# Automatic Maritime Object Detection Using Satellite imagery

Konstantina Bereta
*MarineTraffic*
*Athens, Greece*
Email: konstantina.bereta@marinetraffic.com

Dimitris Zissis
*University of the Aegean*
*Ermoupoli, Syros, Greece*
Email: dzissis@aegean.gr

Raffaele Grasso
*NATO-STO-CMRE*
*La Spezia, Italy*
Email: raffaele.grasso@cmre.nato.int

*Abstract*—**In this paper we present an approach for performing object classification and segmentation in satellite images for the Maritime domain. We employ neural network architectures for object classification and segmentation tasks in order to identify different classes of objects in satellite imagery for the maritime domain, such as vessels, land (e.g., port terminals), clouds, etc. We compare the accuracy of different neural network architectures and present the results of our experimental evaluation.**

## 1. Introduction

Maritime Situational Awareness (MSA) is the systematic monitoring of the maritime domain in order to detect maritime activities and events that impact safety and navigation. In order to be able to obtain an up-to-date picture of the maritime domain, various systems have been developed aiming at monitoring the activities of vessels. These systems are distinguished into two broad categories: cooperative and non-cooperative systems.

Cooperative vessel reporting systems require for the collaboration of the vessels' crews to transmit reports on the position of the vessels, along with other information depending on which system is used (e.g., navigational status, speed, etc.). The most common cooperative reporting system is the Automatic Identification System (AIS) that was established as mandatory for all commercial vessels over 299 Gross Tonnage (GT) and for all passenger vessels regardless of their GT that travel internationally. AIS relies on VHF communication. AIS receivers are installed aboard vessels in coastal (and satellite) stations in order to be able to receive AIS messages. Transceivers, i.e., devices that are able to transmit and receive AIS messages are installed board the vessels. Vessels are obliged to send AIS messages frequently. The duration of time intervals between two consecutive AIS transmissions from the same vessel ranges from a few seconds to a few minutes and depends on the navigational status of the vessel (e.g., the speed, change in heading, etc.). Although AIS is an undoubtedly valuable source of information for the maritime domain, it comes with the following drawbacks: (i) there are places with limited or no coverage, (ii) malfunctions in AIS transponders may happen, and (iii) it is common for vessels that engage in illegal activities to turn their transponder off.

In contrast to collaborative vessel reporting systems, non-cooperative systems do not rely on the collaboration of a vessel's crew in order to receive updates regarding its navigational status. Using these systems, vessels can be monitored independently. An example of such a source is a satellite constellation that produces images that enable the detection of vessels. Although the temporal and spatial coverage of satellite data is nowhere near the respective ones for AIS, using data from satellite images fused with AIS data is valuable. For example, if a vessel "goes dark" by switching off its transponder, it is possible that it can be tracked using a high resolution satellite image.

In this paper we present an approach that implements data fusion and AI techniques in order to perform object identification in satellite data. The contributions of our approach to the state-of-the-art are the following: (i) Our approach fully exploits the capabilities of optical images, contrary to related approaches that consider SAR images. More specifically, we use ESA Sentinel 2 images. (ii) Related publications focus in the detection of a certain type of features that can be extracted from a satellite image, especially vessels. However, more than one type of features can be observed such as offshore facilities, clouds, etc. Most related frameworks employ AI techniques such as Convolutional Neural Networks to identify one of these categories (iii) Most related efforts rely either on image processing techniques or on AI and Machine Learning techniques. We develop a hybrid approach that uses both image processing and machine learning techniques. We employ machine learning techniques, such as Convolutional Neural Networks (CNNs) in order to classify the detected objects (e.g., identify the type of vessel, clouds, etc.) and we use image processing to detect objects in the sea and extract attributes (e.g., location, dimensions, etc.).

## 2. Background and Related work

In this section we present some background and related work in the area of classification and semantic segmentation in satellite imagery, and we also describe state-of-the-art CNN architectures that can be employed for these tasks.

## 2.1. Classification in satellite imagery

In related work, [1] presents a comparison study for vessel detection in SAR imagery. [2] introduces an approach for fusing SAR and AIS data for vessel detection. [3] describes an approach that uses Sentinel-1 imagery to estimate the size of vessels. [4] proposes an technique for ship classification in high-resolution SAR imagery using deep learning. [5] presents data fusion techniques for maritime surveillance. [6], [7] and [8] describe more approaches that perform fusion of AIS and SAR data for ship detection. All aforementioned approaches use SAR imagery, while in the work described in this paper we use optical imagery. Optical imagery is also used in [9] recently proposed a CNN network architecture for ship classification using multi-spectral imagery. Our approach deviates from the methods described in these publications in that (i) fully exploit optical imagery as opposed to SAR imagery, (ii) we do not only perform vessel detection but also segmentation, and we also include other classes that can be found in optical imagery such as land (e.g., terminals) and clouds. Another related work is described in [10], in which a state-of-the-art method that heavily relies on CNNs is employed to distinguish sea and land parcels. For land parcels, we additionally employ land masking techniques, so that we mask out the biggest part of an image that that corresponds to land, leaving only coastline parcels and new-built structures that have not been mapped yet (i.e.g, and included in the dataset that we use for land masking).

## 2.2. Convolutional Neural Networks

Some of the state-of-the art CNN architectures for image classification and segmentation are the following:

- Resnet. Resnet [11] is a CNN which is particularly useful for carrying out challenging image classification tasks (e.g., a lot of different objects appearing in scenes, detailed scenes, objects of the same class appearing in different positions and dimensions, colours, etc.). It contains a large number of layers but achieves very good accuracy.
- VGG-16. VGG-16 [12] is one of the most widespread CNN for image classification. VGG-16 contains less layers than Resnet, which makes it more lightweight, faster but less accurate than Resnet.
- UNet. UNet has been reported to achieve very good results for image classification and segmentation and it is known for its widespread application in the medical domain [13] due to its proved accuracy in the classification and segmentation of medical imagery where details are important.
- Segnet. Segnet [14] follows an Encoder-Decorder architecture similarly to U-net, but Segnet does not include skip connections as U-net does. Segnet does not include any skip connections and one of the main characteristics of its architecture is that the maxpooling indices of the encoder layers are used by the corresponding decoder layers in whiche the upsampling operation is performed.
- Fully Convolutional Network (FCN) [15]. FCN is a typical CNN architecture used for semantic segmentation. It is built by transforming a typical CNN (e.g., VGG-16) into fully convolutional and using transposed convolutions for upsampling.

## 3. Approach

In this section we describe our approach. Section 3.1 describes the training workflow and Section 3.2 predication describes the classification/segmentation workflow.

### 3.1. Training

The training workflow appears in Figure 1. It comprises the following steps:

- **Image acquisition**. The Satellite images are automatically downloaded through the Python API[1] of the Copernicus Open Access Hub Hub[2].
- **Image transformation**. The satellite image is transformed from its source Coordinate Reference System to the WGS84. The reason for this is to be able to perform the land masking operation that is described below. To retrieve the source CRS of the image, the metadata file of the image is used. In the context of this work, we use Sentinel-2 images. These are optical images that come in three different resolutions: 10m, 20m and 60m resolution. We use the 10m resolution images.
- **Land Masking**. Land masking is an operation that masks out the land part of an image. To be able to do this, we use a global coastline shapefile that contains the geometries of the coastlines of all countries in the world. We combine this file with the satellite image using a Python GDAL library and the result is an image where the pixels that correspond to land appear with black colour. However, not all land parcels are masked. This might be either because of the resolution of the coastline file, or by the fact that some new land parcels such as terminals and other port facilities are built. For this reason, we also train our models to include a "land" class so that we are able to identify the new structures.
- **Tiling**. Since satellite images are large files, tiling them into smaller tiles, as for example in $256 \times 256$ tiles facilitates image processing operations.
- **Annotation**. Image annotation is twofold. First, we use AIS data as ground truth in order to spot the vessels that appear on the image. Then, we manually annotate the rest of the vessels that do not have registered AIS signals, if any, and the rest of the

---

1. https://sentinelsat.readthedocs.io/en/stable/api.html
2. https://scihub.copernicus.eu/dhus

classes appearing in the scene (e.g., land parcels, clouds) using the tool labelme[3]. For the segmentation task, all vessel instances need to be manually labeled (even those identified through AIS).

- **Training**. Finally, we use the annotated dataset to train our Convolutional Neural Network. The workflow is orthogonal with respect to the CNN architecture used. We experimented with different CNN architectures for the classification and segmentation tasks, as shown in Section 4. The result of this task is our trained model, which we will use for the prediction phase.
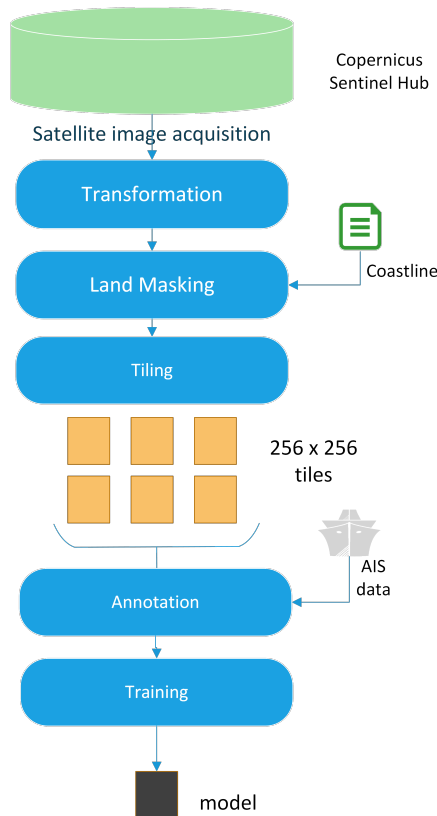


Figure 1. Training workflow

## 3.2. Classification and Segmentation

The workflow for the classification and segmentation tasks is illustrated in Figure 2. It consists of the following steps:

- **Image acquisition**. The workflow is triggered by a request to monitor a vessel. We retrieve its position(s) within the time frame of interest (e.g., its latest positions) from AIS data. This can be for example a vessel that appears out of coverage in the MarineTraffic website[4]. The goal is to locate

3. http://labelme.csail.mit.edu/Release3.0/
4. https://www.marinetraffic.com

| Base model | Segmentation model | Accuracy |
|---|---|---|
| VGG16 | Unet | 0.7647 |
| Vanilla CNN | Unet | 0.929211 |
| Vanilla mini CNN | Unet | 0.99 |
| Resnet50 | UNet | 0.9017 |
| Vanilla CNN | Segnet | 0.87 |

TABLE 1. ACCURACY OF DIFFERENT CNN MODELS

the vessel of interest in satellite imagery, so that we know its position while it was out of AIS coverage. Therefore, we make a request to the Sentinel API for all images acquired in the area of interest within the given time frame and download the results.

- **Transformation, land masking, and tiling**. These steps are the same as the ones of the training phase. The downloaded images are transformed, land masked and tiled.

- **Classification and Segmentation**. The land-masked tiles that result from the previous step are then fed as input to the trained CNN model in order to perform classification and segmentation. The result is a set of image tiles annotated with the instances of the classes of interest (i.e., vessel, cloud, land). The identified featured are then extracted (**feature extraction**). **Metrics calculation**. Finally, we calculate the metrics of the instances that have been classified as vessels. We perform pixel level calculations to identify the dimensions of the vessel, its position (georeferenced in WGS84), the direction and the navigational status (i.e., underway using high/medium speed, stopped/moving with very low speed, etc.), considering the wakes. Given these characteristics in place we are then able to match the vessel of interest with the identified vessels appearing on the Satellite image.

## 4. Experimental Evaluation

For the experimental evaluation we used a dataset consisting of annotated Sentinel-2 image tiles. We tested different combinations of CNN architectures as base and segmentation layers using 5 epoch training and we evaluated their accuracy for the segmentation task for vessel instances. For the implementation and evaluation of the models we used the keras-segmentation Python library[5].

The results are shown below in Table 1.

The results shown in 1 show that, in our case, a simple CNN as base layer combined with UNet for the segmentation layer outperforms all other configurations. However, Segnet combined with a vanilla CNN works also very good, achieving 0.87 accuracy. One of the reasons for that is that Sentinel-2 images have relatively low resolution, and fairly few features per tile with fairly simple shapes, especially taking into account the fact that we include a land masking step that simplifies significantly the number and area of land

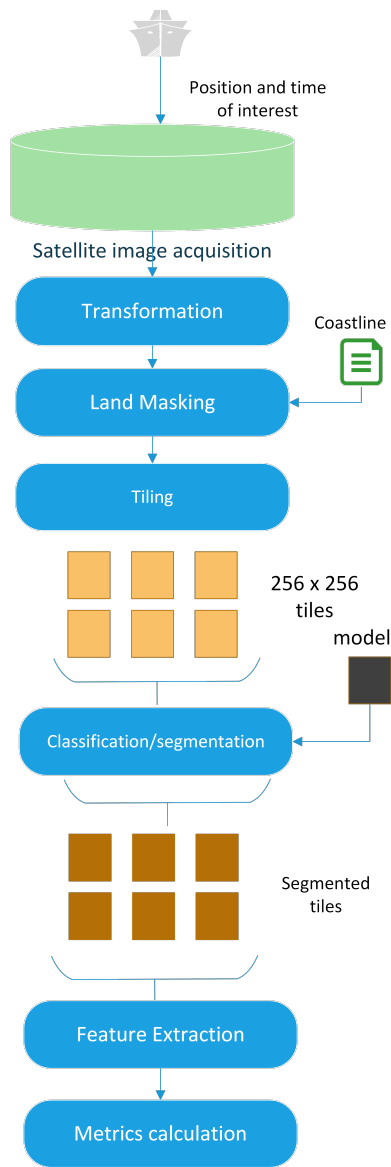5. https://github.com/divamgupta/image-segmentation-keras

Figure 2. Prediction workflow

parcels that may be included in a tile. Also, taking into account the height of the above which Sentinel satellites operate, the same vessels have minimal changes in shape in images taken from different angles, so more sophisticated CNNs can be an overkill, at lease as base layers.

## 5. Conclusions

In conclusion, this paper presents a data-driven approach that identifies objects in the maritime domain such as vessels, land parcels (e.g., new terminals) and other objects that might be seen in satellite imagery and can be considered as noise, such as clouds. Our approach employs state-of-the-art deep learning techniques to perform object classification and segmentation against a dataset that consists of satellite images fused with AIS data and is able to identify vessels extracting also their characteristics, such as position, direction, heading, and vessel type. In the future, we plan to continue our benchmarking work and release a benchmark for testing different CNN algorithms for classification and segmentation against satellite images.

## References

[1] P. Iervolino, R. Guida, P. Lumsdon, J. Janoth, M. Clift, A. Minchella, and P. Bianco, "Ship detection in sar imagery: A comparison study," 07 2017.

[2] L. Achiri, R. Guida, and P. Iervolino, "Sar and ais fusion for maritime surveillance," 09 2018.

[3] M. Stasolla and H. Greidanus, "The exploitation of sentinel-1 images for vessel size estimation," *Remote Sensing Letters*, vol. 7, no. 12, pp. 1219–1228, 2016.

[4] Y. Wang, C. Wang, and H. Zhang, "Ship classification in high-resolution sar images using deep learning of small datasets," *Sensors*, vol. 18, p. 2929, 09 2018.

[5] F. Mazzarella, A. Alessandrini, H. Greidanus, M. Alvarez, P. Argentieri, D. Nappo, and L. Ziemba, "Data fusion for wide-area maritime surveillance," 06 2013.

[6] F. Mazzarella, M. Vespe, and C. Santamaria, "Sar ship detection and self-reporting data fusion based on traffic knowledge," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1685–1689, Aug 2015.

[7] M. Guerriero, P. Willett, S. Coraluppi, and C. Carthel, "Radar/ais data fusion and sar tasking for maritime surveillance," in *2008 11th International Conference on Information Fusion*, June 2008, pp. 1–5.

[8] F. M. Vieira, F. Vincent, J. Tourneret, D. Bonacci, M. Spigai, M. Ansart, and J. Richard, "Ship detection using SAR and AIS raw data for maritime surveillance," in *24th European Signal Processing Conference, EUSIPCO 2016, Budapest, Hungary, August 29 - September 2, 2016*, 2016, pp. 2081–2085.

[9] R. Grasso, "Ship classification from multi-spectral satellite imaging by convolutional neural networks," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[10] Z. Chu, T. Tian, R. Feng, and L. Wang, "Sea-land segmentation with res-unet and fully connected CRF," in *2019 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2019, Yokohama, Japan, July 28 - August 2, 2019*. IEEE, 2019, pp. 3840–3843.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016, pp. 770–778.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, 09 2014.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, 2015, pp. 234–241.

[14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015.* IEEE Computer Society, 2015, pp. 3431–3440.