, such as CenterNet [26], [27], FCOS [28], [29]. Contextual information is leveraged to perform detection in complex scenes [30], [31]. As the models are becoming more accurate, the community has started to push for lighter and faster solutions to enhance the detection speed for real-time purposes using techniques such as weight quantization and pruning (for a detailed review [32]).

However, using SAR images to perform on-board ship detection implies that images are already available. Thus, on-board SAR image focusing should be tackled in (near) real-time, which is a nonnegligible technical and computational challenge. To by-pass this computational burden, recent efforts have been directed toward processing intermediate SAR data to detect ships. Using range-compressed data as input to neural network models is a promising avenue [33], [34], [35], [36].

Moreover, the research community is starting to explore the potential of applying deep learning directly to raw SAR echo data. In theory, neural networks can learn to approximate arbitrary functions and should be able to learn a mapping from raw data to a desired output, given sufficient and representative training data and a suitable model. For example, an end-to-end computational pipeline to perform image focusing and landscape classification was proposed [37] or a trainable Omega-K and sparse optimization-based SAR-GMT imaging network with improved image quality [38].

More specifically, a recent study tried to perform ship detection directly on raw SAR echoes [39]. They illustrated that it

Manuscript received 30 November 2023; revised 2 March 2024 and 16 April 2024; accepted 24 April 2024. Date of publication 9 May 2024; date of current version 30 May 2024. *(Corresponding author: Nicolas Floury.)*

Kevin De Sousa was with Wave Interaction and Propagation section Radio Frequency Payloads and Technology Division, Electrical Department, European Space Research and Technology Centre (ESTEC), European Space Agency (ESA), 2200AG Noordwijk, The Netherlands. He is now with the Centre d'Etudes Spatiales de la Biosphère (CESBIO), Université de Toulouse, CNRS/CNES/IRD/INRAE/UT3, Toulouse, France.

Georgios Pilikos and Nicolas Floury are with the Wave Interaction and Propagation section, Radio Frequency Payloads and Technology Division, Electrical Department, European Space Research and Technology Centre (ESTEC), European Space Agency (ESA), NL-2200AG Noordwijk, The Netherlands (e-mail: Nicolas.Floury@esa.int).

Mario Azcueta is with the Copernicus Sentinel-1 Payload section, Copernicus Space Component Space Segment, Earth Observation Projects Department, Directorate of EO Programmes, European Space Research and Technology Centre (ESTEC), European Space Agency (ESA), NL-2200AG Noordwijk, The Netherlands.

Digital Object Identifier 10.1109/JSTARS.2024.3399021

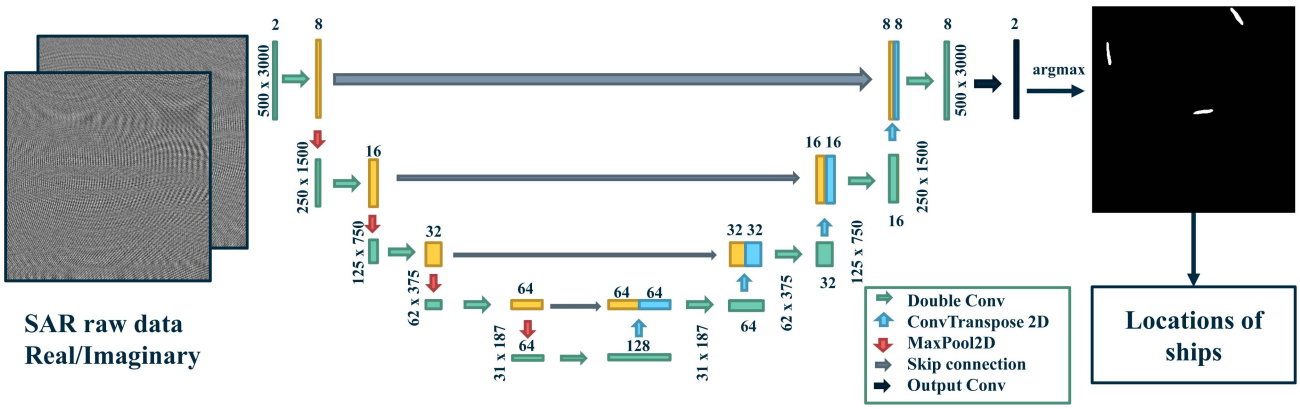


Fig. 1. Visual representation of the proposed algorithm. First, the raw data are processed by a neural network. Then, a binary segmentation map is produced. From this map, the ship structures are extracted to estimate their location.

was possible to perform single ship detection on a simple scene with calm sea background using a bidirectional long short-term memory. Another work demonstrates its feasibility on simulated data and on real data from the European Remote Sensing satellite using residual neural networks (ResNets) [40].

We contribute by proposing a ship detection algorithm that is able to detect multiple ships on real scenes based on convolutional neural networks from raw data. Raw SAR echoes are composed of complex numbers, thus two separate input channels are used, one for the real and another for the imaginary part. These are being processed by a series of convolutional layers to produce an on-board segmentation map (ship or background). In addition, an algorithm is presented, which extracts the ship location based on the centre of mass (precise geolocation could be estimated on-ground later). Using this, we demonstrate that it is possible to detect multiple ships from raw SAR echoes with real data from Sentinel-1 stripmap mode scenes. We also compare the detection accuracy with networks utilizing SLC images as input, discussing the benefits and pitfalls of our approach.

The rest of this article is organized as follows. Section II introduces our proposed ship detection algorithm. Section III describes the experiments performed along with the datasets and the results obtained. Section IV provides a discussion. Finally, Section V concludes this article.

II. Proposed Deep Learning Algorithm

In this section, the proposed ship detection algorithm (see Fig. 1) is explained, along with possible variations. Due to the fact that SAR data are complex in nature, additional considerations are necessary during neural network design. Given that the operations are performed in the raw SAR data domain, we would like to provide to the network, both amplitude and phase information which are paramount to image focusing. However, since extra computation is required to obtain these, it is preferable to operate directly on the real and imaginary parts. The input data are structured such that successive pulses are arranged into a matrix separated into two input channels. The convolutional filters work on blocks of this matrix. We investigated the use of two different convolutional neural network architectures, described next.

A. Residual Neural Network

The first architecture tested is the ResNet [41]. The key idea behind ResNet is the use of residual blocks, which contain skip connections, and sequences of convolutional layers. It applies a series of filters to the input image and produces feature maps. These feature maps capture different aspects of the image, gradually extracting higher-level features as the network increases in depth.

B. U-Net

The U-Net [42] is a specific convolutional neural network that consists of an encoder path, which downsamples the input data, and a decoder path, which upsamples the encoded features to produce a segmentation map. The input data are being processed by a series of layers. Each layer consists of 2-D convolutional filters, followed by normalization and ReLU activation. At each stage of the network, two of these layers are used, denoted by the green arrow in Fig. 1. Max-pooling then follows to downscale the intermediate feature maps, denoted with a red arrow. At the decoder side, the same structure follows, but upscaling is performed using a transposed convolution denoted with a light blue arrow.

C. Loss Function

To optimize the parameters that compose the neural networks, a loss function is required. In this work, we use the cross-entropy loss, which is defined as

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C}\log\left(\frac{\exp(\hat{y}_{n,c})}{\sum_{i=1}^{C}\exp(\hat{y}_{n,i})}\right) y_{n,c} \quad (1)$$

where

$\hat{\mathbf{y}}$ is the output of the model,
C is the number of classes,
$\mathbf{y}$ is the ground truth label, and

N is the total number of pixels in the ground truth.

Cross-entropy could be used for the segmentation of multiple classes (e.g. between ships, sea and land). Special care is needed when there is imbalance of classes. This can be circumvented with the use of weights for classes, inversely proportional to the number of pixels belonging to each class. We focus on ships or sea, to obtain a binary predicted output..

The predicted output of a neural network is given by

$$\hat{\mathbf{y}} = \mathcal{F}(\mathbf{d}) \tag{2}$$

where $\mathbf{d}$ are the raw SAR data, normalized by the maximum amplitude and $\mathcal{F}$ is the neural network mapping from data space to ship detection. To obtain the binary segmentation map, the predicted values for each pixel per channel are compared and the highest value is selected as the class.

D. Estimation of Ship Locations

The binary cross entropy loss function can be used for image segmentation. In our case, we are interested in separating the sea background from the ships, i.e., a binary segmentation. The final aim, however, is to detect the ships' presence and their location, not their shape. To achieve this, the ship structures composed of cluster of points are identified using the following steps.

1) Nonzero elements in the binary map are flagged.
2) Connected nonzero elements are identified. This is achieved by grouping nonzero elements into unique clusters (ships) by iterating through nonzero pixels. Here, we used a search algorithm [43] but any suitable method for the task could be used. Proximity between clusters is not considered.
3) Centre of mass is calculated per ship.
4) List of ship locations are returned.

III. Experiments

In order to investigate our proposed algorithm, two datasets are used. The first one consists of raw SAR data from point target simulations and the second contains real stripmap data.

A. Raw SAR Data Simulation of Point Targets

A synthetic stripmap mode dataset was computed from point targets, distributed randomly in the scene with homogeneous background clutter. Real sea clutter can be modeled using various different models as reviewed in [44]. In this article, a simplistic representation of the sea clutter was modeled as circular complex Gaussian noise [40]. The number and location of the point targets varied randomly. The SAR transmits a chirp, which appears in range. Another chirp appears in azimuth due to the phase history of the collected echoes along the SAR trajectory. These echoes were composed of 1871 azimuth lines and 2500 range samples. Table I lists some key parameters of the simulation. Note that these do not correspond to any actual Sentinel-1 acquisition mode. The sampling frequency, pulse duration, and height were set for ease of experiments. Fig. 2(a) shows the real part of the raw data, Fig. 2(b) shows the corresponding focused image and Fig. 2(c) shows the input of the simulator used to generate the data.

TABLE I
Parameters of the Simulation

Parameters	Values [units]
Centre frequency	5.405 [GHz]
Chirp bandwidth	50 [MHz]
Chirp pulse duration	3 [μs]
Range sampling frequency	150 [MHz]
Pulse Repetition Frequency	1871 [Hz]
Antenna length	12 [m]
Orbit altitude	690 [km]
Mean signal-to-clutter ratio	39.18 [dB]

TABLE II
Synthetic Dataset Description

	Data volumes	Number of targets
Training	700	4053
Validation	200	1228
Test	100	1000
Total	1000	6281

The range-Doppler algorithm was used to focus the raw data [45], which can be decomposed into three steps: range compression, range cell migration correction, and azimuth compression.

In total, 1000 simulation data volumes were generated and divided into three separate datasets: training, validation, and test as described in Table II. Since raw data were simulated from a known input, labeling is not necessary. Instead, the simulator input serves as the ground truth for the locations.

A ResNet was used to detect the presence and location of the point targets from both raw data and SLC images focused from the respective simulated raw data. The model trained using SLC images as input converges faster than the one trained using raw data as input. This could be due to the fact that the functional mapping from SLC images to ships is simpler than the raw data to ships mapping. Instead of training with the same number of epochs, we first trained a model with SLC images as input and empirically chose a cross entropy loss of 2.5×10^{-5} as a threshold. This enables the models trained using raw data as input to reach comparable detection accuracy to those trained using SLC images as input, albeit for longer training time.

All the presented models have been optimized using the ADAM optimizer [46] and trained using the cross entropy loss defined in (1). Table III tabulates the results obtained with different metrics, including the inference time and the focusing time when running on NVIDIA Quadro RTX 8000 GPU. Note that for the inference time, only the time to process the input to the model output is recorded. The loading time of data into the GPU is not taken into account. In addition, estimating the center of mass for each ship structure is not included, as it can be done on the ground with precise geolocation.

With regards to detection accuracy, the F1-score is used as metric. The common practice of using the standard F1-score refers to image segmentation tasks. If the goal is to accurately segment the shapes of the ships, then it would be more suitable to use this. However, this is not what we are interested here, but rather we would like to detect the ships. Thus, we utilize a relaxed F1-score where a correct detection is considered when the centre of mass of a prediction is within the ground truth label. We denote

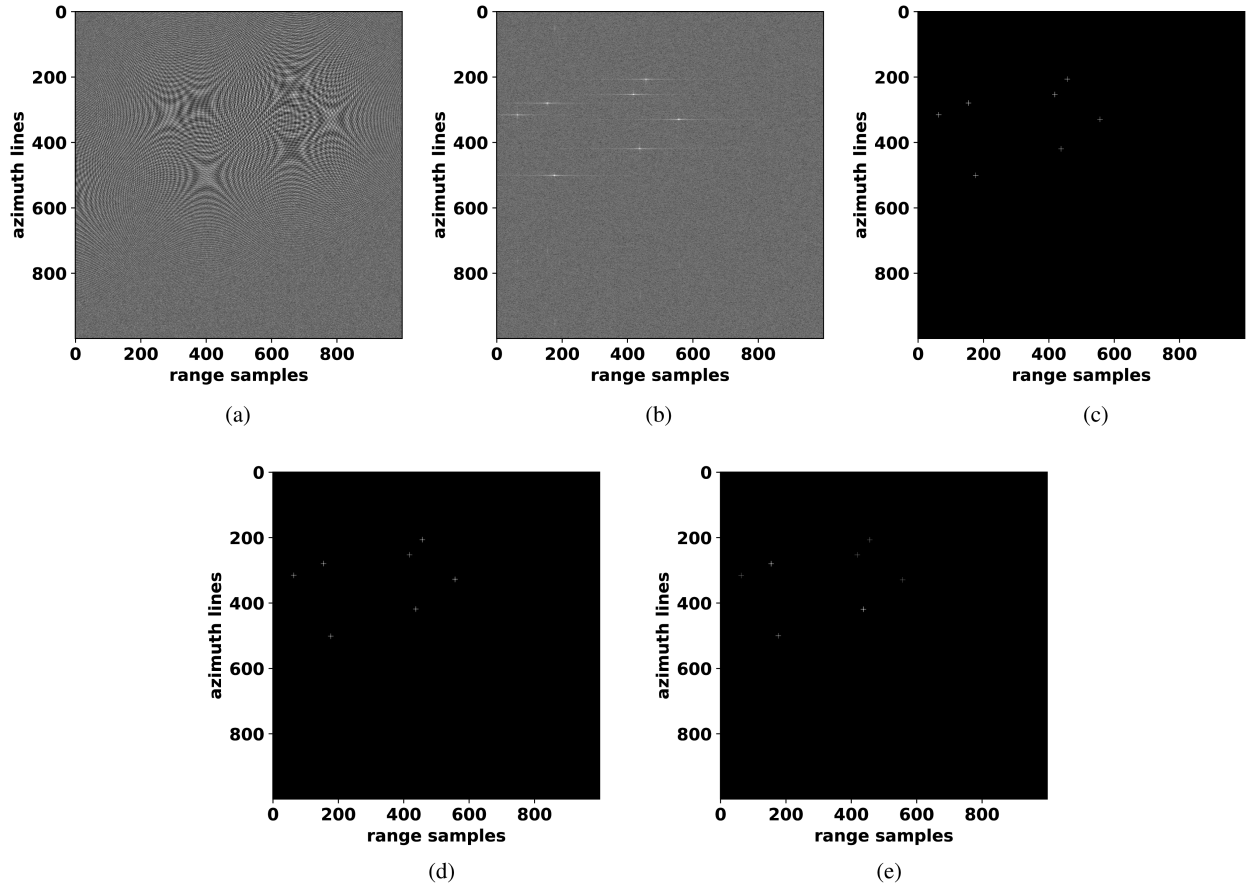


Fig. 2. Portion of data volume from raw SAR data simulator and corresponding results. (a) Real part of raw data. (b) SLC image. (c) Simulation input (ground truth). (d) Prediction using the proposed algorithm from raw data. (e) Prediction using SLC image as input to trained model.

TABLE III
SYNTHETIC MODEL METRICS

Datatype (Channels, Layers)	Inference Time [ms]	Focusing Time [ms]	$F1_{com}$	Training epochs
RAW data (8, 6)	70.908	0	0.885	312
RAW data (32, 6)	159.366	0	0.909	177
RAW data (8, 14)	173.431	0	0.993	168
SLC (8, 6)	59.517	130.755	0.846	6
SLC (32, 6)	151.859	128.053	**0.999**	10

The bold values correspond to the best performance for the $F1_{com}$ metric, which is the metric used to compare our models.

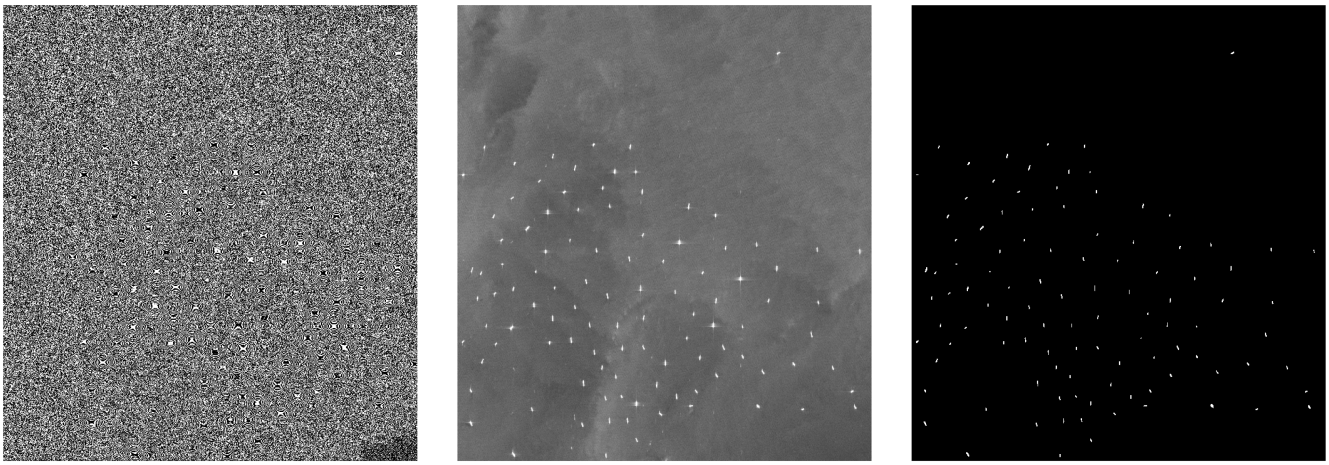
the relaxed F1-score as $F1_{com}$. It can be seen that it is possible to achieve very high accuracy using models with either raw data or SLC images as input. However, the training time of models using raw data as input is much longer, which is performed on-ground. For on-board ship detection, the important metric to consider is the total running time (inference + focusing time), which is considerably reduced when by-passing image focusing. Fig. 2(d) and (e) illustrates the output of the ResNet with 14 layers and eight channels using raw data and SLC images as input, respectively.

B. Real Sentinel-1 Stripmap Mode

Following the testing of our hypothesis to detect point targets from raw SAR data simulations, we move on to real data. To achieve this, we train neural networks from scratch using real Sentinel-1 Stripmap mode (S6) VV polarization data, acquired from the Copernicus Open Access Hub (Copernicus Service Information 2022). Our study focused on the Sao Paulo region, home to one of the busiest cargo container ports in Latin America, situated along the Atlantic Ocean coast. Fig. 3(a) illustrates the real part of the raw SAR data. Fig. 3(b) shows the respective SLC image and Fig. 3(c) is the test binary segmentation label, manually created.

Note that a dim replica of some of the brightest ships in the SLC image in Fig. 3 is visible in the vertical direction. This is a known artefact of SAR images called azimuth ambiguities, and can be mitigated by trading off azimuth resolution with the replica's brightness. Ambiguous ships were specifically not considered during the manual labeling process. A more systematic labeling procedure would be advantageous and subject to future work.

A total of 12 scenes from Sao Paulo were utilized, with each scene cropped to 12 000 $\times$ 19 950 dimensions. To mitigate class imbalance, a section of the sea was excluded. Within each scene,



(a) (b) (c)

Fig. 3. Evaluation scen (contains modified copernicus sentinel data [2022]). (a) Real part of raw SAR data. (b) SLC scene. (c) Binary segmentation label.

TABLE IV
REAL DATASET DESCRIPTION FOR INPUT SIZE 500 × 3000

	Number of images with ships	Total number of images	Number of targets
Training	189	448	713
Validation	52	113	169
Test	21	51	111
Total Samples	262	612	993

The bold values correspond to the best performance for the $F1_{com}$ metric, which is the metric used to compare our models.

8500 azimuth lines and 9000 range samples were employed. The dataset description for an input size of 500 × 3000 can be found in Table IV. The mean and standard deviation signal-to-clutter ratio is 42.81 and 4.95 dB, respectively, between the ships and the sea clutter.

Applying the ResNet directly did not yield satisfactory results, given the higher degree of complexity of real data. Thus, we changed the architecture to a U-net and investigated different configurations, to identify the most suitable with respect to the limitations of computational complexity, training energy consumption and detection accuracy.

1) Architecture Design: There are various hyperparameters that could be changed during architecture design. Nevertheless, here we focus on the three given in the following.

1) *The number of depth layers:* The amount of times the input has been downscaled (e.g., 4 in Fig. 1).
2) *The base channels:* The number of channels after the first convolutional layer (e.g., 8 in Fig. 1). The channels are then increased by multiples of 2 in the encoder and decreased in the decoder, respectively. The number of starting channels will impact the size of the latent space at the bottleneck.
3) *The input size:* Number of range samples and azimuth lines fed into the model.

Table V includes the relaxed F1-score achieved by different configurations and corresponding inference time as defined before, which is illustrated in Fig. 4.

Increasing the amount of channels and thus increasing the capacity of the model does not always translate to improved detection accuracy. This could be due to the fact that bigger models require more training data to be trained sufficiently.

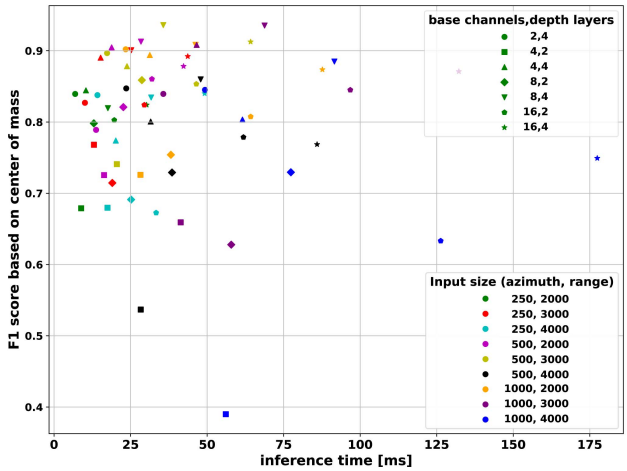


Fig. 4. Visual representation of the different model configurations investigated based on their relaxed F1-score against their inference time.

Increasing the range samples used is not always helpful. Too big and the number of training data are reduced significantly when the number of scenes is fixed. On the other hand, too small and it might not be able to capture all the energy spread by the chirp. Thus, it should be chosen depending on the specific SAR mode characteristics, in particular the chirp length. Furthermore, by increasing the number of layers, number of base channels and the input size, the inference time increases. A tradeoff between detection accuracy and inference time exists and needs thorough consideration.

In our study, we highlight three configurations as follows:

1) (500 × 2000) and (four base channels and four depth layers) achieves $F1_{com} = 0.903$ with inference time = 18.86 ms.
2) (500 × 3000) and (eight base channels and four depth layers) achieves $F1_{com} = 0.937$ and inference time = 35.84 ms.

TABLE V
RELAXED F1SCORE AND INFERENCE TIME

Size of input data		(base channels, number of depth layers)													
		(2, 4)		(4, 2)		(4, 4)		(8, 2)		(8, 4)		(16, 2)		(16, 4)	
azimuth lines	range samples	time [ms]	$F1_{com}$	time [ms]	$F1_{com}$	time [ms]	$F1_{com}$	time [ms]	$F1_{com}$	time [ms]	$F1_{com}$	time [ms]	$F1_{com}$	time [ms]	$F1_{com}$
250	**2000**	6.91	0.836	8.85	0.681	10.41	0.844	13.06	0.801	17.67	0.823	20.09	0.807	30.79	0.826
250	**3000**	10.09	0.826	13.04	0.771	15.25	0.89	19.06	0.717	25.25	0.901	29.53	0.826	43.95	0.895
250	**4000**	14.23	0.844	17.52	0.678	20.27	0.783	25.24	0.691	31.25	0.838	33.84	0.678	48.52	0.844
500	**2000**	13.81	0.793	16.35	0.726	18.86	0.903	22.63	0.825	28.1	0.907	31.20	0.86	42.38	0.875
500	**3000**	17.07	0.897	21.39	0.743	23.95	0.882	28.53	0.859	35.84	0.937	46.07	0.855	63.77	0.914
500	**4000**	24.25	0.859	28.09	0.536	32.01	0.798	39.00	0.73	47.88	0.858	62.14	0.779	86.05	0.768
1000	**2000**	23.91	0.903	28.23	0.727	31.04	0.899	38.07	0.762	46.73	0.913	64.31	0.815	87.46	0.88
1000	**3000**	35.63	0.839	41.30	0.66	46.77	0.911	57.61	0.629	69.20	**0.94**	96.96	0.846	133.41	0.875
1000	**4000**	48.58	0.848	57.29	0.391	63.72	0.813	77.40	0.729	91.79	0.894	126.30	0.645	177.44	0.762

The bold values correspond to the best performance for the $F1_{com}$ metric, which is the metric used to compare our models.

TABLE VI
COMPARISON BETWEEN MODEL TRAINED USING RAW DATA AS INPUT AGAINST SLC IMAGES

	F1	$F1_{com}$	Inference time [ms]	Focusing time [ms]	Power consumption [W]	Training time per epoch [s]	Energy used for training [kWh]	Model Memory [MB]	Model Parameters [#]
RAW	0.840	0.937	35.84	0	111.7	446	5.83 (for total of 421 epochs)	4.846	1270394
SLC	0.912	0.976	34.48	37.84	104.6	431	6.22 (for total of 497 epochs)	4.845	1270194

3) (1000 × 3000) and (eight base channels and four depth layers) achieves $F1_{com} = 0.94$ and inference time = 69.20 ms.

The second one is chosen, for our purposes, as the best compromise between detection accuracy and inference time. This is used as a benchmark to compare against an equivalent model trained using SLC images as input.

2) Comparison Between Raw Data or SLC Images as Input: In order to evaluate the performance of our proposed model that uses raw data as input, we compare it against a model trained using SLC images as input and the same targets. The model with eight base channels and four depth layers is used with input size of 500 × 3000. The same model specifications are used to train a neural network using the amplitude of SLC images in dB, as single channel input. Both models were trained for five hundred epochs and the model with the lowest loss was selected, respectively. Here, we do not impose a threshold on the loss function such that the best possible model can be obtained, taking into account preselected energy requirements (proportional to number of epochs). During the training of neural networks, many performance metrics should be considered. Table VI lists some first efforts to assess tradeoffs between detection accuracy, inference time and energy consumption during training.

Our proposed model, when trained on raw data, demonstrates a relaxed F1-score of 0.937, which is lower than its SLC equivalent, reaching 0.976. With the standard F1-score, the model trained using SLC images as input achieves a score of 0.912, whereas the model that uses raw data as input obtains 0.840. With regards to inference time, both models are comparable. However, our approach does not require any image focusing, approximately halving the total running time. A visual example is shown in Fig. 5 where Fig. 5(d) illustrates the output prediction of the model trained using raw data as input and Fig. 5(e) the output prediction of the model trained using SLC images as input. The raw data used as input is included in Fig. 5(a). The centre of each of the responses in the raw data corresponding to ships is shifted to the right by half of the chirp length and careful consideration is needed when splitting it into portions. In this work, nonoverlapping input data were used, but further considerations are required on the best way to pass the data to the networks.

To get accurate models, it is essential to train for sufficiently long time. However, it is also important to monitor the energy needed. In Table VI, we present the power consumption during training, and the total energy is calculated based on the number of epochs required for the best model. This depends on the specific hardware used for training and can be reduced.

C. On-Board Data Computation

The original U-net model [42] was not designed for on-board, light-weight applications. If it was used in our case, with two inputs, it would have been composed of 31 031 234 parameters. In order to perform on-board SAR data analysis, the proposed neural network model should be as small as possible but simultaneously achieve high ship detection accuracy.

From the previous tradeoff analysis of 63 different model configurations, we chose a model that had 1 270 394 parameters. This is a reduction by 95.91%. The model design was specifically performed in such a way to facilitate on-board data processing. Notably, the model is fully convolutional, with reduced parameters. Fully connected layers are not included to reduce the difficulty of the on-board implementation.

Prior to implementing the model in hardware, network quantization and pruning could be performed to reduce the complexity and memory requirements. After each model alteration, the resulting detection accuracy should be validated. Nevertheless, hardware implementation is left for potential future studies.

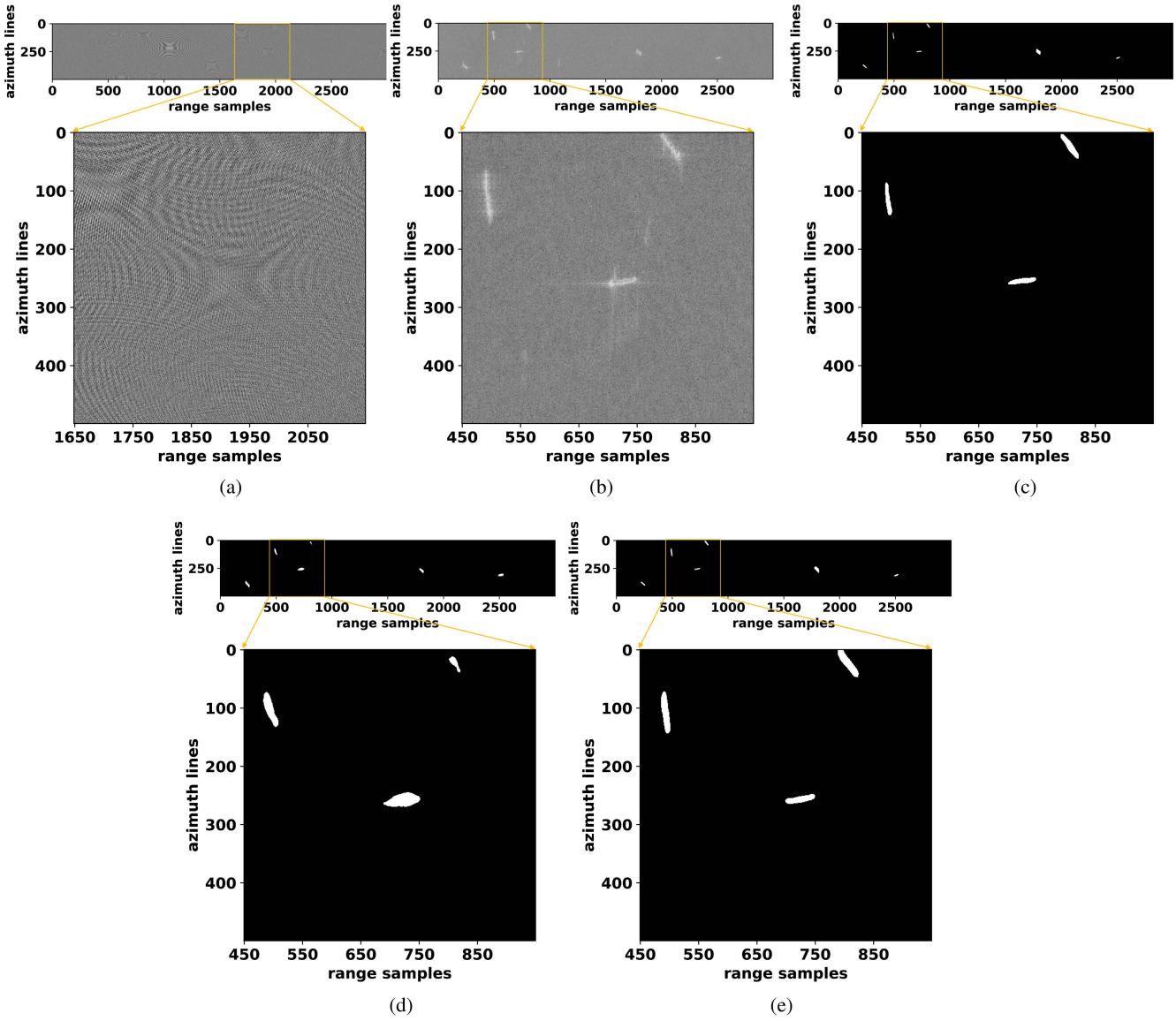


Fig. 5. Visual example using real Sentinel-1 data (contains modified Copernicus Sentinel Data [2022]). A zoom-in portion of the 500 × 3000 input is shown with the corresponding SLC image, ground truth, and predictions. Note the shift to the right in the range-axis of the raw data is equivalent to half of the chirp length. (a) Real part of raw data. (b) SLC image. (c) Label. (d) Model prediction using raw data. (e) Model prediction using SLC images.

IV. Discussion

By performing experiments on both simulated and real data, it was observed that it is possible to detect multiple ships directly from raw SAR data. The functional mapping from raw data to ship detection can be represented using convolutional neural networks, provided that sufficient and representative data are available as well as enough model capacity and structure.

Starting with simulated data, it was possible to detect targets from raw data computed from simulations and achieve similar performance as models using SLC inputs, at the expense of longer training times. Nevertheless, the key parameter is the running time, which is significantly improved by our proposed approach, since no image focusing is required.

Moving to the real Sentinel-1 dataset, the scene is more complex. The number of model parameters had to be increased as well as the neural network architecture changed. A U-net was utilized to leverage the downscaling and upscaling operations, useful for data segmentation tasks.

It was observed that the U-net was also able to detect the ships from the real raw data accurately, being comparable to a U-net using SLC images as input. Nevertheless, if the shapes of the ships produced in the segmentation map are intended to be used explicitly as components for further post-processing (e.g., ship classification), the model using SLC images as input could be more suitable as shown by the difference in F1-score in Table VI. This is because in this current analysis, the model trained using SLC images produces more accurate shapes. Further development is needed to investigate whether bigger, more sophisticated models using raw data as input could improve this performance.

During our analysis, we noted some important factors that can affect the prediction accuracy. The size of the neural network

window (input data size) should be careful chosen. This choice should be done by considering the length of the transmitted chirp, the SAR mode characteristics and the amount of training data available. The depth of the model also matters as it processes more spatial information concurrently.

Overfitting is another factor that needs to be considered. In this work, only training data from the Sao Paulo region was used. This sea background is relatively calm and near a port where ships are likely stationary. This stripmap mode scene was used as proof-of-concept but further work is essential in order to generalize our approach. To achieve this, it is necessary to create a larger dataset with more diverse scenes (open ocean, different levels of background noise, and different number of ships) and various acquisition modes (e.g., extending this work to Interferometric Wide mode). In addition, when manually creating a dataset, labeling is prone to human error. A more systematic approach to training data generation should be considered during further developments.

Running instances of these deep learning models on specific hardware targets suitable for space missions would be very useful. Choosing when and how best to perform the centre of mass estimation and the precise geolocation of ships is also essential, as well as investigating different inputs to the network (e.g., using complex-valued neural networks [47]). Finally, the evaluation of real-time capabilities as well as investigation of any potential changes to detection accuracy (e.g., after weight quantization or pruning) are required.

V. Conclusion

In order to reduce the running time of an on-board ship detection module, it is beneficial to develop algorithms that operate directly on raw SAR echoes. In this work, we developed a novel ship detection algorithm based on convolutional neural networks and demonstrated its performance using simulations and real stripmap mode Sentinel-1 data.

Preliminary results illustrated that comparable detection performance is achieved by our proposed model using raw data as input and models that use SLC images as input. These results serve as proof-of-concept with further investigation required to generalize to more complex scenes and acquisition modes. Depending on the complexity of the observed scene, models could be applied either directly to raw SAR echoes or at various stages of the processing chain (e.g., after range compression). By doing so, more responsive and intelligent SAR systems could be achieved and this work demonstrated the great potential of deep learning toward this development.

Acknowledgment

The author Kevin De Sousa would like to thank the ESTEC Data Systems and Microelectronics division for generously granting him access to their TEC-ED On-Board Processing and Artificial Intelligence Facility. The authors would also like to thank the Sentinel-1 payload team and the Copernicus Open Access Hub for the Sentinel-1 data.

References

[1] X. X. Zhu et al., "Deep learning meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021.

[2] A. G. Mullissa, C. Persello, and A. Stein, "PolSARNet: A deep fully convolutional network for polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5300–5309, Dec. 2019.

[3] G. Pilikos, M. Azcueta, R. Camarero, and N. Floury, "Raw data compression for synthetic aperture radar using deep learning," in *Proc. 8th Int. Workshop On-Board Payload Data Compression*, 2022, pp. 1–8.

[4] C. Schwegmann, W. Kleynhans, B. Salmon, L. Mdakane, and R. Meyer, "Very deep learning for ship discrimination in synthetic aperture radar imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 104–107.

[5] C. Bentes, A. Frost, D. Velotto, and B. Tings, "Ship-iceberg discrimination with convolutional neural networks in high resolution SAR images," in *Proc. EUSAR: 11th Eur. Conf. Synthetic Aperture Radar*, 2016, pp. 1–4.

[6] P. Iervolino, R. Guida, and P. Whittaker, "A model for the backscattering from a canonical ship in SAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 3, pp. 1163–1175, Mar. 2016.

[7] S. Brusch, S. Lehner, T. Fritz, M. Soccorsi, A. Soloviev, and B. Van Schie, "Ship surveillance with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1092–1103, Mar. 2011.

[8] K. Eldhuset, "An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1010–1019, Jul. 1996.

[9] X. Leng, K. Ji, S. Zhou, and X. Xing, "Ship detection based on complex signal kurtosis in single-channel SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6447–6461, Sep. 2019.

[10] Z. Lv, J. Lu, Q. Wang, Z. Guo, and N. Li, "ESP-LRSMD: A two-step detector for ship detection using SLC SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 5233516.

[11] J. Cui, H. Jia, H. Wang, and F. Xu, "A fast threshold neural network for ship detection in large-scene SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6016–6032, Jul. 2022.

[12] M. Kang, X. Leng, Z. Lin, and K. Ji, "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," in *Proc. IEEE Int. Workshop Remote Sens. With Intell. Process.*, 2017, pp. 1–4.

[13] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 860.

[14] Y. Sun, Z. Wang, X. Sun, and K. Fu, "SPAN: Strong scattering point aware network for ship detection and classification in large-scale SAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1188–1204, Jan. 2022.

[15] C. Chen, C. He, C. Hu, H. Pei, and L. Jiao, "A deep neural network based on an attention mechanism for SAR ship detection in multiscale and complex scenarios," *IEEE Access*, vol. 7, pp. 104848–104863, 2019.

[16] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.

[17] Y. Liu, G. Yan, F. Ma, Y. Zhou, and F. Zhang, "SAR ship detection based on explainable evidence learning under intraclass imbalance," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Mar. 2024, Art. no. 5207715.

[18] T. Yue, Y. Zhang, J. Wang, Y. Xu, and P. Liu, "A weak supervision learning paradigm for oriented ship detection in SAR image," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Mar. 2024, Art. no. 5207812.

[19] C. W. Yuanyuan Wang and H. Zhang, "Combining a single shot multibox detector with transfer learning for ship detection using Sentinel-1 SAR images," *Remote Sens. Lett.*, vol. 9, no. 8, pp. 780–788, 2018.

[20] Y.-L. Chang, A. Anagaw, L. Chang, Y. C. Wang, C.-Y. Hsiao, and W.-H. Lee, "Ship detection based on YOLOv2 for SAR imagery," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 786. [Online]. Available: https://www.mdpi.com/2072-4292/11/7/786

[21] X. Xu, X. Zhang, and T. Zhang, "Lite-yolov5: A lightweight deep learning detector for on-board ship detection in large-scene Sentinel-1 sar images," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 1018.

[22] S. Jiang and X. Zhou, "DWSC-YOLO: A lightweight ship detector of SAR images based on deep learning," *J. Mar. Sci. Eng.*, vol. 10, no. 11, 2022, Art. no. 1699.

[23] T. Zhang, X. Zhang, J. Shi, and S. Wei, "Hyperli-Net: A hyper-light deep learning network for high-accurate and high-speed ship detection from synthetic aperture radar imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 167, pp. 123–153, 2020.
[24] Z. Sun, X. Leng, Y. Lei, B. Xiong, K. Ji, and G. Kuang, "BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4209.
[25] X. Ren, Y. Bai, G. Liu, and P. Zhang, "YOLO-Lite: An efficient lightweight network for SAR ship detection," *Remote Sens.*, vol. 15, no. 15, 2023, Art. no. 3771. [Online]. Available: https://www.mdpi.com/2072-4292/15/15/3771
[26] Z. Cui, X. Wang, N. Liu, Z. Cao, and J. Yang, "Ship detection in large-scale SAR images via spatial shuffle-group enhance attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 379–391, Jan. 2021.
[27] H. Guo, X. Yang, N. Wang, and X. Gao, "A CenterNet model for ship detection in SAR images," *Pattern Recognit.*, vol. 112, 2021, Art. no. 107787. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320320305902
[28] J. Fu, X. Sun, Z. Wang, and K. Fu, "An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1331–1344, Feb. 2021.
[29] Z. Sun et al., "An anchor-free detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7799–7816, 2021.
[30] H. Jia, X. Pu, Q. Liu, H. Wang, and F. Xu, "A fast progressive ship detection method for very large full-scene SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Feb. 2024, Art. no. 5206615.
[31] C. Zhang, G. Gao, J. Liu, and D. Duan, "Oriented ship detection based on soft thresholding and context information in SAR images of complex scenes," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Dec. 2024, Art. no. 5200615.
[32] J. Li, J. Chen, P. Cheng, Z. Yu, L. Yu, and C. Chi, "A survey on deep-learning-based real-time SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3218–3247, Mar. 2023.
[33] T. Loran, A. B. C. Da Silva, S. K. Joshi, S. V. Baumgartner, and G. Krieger, "Ship detection based on faster R-CNN using range-compressed airborne radar data," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Dec. 2022, Art. no. 3500205.
[34] X. Leng, J. Wang, K. Ji, and G. Kuang, "Ship detection in range-compressed SAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, 2022, pp. 2135–2138.
[35] S. K. Joshi, S. V. Baumgartner, A. B. Da Silva, and G. Krieger, "Range-doppler based CFAR ship detection with automatic training data selection," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1270.
[36] H. Zeng et al., "An Incept-TextCNN model for ship target detection in SAR range-compressed domain," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, Jan. 2024, Art. no. 3501305.
[37] A. Rittenbach and J. P. Walters, "Demonstration of a fully neural network based synthetic aperture radar processing pipeline for image formation and analysis," *Sensors, Systems, and Next-Generation Satellites XXV*, vol. 11858, 2021, Art. no. 118580K, doi: 10.1117/12.2599955.
[38] H. Zhang, J. Ni, S. Xiong, Y. Luo, and Q. Zhang, "Omega-KA-Net: A SAR ground moving target imaging network based on trainable omega-K algorithm and sparse optimization," *Remote Sens.*, vol. 14, no. 7, 2022, Art. no. 1664.
[39] X. Leng, K. Ji, and G. Kuang, "Ship detection from raw SAR echo data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5207811.
[40] G. Cascelli et al., "Use of a residual neural network to demonstrate feasibility of ship detection based on synthetic aperture radar raw data," *Technologies*, vol. 11, no. 6, 2023, Art. no. 178. [Online]. Available: https://www.mdpi.com/2227-7080/11/6/178
[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[42] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Interv.–MICCAI: 18th Int. Conf.*, 2015, pp. 234–241.
[43] A. Bundy and L. Wallen, "Breadth-first search," *Catalogue of Artif. Intell. Tools*. Berlin, Heidelberg, Germany: Springer, 1984, pp. 13–13.
[44] G. Gao, "Statistical modeling of SAR images: A survey," *Sensors*, vol. 10, no. 1, pp. 775–795, 2010.
[45] J. Bennett, I. Cumming and R. Deane, "Digital processing of synthetic aperture radar data," London, U.K.: Artech House, 2005.
[46] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, *arXiv:1412.6980*.
[47] R. Mohammadi Asiyabi, M. Datcu, A. Anghel, and H. Nies, "Complex-valued end-to-end deep network with coherency preservation for complex-valued SAR data reconstruction and classificationn," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Apr. 2023, Art. no. 5206417.

Kevin De Sousa received the master's degree in electrical engineering from Université catholique de Louvain, Louvain-la-Neuve, Belgium, in 2021. He is currently working toward the Ph.D. degree in remote sensing with the Centre d'Etudes Spatiales de la Biosphère (CESBIO) Laboratory, Université Paul Sabatier 3, Toulouse, France.

Between 2021 and 2023, he worked with the European Space Agency, focusing on ship detection using SAR data. His research focuses on retrieving biophysical variables from remote sensing data (SAR and optical images) using physics-informed neural networks.

Georgios Pilikos received the Ph.D. degree in physics from the University of Cambridge, Cambridge, U.K., in 2019.

After completing the Ph.D. degree, he worked with the Computational Imaging group of Centrum Wiskunde en Informatica, Amsterdam, The Netherlands. He has worked on various research projects related to the development of machine learning models and algorithms for wave-based imaging applications for over ten years. In 2021, he joined the European Space Agency, where he is part of the Wave Interaction and Propagation section. His current research interests include machine learning, on-board data processing, and electromagnetic scattering models applied to microwave remote sensing.

Mario Azcueta received the Ingeniero degree in electronic engineering from the University of Buenos Aires, Buenos Aires, Argentina.

Between 2011 and 2020, he worked for the SAOCOM mission with the National Commission of Space Activities from Argentina (CONAE) focusing on synthetic aperture radar (SAR) image processing, interferometry and calibration. He was responsible for the SAOCOM-1 A external calibration until the completion of its commissioning phase. During 2016–2017, he was a Research Fellow with Politecnico di Milano, Milan, Italy, focusing on bistatic SAR calibration methodologies for the SAOCOM-CS mission. From 2020 to 2022, he led the development of the operational L1 SAR processor for the Noctua project with MetaSensing BV, Noordwijk, The Netherlands. Since 2022, he has been at ESA-ESTEC as SAR performance engineer for the Sentinel-1 mission.

Nicolas Floury received the M.Sc. degree in engineering from Télécom Paris, Paris, France, in 1993, and the Ph.D. degree in applied physics from the University Paris Diderot, Paris, France, in 1999.

Since then, he has been with the European Space Research and Technology Centre, European Space Agency, Noordwijk, The Netherlands, where he is the Head of the Wave Interaction and Propagation Section. His research interests include signal processing and electromagnetic modeling applied to microwave interaction with natural media.