Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)
Special Track on AI for Good (Demos)

Argo: Towards Small Vessel Detection for Humanitarian Purposes

Elisabeth Moser[1,4*], **Selina Meyer**[2,4], **Maximilian Schmidhuber**[2,4], **Daniel Ketterer**[3,4] and **Matthias Eberhardt**[1]

[1]University of Applied Sciences, Regensburg, Germany
[2]Chair for Information Science, University Regensburg, Germany
[3]Esslingen University of Applied Sciences, Germany
[4]Space-Eye e.V., Regensburg, Germany
{elisabeth.wittmann, selina.meyer, maximilian.schmidhuber, daniel.ketterer}@space-eye.org, matthias.eberhardt@st.oth-regensburg.de

Abstract

Refugees trying to get to Europe via the Mediterranean often face human rights violations. The present situation is not in line with the UN's SDG's 10 and 16. We present *Argo*: a semi-automatically created vessel classification dataset focused on small boats, with the aim to enable NGOs and the public to detect refugee boats in satellite imagery. We achieve a classification recall of 91% on small ships. With a tool developed on top of the results presented here, NGOs could collect information and hold institutions participating in illegal activities accountable.

1 Introduction

For years, refugees have been risking their lives on sea routes along the Mediterranean, usually travelling in unsuitable inflatable rubber boats to reach shelter and seek asylum in Europe. Along their route, they often face illegal pushbacks preventing them from getting to safety and leading to the deaths of thousands of people [Tondo, 2021]. Little information is available to the public about those events. As a result, institutions and actors involved are hardly held accountable [Parliamentary Assembly, 2019]. This situation is in stark contrast with human rights principles and inhibits the United Nations Sustainable Development Goals (SDG's) 10 (reducing inequality) and 16 (peace, justice and strong institutions). It corrupts our peace and justice system and strips refugees of their right to seek asylum, thus reinforcing already existing inequalities between members of developed and developing nations [FRA, 2020]. Enabling NGOs and the public to identify refugee boats in distress on satellite data would be beneficial to the SDG's, as it would provide a way for institutions participating in illegal pushback operations to be held accountable for their actions.

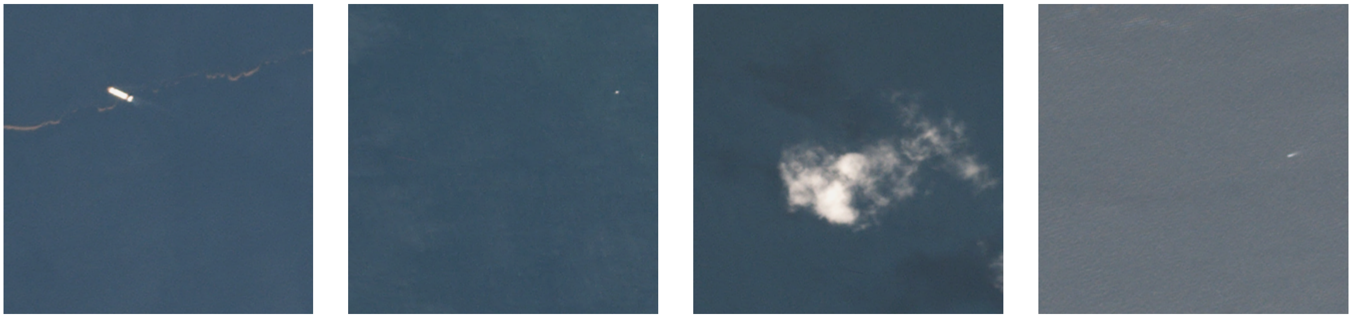
Currently, available satellite data around small vessel detection and classification is scarce, as most free to use data sources do not offer high enough resolution [Liu *et al.*, 2017]. Existing high-resolution datasets are usually created through labour-intensive manual annotation and do not focus on small vessels [Yang *et al.*, 2018; Lin *et al.*, 2017; Zhang *et al.*, 2019; Airbus, 2019]. While large vessels such as oil tanks and battleships are easy to spot [Liu *et al.*, 2018], small vessels remain hard to classify [Kanjir *et al.*, 2018]. One reason for this is that in order for vessels to be detected, image resolution needs to be at least one third of the vessel length [Bannister and Neyland, 2015]. Thus, for vessels smaller than 20 m, which are most common for transporting refugees, a spatial resolution of less than 7x7 m is required. Publicly available satellite images, such as Sentinel-2 data, yield no less than 10 m spatial resolution [European Space Agency, 2015], rendering the detection and classification of small vessels virtually impossible.

In this work, we present *Argo*, a new semi-automatically created high-resolution dataset with ground truth, including imagery of small vessels. We describe a scalable dataset creation process, using Automatic Identification System (AIS) data, which all passenger ships, as well as large cargo ships are required to broadcast at regular intervals [International Maritime Organization, 2019]. We also show that the resolution of our dataset is high enough to classify vessels smaller than 20 m in a real, uncontrolled setting. In future work, we plan to expand this dataset with imagery from the Mediterranean using different data sources in order to support NGOs and humanitarian efforts in the reliable detection of small vessels. We also plan to identify different signals of distress in order to be able to detect pushbacks and refugee boats in dangerous situations.

2 Related Work

Traditionally, Synthetic-Aperture Radar (SAR) data was most commonly used to detect ships and boats at sea. Recently, optical satellite imagery has become more common as a data source [Kanjir *et al.*, 2018]. Approaches also vary between vessel detection and vessel classification. Detection returns accurate coordinates of detected ships on large scale images but requires labour-intensive human annotation for dataset creation. In contrast, classification only returns information on whether or not a ship is present in a given picture, which greatly simplifies the training and dataset creation process [Ulusoy and Bishop, 2006]. Both approaches have been used in literature. For instance, [Kanjir, 2019] used optical

*Contact Author



(a) True Positive (b) False Positive (c) True Negative (d) False Negative

Figure 1: Examples of the network's predictions [Planet Team, 2017].

Sentinel-2 data and a decision tree-based approach to detect vessels at sea. They achieved mixed results on boats longer than 20 m but could not detect smaller vessels due to low spatial resolution.

[Topputo *et al.*, 2016] combined optical and SAR data and used AIS data to distinguish between known and unknown - possibly illegal - vessels and validate classifications. Although they achieved good results, they faced the problem that their system learned to ignore boats without AIS signal - which includes mostly smaller boats that are legally not required to carry AIS. [Milios *et al.*, 2019] used the same approach, combining AIS-data and Sentinel-1 imagery to create a labelled vessel dataset. They reached high scores in classification but faced the same problem as [Topputo *et al.*, 2016]. Thus combining AIS data with available satellite imagery is suitable to counteract data scarcity for vessel detection datasets in general but is not suitable for small boat detection without additional adaptation.

In an experiment explicitly focused on small vessel detection, [Lanz *et al.*, 2021] collected data by placing a rubber boat commonly used by refugees on a lake and collecting SAR data on it. For detection, they used intensity-based, sublook-based and polarimetric-based approaches. Their initial results are promising but have not yet been confirmed to yield comparable results on imagery at sea. [Zhang *et al.*, 2019] detected ships on optical data with a spatial resolution of 1 m. Their models were effective on both offshore and inland rivers and are capable of detecting both smaller and bigger vessels, proving that data with high spatial resolution is required to detect smaller boats effectively. Unfortunately, their dataset is not publicly available.

3 Dataset Creation

In order to collect ground truth imagery of small boats at sea, we created *Argo*, a labelled classification dataset with over a hundred examples of boats smaller than 20 m, using an intersection of AIS data with satellite imagery as presented by [Topputo *et al.*, 2016] and [Milios *et al.*, 2019]. Since identifying specific positions to create a detection dataset would entail much more manual labour, we focused on classification data, making the dataset creation process highly scalable. We used PlanetScope-4-Band imagery [Planet Team, 2017] taken in the coastal region between Miami and Palm Beach (US) in 2017. This region was selected due to the availability of high temporal resolution historical AIS data from marinecadastre.gov [BOEM and NOAA, 2022] and the high number of smaller boats. To create positive classification examples including vessels, the metadata of available satellite imagery was intersected with the AIS data, and small images of 1300x2400 pixels around each detected ship were downloaded. For negative examples, the AIS database was used to verify that no vessels are present in the perimeter. This produced a balanced dataset of 1750 annotated PlanetScope 4-Band images at 3.5 m/pixel resolution.

During the first experiments, it became apparent that the created dataset contained some systematic errors:

- Clouds covering the scene (labelled as ship)
- Additional vessels without AIS signal (labelled as no ship)
- No visible vessel in the imagery (labelled as ship)

Three human reviewers checked the *Argo* dataset to mitigate those errors. Each reviewer checked each image at least three times. Their annotations were aggregated and data marked as mislabelled at least 80% of the time was relabelled. There were 129 ship images, which were relabelled to no ship, and 50 samples vice versa. This corresponds to a total of less than 10% of the dataset. Thus, the intersection of AIS data with satellite imagery already provides a relatively good estimate but human supervision is still advised to eliminate the error sources mentioned above.

The *Argo* dataset is made available with PlanetScope imagery as reflectance corrected BGRN NumPy binary files and the AIS metadata of the ships, as well as the human corrected labels are provided as a .csv file [1].

4 Experiments

Fine-tuning of pretrained image classification networks is a common approach to tackle image classification tasks where only small datasets are available. We chose a pretrained ResNet34 [He *et al.*, 2016] from the PyTorch model zoo, pretrained on the ImageNet Challenge data [Deng *et al.*, 2009] as the basis for the presented experiments and used Fastai

[1] https://doi.org/10.5281/zenodo.6058710

[Howard and Gugger, 2020] with a PyTorch backend. We employed handcrafted colour adaptation of the RGB channels to increase contrast and create a colour distribution more similar to the ImageNet data. First, a crop of 400x400 pixels in the middle of the image was taken with a random deviation of up to 20 pixels. In addition, we applied the following commonly used preprocessing steps: flipping, rotation, zoom, perspective warping and adaptions to contrast and lighting. Finally, we compared three different modes of fine-tuning:

- ResNet34 tuning only the final layer on RGB data,
- ResNet34 tuning all layers on RGB data,
- ResNet34 adapted for RGBN data, tuning all layers.

Whilst the tuning of all or just the final layer on RGB data was achieved by removing the NIR channel from the 4-Band PlanetScope data, training on the full RGBN data required some engineering on the networks architecture. The red channel filter in the first convolutional layer is duplicated for the NIR-channel to adapt the ResNet34 for 4-channel input data. We chose the red filter, as it most closely resembles the NIR data. Therefore, the first convolutional layer was replaced with a 4-channel convolutional layer, with weights initialised to the same values as the original filter's red, green, blue and red channel filters.

5 Results

Each of the experiments was repeated five times to ensure the reliability of the presented metrics. We performed five-fold cross-validation with stratified folds, which were kept the same for all experiments and runs. We present the achieved mean accuracy, F1-score and ROC-AUC score in table 1. Fine-tuning all layers to RGB data performed best, reaching a mean accuracy of 91.9% and a mean F1-score of 91.3%. The RGBN network and the RGB network where only the last layer was trained performed slightly worse with F1 scores of around 88%. The high ROC-AUC scores hint at decision boundaries that can be well adapted to provide few false positives while having almost no false negatives. In figure 1, we display representative examples of wrong and correct classifications[2]. We find that recall was stable regardless of vessel size, with a recall value of over 90% for vessels smaller than 20 m and at most 4% differences in recall between groups of different vessel lengths (see figure 2).

Network	Acc	F1	ROC-AUC
RGB last layer only	0.892	0.880	0.950
RGB all layers	**0.919**	**0.913**	**0.970**
RGBN all layers	0.888	0.876	0.943

Table 1: Mean performance of the three tested networks after five runs of five-fold cross-validation. Bold scores are significant at $p < 0.05$.

[2]More examples of correctly and incorrectly classified data and more detailed results are available at https://wandb.ai/elizamanelli/ship/reports/Ship-Dataset--VmlldzoxNTM5MDAw

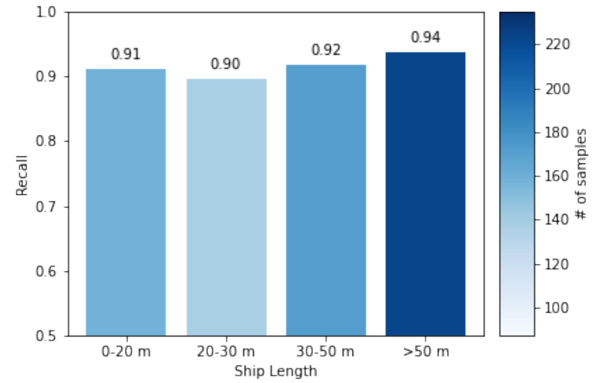


Figure 2: Recall of the trained network by ship length. Samples with no corresponding ship length in AIS data were dropped. For better readability the y-axis is cropped at a Recall of 0.5.

6 Discussion

The overall performance of the networks is very promising and shows that the *Argo* dataset is well suited for the task. The higher performance of the RGB networks over the RGBN network suggests that the current integration process of the NIR channel into the network is still sub-optimal. Since we are mainly interested in spotting small vessels, we focused on the comparability of results across different vessel lengths in evaluation. The performance does not drop for smaller vessel lengths. Comparing these results with the performance of existing datasets might yield further insights in the future.

Opting for classification instead of detection might seem questionable due to the large area considered for the final detection task. However, the creation process of a classification dataset can be automated a lot better. Therefore, it is easier to create additional classification datasets using this method for different satellite imagery sources. Thus, this approach allows the adaptation of pretrained classification networks for multi-channel satellite data. In addition, the intersection of AIS data and satellite imagery can be used to create custom datasets based on prerequisites like a maximum vessel length or a low speed over ground. In a final system, a classifier trained on such data could be used to confirm or reject predictions by a detection network with high recall.

7 Conclusion

We presented an approach to creating a vessel classification dataset without intensive manual annotation and published the *Argo* dataset of PlanetScope-4-Band annotated imagery [Planet Team, 2017] with this paper, proving the feasibility of this approach. We showed the applicability of the dataset to the problem of small vessel classification by fine-tuning pretrained networks. Good overall performance was reached with an F1-score of 91.3% and a ROC-AUC score of 97%. We emphasise that the performance remained high for small vessels with less than 20 m in length. With this contribution, we seek to empower NGOs to shed light on an area where the UN's SDGs are not sufficiently implemented and enable further research on small vessel detection.

References

[Airbus, 2019] Airbus. Airbus ship detection challenge. https://www.kaggle.com/c/airbus-ship-detection/overview, 2019. Accessed February 9th 2022.

[Bannister and Neyland, 2015] Nigel P. Bannister and David L. Neyland. Maritime domain awareness with commercially accessible electro-optical sensors in space. *International Journal of Remote Sensing*, 36(1):211–243, 2015.

[BOEM and NOAA, 2022] BOEM and NOAA. Bureau of ocean energy management (BOEM) and national oceanic and atmospheric administration (NOAA): MarineCadastre.gov - AIS. https://marinecadastre.gov/ais/, February 2022. Accessed February 10th 2022.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[European Space Agency, 2015] European Space Agency. Sentinel-2 user handbook. https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook, July 2015. Accessed February 9th 2022.

[FRA, 2020] FRA. 1.9. push-backs and pull-backs at sea. In *Handbook on European law relating to asylum, borders and immigration*, pages 46–48. European Union Agency for Fundamental Rights and Council of Europe (FRA), 2020.

[He *et al.*, 2016] 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.

[Howard and Gugger, 2020] Jeremy Howard and Sylvain Gugger. Fastai: A layered api for deep learning. *Information*, 11(2):108, Feb 2020.

[International Maritime Organization, 2019] International Maritime Organization. AIS transponders. https://www.imo.org/en/OurWork/Safety/Pages/AIS.aspx, 2019. Accessed February 12th 2022.

[Kanjir *et al.*, 2018] Urška Kanjir, Harm Greidanus, and Krištof Oštir. Vessel detection and classification from spaceborne optical images: A literature survey. *Remote Sensing of Environment*, 207:1–26, 2018.

[Kanjir, 2019] Urška Kanjir. Detecting migrant vessels in the mediterranean sea: Using sentinel-2 images to aid humanitarian actions. *Acta Astronautica*, 155:45–50, 2019.

[Lanz *et al.*, 2021] Peter Lanz, Armando Marino, Thomas Brinkhoff, Frank Köster, and Matthias Möller. The inflate-SAR campaign: Testing SAR vessel detection systems for refugee rubber inflatables. *Remote Sensing*, 13(8):1487, 2021.

[Lin *et al.*, 2017] Haoning Lin, Zhenwei Shi, and Zhengxia Zou. Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 14(10):1665–1669, 2017.

[Liu *et al.*, 2017] Zikun Liu, Liu Yuan, Lubin Weng, and Yiping Yang. A high resolution optical satellite image dataset for ship recognition and some new baselines. In *International conference on pattern recognition applications and methods*, volume 2, pages 324–331. SciTePress, 2017.

[Liu *et al.*, 2018] Wenchao Liu, Long Ma, and He Chen. Arbitrary-oriented ship detection framework in optical remote-sensing images. *IEEE geoscience and remote sensing letters*, 15(6):937–941, 2018.

[Milios *et al.*, 2019] Aristides Milios, Konstantina Bereta, Konstantinos Chatzikokolakis, Dimitris Zissis, and Stan Matwin. Automatic fusion of satellite imagery and ais data for vessel detection. In *2019 22th International Conference on Information Fusion (FUSION)*, pages 1–5. IEEE, 2019.

[Parliamentary Assembly, 2019] Parliamentary Assembly. PACE - resolution 2299 (2019) - pushback policies and practice in council of europe member states. https://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=28074, June 2019. Accessed February 10th 2022.

[Planet Team, 2017] Planet Team. Planet application program interface: In space for life on earth. https://api.planet.com, 2017.

[Tondo, 2021] Lorenzo Tondo. Revealed: 2,000 refugee deaths linked to illegal eu pushbacks. https://www.theguardian.com/global-development/2021/may/05/revealed-2000-refugee-deaths-linked-to-eu-pushbacks, May 2021. Accessed February 11th 2022.

[Topputo *et al.*, 2016] Francesco Topputo, Mauro Massari, Riccardo Lombardi, Marco Gianinetto, Andrea Marchesi, Martina Aiello, Stefano Tebaldini, and Francesco Banda. Space shepherd-using space systems to save human lives. *Research Publications at Politecnico di Milano*, 2016.

[Ulusoy and Bishop, 2006] Ilkay Ulusoy and Christopher M Bishop. Comparison of generative and discriminative techniques for object detection and classification. In *Toward Category-Level Object Recognition*, pages 173–195. Springer, 2006.

[Yang *et al.*, 2018] Xue Yang, Hao Sun, Kun Fu, Jirui Yang, Xian Sun, Menglong Yan, and Zhi Guo. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sensing*, 10(1), 2018.

[Zhang *et al.*, 2019] Shaoming Zhang, Ruize Wu, Kunyuan Xu, Jianmei Wang, and Weiwei Sun. R-CNN-based ship detection from high resolution remote sensing imagery. *Remote Sensing*, 11(6):631, 2019.