A FULLY CONVOLUTIONAL NEURAL NETWORK FOR LOW-COMPLEXITY SINGLE-STAGE SHIP DETECTION IN SENTINEL-1 SAR IMAGES

Davide Cozzolino, Gerardo Di Martino, Giovanni Poggi, Luisa Verdoliva

DIETI, Università Federico II di Napoli, Naples - ITALY

ABSTRACT

Ship detection is a fundamental task for SAR-based maritime surveillance. Besides providing high reliability, a good detector is required to be computationally light, in order to analyze huge areas in a reasonable time. We propose a fully convolutional neural network for ship detection in SAR images. Thanks to a relatively simple architecture, complexity remains low enough to allow for a single-stage approach, thus avoiding the possible errors of CFAR pre-screening. Experiments on a Sentinel-1 dataset prove the proposed CNN to be much more reliable than CFAR detection.

Index Terms— SAR, ship detection, CNN, machine learning.

1. INTRODUCTION

Ship detection in SAR images is important for a number of maritime surveillance applications, including border security, ship surveillance, and fisheries control. With the wide availability of Sentinel-1 images, acquired in wide-swath TopSAR mode with a revisit frequency of few days, the need for reliable automatic ship detection tools becomes only more pressing.

In low-wind conditions, ship detection from X and C-band SAR images can be considered as a quite simple task, which can be reliably solved by analyzing the image intensity levels. As a matter of fact, CFAR (constant false alarm rate) techniques [1, 2], which perform very basic operations on image amplitudes in sliding-window modality, represent the most popular tools for ship detection in SAR images.

However, some peculiar issues contribute to increase the difficulty of the problem at hand. First, in many applications, a very low false alarm rate (FAR) is required for ship detection systems, less than 10^{-4}. Considering the heavy-tailed distribution of sea reflectance, enforcing a very-low FAR may cause a significant increase of missed detections, even in favorable weather conditions. In addition, the signal backscattered from the sea surface can be strongly heterogeneous due to a number of diverse phenomena, including oceanographic processes, specific meteorological conditions, or presence of films on the sea surface. Finally, a peculiar problem of sea SAR images is the presence of ambiguities [3]. Due to the non-ideality of antennas, aliasing effects are present in the final SAR image, with bright areas (including ships themselves) replicated with deterministic azimuth and range shifts. Although these replicas, or "ghosts", are much weaker than the original objects, they may be wrongly identified as actual targets by conventional CFAR algorithms when occurring on a low-scattering sea background. Therefore, in this case, ambiguities cause performance impairment especially in *low-wind* conditions. Figure 1 shows examples of patches with ships, ambiguities, and sea.

From the above discussion, it is evident that the use of the sole amplitude distribution is not sufficient to meet the strict performance requirements of maritime surveillance applications, especially for telling apart ships from ambiguities. Indeed, ghosts generated by urban areas and big harbors or by the ships themselves represent the main challenge to correct ship detection. Some papers have addressed this specific problem. In [3] an ad hoc pre-processing, based on a modified Wiener-filtering approach, is proposed. However, it requires the availability of single look complex images acquired in stripmap mode, and is thus not suitable for application to wide-swath modes (ScanSAR and TopSAR). In [2] and in [4], instead, ambiguities are recognized based on their displacement from actual targets. This approach, however, requires the knowledge of several sensor parameters, thus being not suitable for implementation by non-expert SAR users. In all cases, these procedures are significantly more complex than basic CFAR detection, making them unsuitable for direct use on the large images involved in real-world applications.

To limit complexity, a two-stage procedure is often considered [2, 4], with a *pre-screening* stage where a simple high-FAR detector is used, followed by a *discrimination* stage, where contextual information is exploited to reduce the FAR at the expense of intense computation. Leveraging on the fact that targets are only sparsely present in the images, the complex discrimination phase is used only on a small subset of the whole image, allowing for an acceptable overall computational burden.

In this paper, we propose to perform ship detection in a single stage through a convolutional neural network (CNN), avoiding the possible errors introduced by pre-screening [5] and achieving both low-complexity and high reliability. Indeed, at the medium/high resolutions adopted by actual SAR

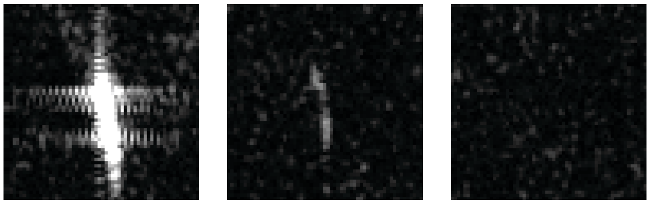

Fig. 1. Examples of patches with, from left to right, ship, ghost (replica of the previous ship), and sea.

sensors, the image patches under analysis exhibit a number of informative textural and structural patterns, such as ship and sea cross-section textures, or ship wakes [1]. These patterns represent precious hints for ship detection, and should be properly taken into account, together with object shape and signal amplitude. Deep neural networks hold the potential to exploit efficiently all these diverse pieces of information, improving ship detection performance and, in particular, ship-ambiguity discrimination. In addition, with suitable architectural choices, a deep neural network can be implemented with limited complexity. To this end, we resort to a fully-convolutional network, with a relatively light structure.

Note that deep learning has been already considered for ship detection. In [6] encouraging results are reported, however, a very deep "highway" architecture is adopted, with fully connected layers and, presumably, a high complexity, motivating a two-stage approach, with the neural network applied only to the final discrimination. In [7], instead, only the specific problem of ship-iceberg discrimination is addressed.

In the rest of the paper, after providing some basic concepts on CNNs (Section II), we describe the proposed architecture (Section III), present some experimental results on Sentinel-1 SAR images (Section IV), and eventually draw conclusions (Section V).

2. CONVOLUTIONAL NEURAL NETWORKS

We recall here a few high-level concepts on CNN and deep learning, instrumental to our needs. For a more comprehensive introduction, the reader is referred to specific sources [8].

A neural network is composed by a large number of densely interconnected, simple, processing units, the artificial neurons. The generic neuron computes its output as a nonlinear function ϕ of the weighted sum of its inputs x_i and of an offset θ

$$y = \phi(\sum_i w_i x_i - \theta) \tag{1}$$

When input patterns match the neuron weights, a large output is observed, hence, the neuron acts as a feature extractor.

Neurons are organized in layers, like in the visual cortex of mammals, with the outputs of layer-n neurons acting as inputs to layer n+1 neurons. The rationale is that a processing architecture with multiple layers of massively interconnected units may be well fit to address complex pattern recognition problems. Indeed, if neurons of the first layer are activated by elementary input patterns, those of the second layer, combining the first layer outputs, are sensitive to patterns of patterns, hence extract more complex input features. By using a sufficient number of layers, more and more abstract features are matched, allowing for the recognition of arbitrarily complex patterns, and the solution of challenging signal analysis problems. Armed with a suitable training algorithm to learn the neuron weights, and with enough labeled training data, a large number of problems can be addressed in a unified manner.

This intuition has been fully corroborated, in the last few years, by a long series of impressive results in all fields of computer vision and image processing, including remote sensing. Besides the fast growth of computing power and the diffusion of large datasets of labeled images, a major enabling factor has been the adoption of efficient network architectures, like the long-known [9] convolutional networks.

With CNNs, the neurons of layer $(n + 1)$ are not connected anymore with all neurons of layer n but only with a local subset of them, implementing the concept of *receptive field*, drawn again from studies in neurophisiology. In addition, all neurons of a layer are identical to one another, except for their receptive fields. With such constraints, the number of free network parameters reduces drastically, allowing fast and effective training. Moreover, the layer output is computed by means of a simple convolution (apart from the nonlinearity).

In a typical CNN architecture besides the convolutional layers mentioned above, other layers may be used to achieve other goals, such as input conditioning (normalization layers), size reduction (pooling layers), and global processing (fully connected layers).

3. PROPOSED ARCHITECTURE

To address the ship detection problem we use the CNN architecture described in Fig.2. The network takes in input a square patch of 64×64 pixels, and outputs the probabilities of target and background classes. Given the resolution of the target images (Sentinel-1) the patch size is large enough to accommodate even the largest ships with at least part of the wake. In any case, the output probability/label is associated only with the central part of the patch. Input patches are preliminarily normalized to zero mean and unit energy, an important step in the presence of high-dynamic data, such as single-look SAR. Then, three convolutional layers follow, which extract more and more complex features. In the figure, we also report the layer main hyperparameters, namely, number of feature maps, size of analysis window, and type of nonlinearity. All filters have relatively small size to limit the number of weights to learn. In all cases a rectified linear unit (ReLU) nonlinearity is used. Each convolutional layer is followed by a max-pooling layer, operating on 2×2 windows, with a step 2 subsampling aimed at reducing the input size, thus allowing subsequent fil-

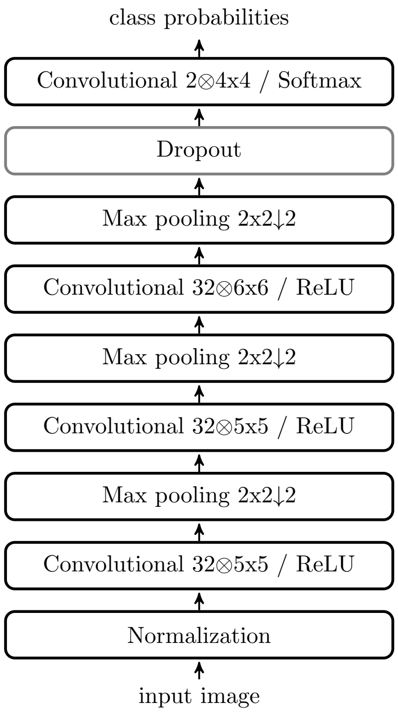

Fig. 2. Proposed CNN architecture.

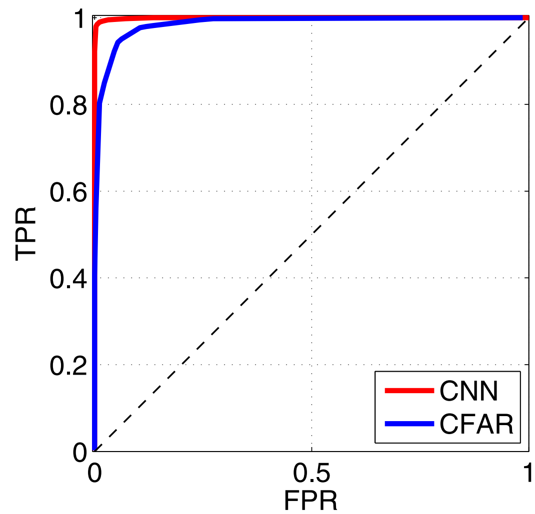

Fig. 3. Receiver operating curves for ship-vs-ghost problem.

ters to see ever larger portions of the input patch. The filters of the last convolutional layer operate on the whole input feature maps, and output two values that are converted in probability through softmax. To avoid overfitting to the training set, a dropout layer is used, with 50% of neurons removed at random at each iteration. The loss function is the cross-entropy between ground truth label and CNN output.

Overall, we are using a relatively simple architecture, to limit both complexity and number of free parameters. Indeed, in most remote-sensing applications, the main problem is the scarcity of labeled training data to learn the network weight, and ship detection makes no exception. We perform training through backpropagation, with the gradient-based optimization method proposed in [10]. Learning rate and weight decay are fixed at 10^{-4} and 0.004, respectively, and the other parameters are set to default values under the CAFFE framework as suggested in [10]. Training operates on minibatches of 192 patches (each one including 16 ships, 16 ghosts, and 160 sea patches) for a total of 32640 iterations.

4. EXPERIMENTAL RESULTS

Our dataset includes 16 Sentinel-1 Interferometric Wide Swath (IW) images in Ground Range Detected (GRD) format with a pixel spacing of 10m$\times$10m in azimuth-ground range. Images have been acquired under a wide variety of sea conditions. To obtain a sufficiently large number of training and testing samples, areas with a rather high density of ships have been selected. The images have been labeled via visual inspection by an expert SAR interpreter, identifying a total of 1066 ships, with a wide variety of shapes and orientations. Regarding ghosts, Sentinel-1 data present a significant amount of both azimuth and range ambiguities [4], thus representing a good test-bed for the proposed approach. On the available dataset, a total of 525 ghosts have been labeled. Two of the images have been used for testing purposes only. The other 14 images have been used for training purposes only. More in detail, 680 ships and 170 ghosts have been selected for training, producing, after very moderate augmentation, 1360 patches of both types. The other 386 ships and 355 ghosts were used for testing. In addition 54400 sea patches are included in the training set, with many more available for testing.

During training, all non-ship patches are labeled as background, therefore the CNN learns to behave in the same way in the presence of sea patches and ambiguities. However, it is interesting to study separately the detection performance in the two cases of ship-vs-sea (hence, no ambiguities in the test set) and ship-vs-ambiguity (no sea patches in the test set). In the first, much simpler, case, the CNN does not provide appreciable gains with respect to a CFAR detector, which is already near-perfect.[1] When considering the ship-vs-ambiguity case, instead, the proposed CNN proves much more reliable than the CFAR detector, ensuring near-perfect discrimination, as shown by the red ROC in Fig.3. The CFAR detector has a much worse ROC (blue in Fig.3): to avoid mis-detections, a FAR in the order of 0.25 must be accepted.

Finally, in Fig.4, we show the color-coded output of both the reference CFAR method and the proposed CNN for a se-

[1] We are working to carry out experiments with a much larger number of patches, to explore the very-low FAR region where significant differences may emerge.

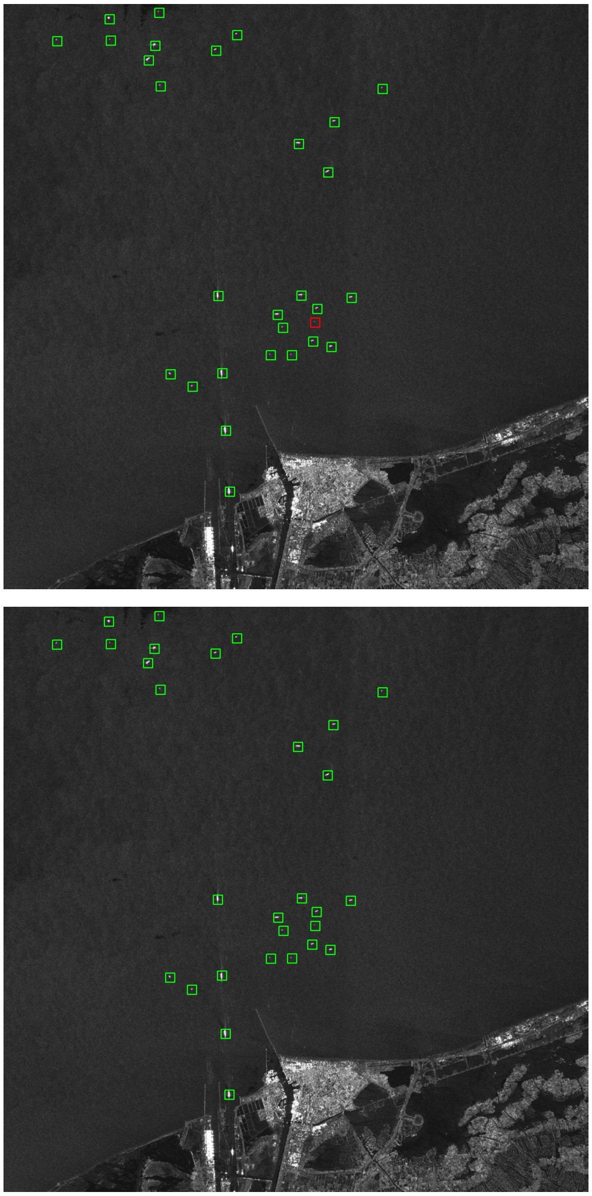

Fig. 4. Ship detection results for CFAR (up) and proposed CNN (bottom).

lected region of the test image of the Suez Canal. The green boxes indicate correctly detected ships, while the red box in the CFAR image indicate a mis-detection. In this specific example, no false alarms are visible, but they are not infrequent in the CFAR case in the presence of ambiguities.

5. CONCLUSIONS

We proposed a fully convolutional neural network for ship detection in SAR images. Although preliminary experiments on Sentinel-1 images are encouraging, there is much room for further improvements. A key issue is to obtain a favourable ratio between training set size and network complexity, so as to ensure correct training. To this end we want to test simpler architectures, and various forms of data augmentation. Extensive testing is then necessary to assess performance against state of the art references.

6. REFERENCES

[1] D.J. Crisp, "The state-of-the-art in ship detection in synthetic aperture radar imagery," Tech. Rep., DTIC Document, 2004.

[2] S. Brusch, S. Lehner, T. Fritz, M. Soccorsi, A. Soloviev, and B. van Schie, "Ship surveillance with TerraSAR-X," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 3, pp. 1092–1103, mar 2011.

[3] G. Di Martino, A. Iodice, D. Riccio, and G. Ruello, "Filtering of azimuth ambiguity in stripmap synthetic aperture radar images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 9, pp. 3967–3978, 2014.

[4] C. Santamaria, M. Stasolla, V. Fernandez Arguedas, P. Argentieri, M. Alvarez, and H. Greidanus, "Sentinel-1 maritime surveillance: Testing and experiences with long-term monitoring," Tech. Rep., 2015.

[5] X. Leng, K. Ji, K. Yang, and H. Zou, "A bilateral CFAR algorithm for ship detection in SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1536–1540, July 2015.

[6] C.P. Schwegmann, W. Kleynhans, B.P. Salmon, L.W. Mdakane, and R.G.V. Meyer, "Very deep learning for ship discrimination in synthetic aperture radar imagery," in *IEEE International Geoscience and Remote Sensing Symposium 2016*, 2016, pp. 104–107.

[7] C. Bentes, A. Frost, D. Velotto, and B. Tings, "Ship-iceberg discrimination with convolutional neural networks in high resolution SAR images," in *European Conference on Synthetic Aperture Radar*, 2016, pp. 491–494.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[9] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mecha- nism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[10] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference for Learning Representations*, 2015.