DEEP LEARNING FOR VESSEL DETECTION AND IDENTIFICATION FROM SPACEBORNE OPTICAL IMAGERY

Giona Matasci*, Jonathan Plante, Kevin Kasa, Payam Mousavi, Andrew Stewart,
Andrew Macdonald, Anne Webster, Jennifer Busler

MDA, 13800 Commerce Parkway, Richmond, BC, Canada V6V 2J3 – giona.matasci@mda.space

KEY WORDS: Ship detection, Ship tracking, Re-identification, CNN, RetinaNet, Twin networks, Very high resolution imagery, Optical remote sensing.

ABSTRACT:

We present a deep learning-based vessel detection and (re-)identification approach from spaceborne optical images. We introduce these two components as part of a maritime surveillance from space pipeline and present experimental results on challenging real-world maritime datasets derived from WorldView imagery. First, we developed a vessel detection model based on RetinaNet achieving a performance of 0.795 F1-score on a challenging multi-scale dataset. We then collected a large-scale dataset for vessel identification by applying the detection model on 200+ optical images, detecting the vessels therein and assigning them an identity via an Automatic Identification System association framework. A vessel re-identification model based on Twin neural networks has then been trained on this dataset featuring 2500+ unique vessels with multiple repeated occurrences across different acquisitions. The model allows to naturally establish similarities between vessel images. It returns a relevant ranking of candidate vessels from a database when provided an input image for a specific vessel the user might be interested in, with top-1 and top-10 accuracies of 38.7% and 76.5%, respectively. This study demonstrates the potential offered by the latest advances in deep learning and computer vision when applied to optical remote sensing imagery in a maritime context, opening new opportunities for automated vessel monitoring and tracking capabilities from space.

1. INTRODUCTION

Understanding maritime activities at sea and on water bodies in general is crucial for many private and public entities (governmental maritime authorities, shipping companies, naval forces, etc.). Activities such as fishing, cargo transportation, passenger and recreational traffic, need to be monitored and regulated in order to prevent or deter Illegal, Unreported and Unregulated (IUU) fishing or human trafficking. Over the last several decades, global ship traffic has dramatically increased (Tournadre, 2014) stressing the need for advanced and effective monitoring tools.

Space-borne remote sensing has become a widely adopted approach for maritime surveillance (Kanjir et al., 2018). Satellite imagery offers panoptic views of large swaths of ocean that are difficult to cover via navigation/patrolling. For instance, it is a crucial tool to support search-and-rescue operations. Traditionally, the main sensing modality used for maritime surveillance has been Synthetic Aperture Radar (SAR) as it does not rely on daylight and is able to penetrate clouds (Stasolla et al., 2016). In recent years, the sharp increase in the number of satellite platforms with optical sensors, led by the deployment of constellations such as WorldView (WV) and PlanetScope, has generated widespread interest in passive sensing capabilities as well. Such sensors operating in the visible and near-infrared regions of the electromagnetic spectrum offer an increased spatial resolution (up to 30 cm) which greatly improves the monitoring capabilities by increasing both the accuracy and the richness of the retrieved information.

The main components of a system for maritime surveillance based on spaceborne imagery can be summarized as follows (Kanjir et al., 2018):

- Vessel detection: locating all the vessels present in the image, i.e., the main task feeding all others.
- Vessel classification: determining the class of the detected vessels (e.g., fishing, tanker, cargo).
- Vessel characterization: deriving additional vessel attributes such as vessel dimensions (length and width), heading, etc.
- Vessel identification: determining the vessel identity, e.g. its Maritime Mobile Service Identity (MMSI) number.
- Vessel tracking: correlating and linking subsequent vessel contacts in order to establish a vessel track with its positions over time.

Focusing on vessel identification, the task can either be thought of in absolute terms with the retrieval of the MMSI number of the vessel of interest or in relative terms by re-identifying a given vessel among a list of candidates.

Such capabilities open the opportunity to uniquely (re-)identify the same vessel across multiple images, supporting the task of vessel tracking (Tunaley, 2004). This could happen at a short temporal scale, where multiple images are acquired within a few hours over the same area or along a maritime corridor where ships have to be re-identified and matched across acquisitions. Tip-and-cue scenarios could be implemented where optical acquisitions are tasked based on preliminary detections from larger-swath SAR images successively collected along a predicted vessel track. A similar approach could be applied at a larger temporal and spatial window, for instance in the case of the search and tracking of a suspicious vessel that might have visited multiple ports or entered a monitored zone (vessel on a "watch list"), days or even months apart. Vessel identification could also be applied when an operator is interested in inspecting a few vessels appearing in a given image, to check for their identity/characteristics from an existing vessel database.

* Corresponding author

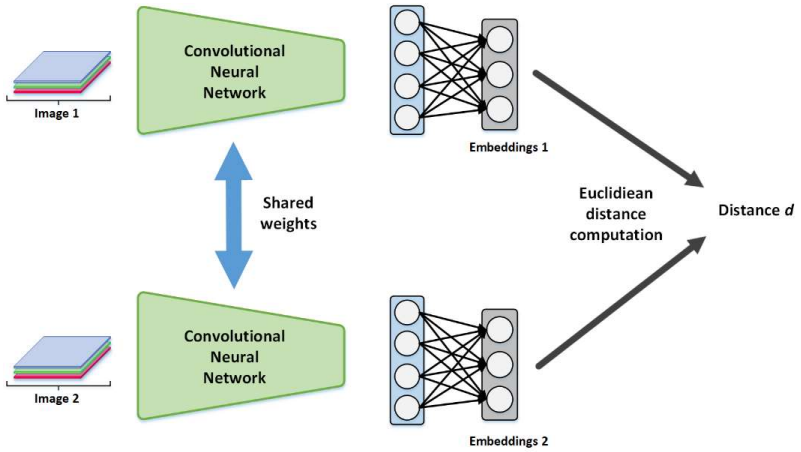


Figure 1. Twin network architecture: CNN backbone and global average pooling operation (green) followed by a series of fully-connected layers (blue) that output the embeddings feature vectors (grey), based on which a distance *d* between input images can be computed.

With the advent of deep learning, Convolutional Neural Networks (CNN) have powered major breakthroughs in automated image interpretation, with direct applications to maritime domain monitoring. The most recent successful vessel detection approaches (Yang et al., 2018; Zhang et al., 2019; Zhu et al., 2020) rely on the latest object detection methods developed in the field of computer vision such as Faster R-CNN (Ren et al., 2015) or RetinaNet (Lin et al., 2017). In vessel identification, previous efforts involving satellite images were based on traditional techniques (e.g., spectral signature analysis) and on coarse spatial resolution platforms such as Sentinel-2 (Heiselberg, 2016). More modern approaches based on deep learning are so far concerned with vessel (re-)identification from natural images, such as those acquired by in-situ port cameras, devices onboard vessels, etc. (Gundogdu et al., 2017; Qiao et al., 2020).

Twin neural networks (Hoffer and Ailon, 2015; Koch, 2015), originally referred to as "Siamese", have been a key step forward in the field of contrastive/metric learning (Chen et al., 2020). They were key in the recent progress made to solve low-shot/one-shot image recognition problems (image retrieval). Such framework allows to establish similarities between images by generating concise image descriptions (an embedding space) that can be compared with simple operations (e.g., dot products, distance computations). This progress in the field of computer vision is behind the advances in re-identification techniques, applied to persons (Ye et al., 2021; Zhong et al., 2019) or vehicles (Chu et al., 2019; Zheng et al., 2019), for instance. Such approaches bear many similar traits to the maritime surveillance problem considered here and inspired our analysis. More recently, the fast-developing sub-field of self-supervised learning also leveraged these principles to enable training of twin networks with artificial positive samples obtained via data augmentation, thus without explicitly providing labels (Caron et al., 2020; Ji et al., 2019). Recent state-of-the-art results in image recognition have been obtained with this family of techniques (Chen et al., 2020).

The goal of the present work is to explore the vessel detection and re-identification tasks that are part of the general maritime surveillance from space pipeline outlined above, specifically when using Very High Resolution (VHR) spaceborne optical images (spatial resolution between 0.3 and 0.5 meters). Our main contribution is a vessel re-identification method based on Twin neural networks trained on a real-world maritime vessel dataset. The pipeline ingests satellite imagery and automatically detects vessels, which can be then further analyzed by other downstream tools, vessel identification being one of them. A central part of this study was the large scale data collection, for which we leveraged a state-of-the-art deep learning-based object detection method, RetinaNet, to locate vessels in a large set of optical remote sensing images. Indeed, accurate detections are key to collect a high-quality training dataset for vessel identification (minimizing false positives) but also at inference time in an operational setting (making sure all relevant vessels in a scene are detected). This detection step was followed by the assignment of an identity to the detected vessels by leveraging an Automatic Identification System (AIS) association framework. Based on this ground truth, a Twin network has been trained to return a meaningful ranked list of the most similar vessel images in a database when provided an input image for a specific vessel.

2. METHODS

2.1 RetinaNet for vessel detection

In our pipeline, for the main upstream task of vessel detection, we developed a model based on the RetinaNet architecture (Lin et al., 2017). The architecture features three main components. First, a backbone CNN acts as main feature extractor (e.g. ResNet). Next, a Feature Pyramid Network (FPN) takes several layers of the pyramidal hierarchy of the CNN and builds lateral connections with upsampled layers to obtain high resolution layers. Finally, a classification and a regression head are attached to each one of these FPN layers to produce the final bounding box and object class predictions. The often large foreground-background class imbalance is mitigated by the use of the focal loss function. The multiple prediction grids allow the detector to handle the multi-scale nature of challenging detection problems such as those encountered in maritime environments (objects of different sizes).

This single-stage object detector was adapted to handle the specificities of remote sensing datasets. Factors such as the size and shape of the objects at hand (e.g., elongated vessels, very small crafts) or the relatively small size of the labeled datasets (compared to traditional ones available in computer vision) had to be considered and addressed. These considerations resulted in changes in network architecture, anchor box configuration and image augmentation strategies.

2.2 Twin networks for vessel identification

To determine if two images are representing the same vessel or not we implemented a Twin neural network. The network encodes the two images into a low-dimensional embedding space. Based on these embedding vectors, a distance or a similarity metric can be computed to determine how close the two images are to each other. Figure 1 shows an overview of the network's architecture. The network is composed of two encoder branches with a CNN backbone (e.g., ResNet), followed by a global average pooling layer (to be independent of the input image size) and by a series of fully-connected layers that output a feature vector of dimension n for each image, the embeddings. Two images are input in parallel to the network and go through the same operations, as the weights of the encoders are shared. The Euclidian distance d between these embedding vectors is then calculated and used to determine if the two images in the pair are the same individual (positive sample) or are representing two different objects (negative sample). Positive image pairs should have small distances while negative pairs will have large distances.

When training the network, we present it with three images: an anchor image, another image of the same vessel, and an image of a different vessel. This results in two image pairs: positive (anchor and same vessel image) and negative (anchor and different vessel image). We then minimize the contrastive loss:

$$(1-Y)\frac{1}{2}d^2 + (Y)\frac{1}{2}\{max(0, m-d)\}^2 , \qquad (1)$$

where Y represents the label of the image pair and m is a margin value. A positive image pair has a label of 0 and a negative image pair a label of 1. This results in minimizing the distance within positive pairs and maximizing the distance within the negative pairs. The margin has the effect of reducing the importance of very dissimilar pairs (distance score beyond the margin) during the training procedure. The choice of using the anchor-positive-negative setup was inspired by another popular loss function: the triplet loss, which has a similar formulation (Weinberger and Saul, 2009). However, preliminary experiments resulted in superior results when using the contrastive loss. We also note how in practice, as the weights of the two branches of the network are shared, we use one single encoder (the twin network formulation is general purpose and would accommodate the use of different image sources, one per branch).

3. EXPERIMENTS

3.1 Vessel detection

3.1.1 Data: To train our vessel detection model, we used a total of 9 VHR images acquired by the WorldView-2 and WorldView-3 platforms between 2015 and 2019 over 7 Areas of Interest (AOI) defined in ports or high maritime traffic areas around the World (e.g., Singapore, Strait of Hormuz). The processing level included ortho-rectification, pan-sharpening and atmospheric compensation. The images were sliced to retain only the RGB bands (bands 5-3-2).

The images were manually labelled and the location of 9004 vessels was recorded. To build the final annotated vessel detection dataset, we added 3766 vessels from the xView dataset (Lam et al., 2018), for a total of 12,770 ships. The dataset is challenging as it includes vessels of significantly different sizes, from large container ships hundreds of meters long to small motorboats spanning just a few meters.

We chipped the large WV images with a chip size of 1024 x 1024 pixels to obtain the smaller images fed to the network. The chips were split to have 70% of the vessels for training, 10% for validation and 20% for test set. Additionally, with the goal of increasing the robustness of our model (reducing the number of false alarms), empty chips were added to each set with a proportion of 20% of the total set size.

3.1.2 Model setup and training: In our preliminary experiments, best performances were observed by using a ResNet-50 backbone and finer grids of the FPN as prediction layers (small spacing between grid cells). The associated anchor configurations featured more extreme aspect ratios and reduced scales, to better capture the vessels' shape. To virtually increase the size of the training set, we adopted an augmentation strategy involving random rotations and flips of the training chips and corresponding bounding boxes. During training we monitored the performance of the validation set via the Average Precision (AP) metric to select the best model. We relied on a Keras implementation of RetinaNet (Gaiser and de Vries, 2019) and adapted it to meet our needs. To further demonstrate the validity of our approach, we also provide a comparison with another state-of-the-art object detection model, Faster R-CNN. The implementation of this two-stage detection network also used a ResNet-50 as backbone and featured a FPN (Wu and others, 2016). It adopted the same anchor box configurations and image augmentation strategies as RetinaNet.

3.1.3 Results: Table 1 shows a summary of the model performance for the two compared models. RetinaNet generalized well on the independent test set, with an AP of 0.799 and an F1-score (at threshold t = 0.5) of 0.795. Faster R-CNN showed promising yet inferior performances (AP = 0.628, F1-score = 0.727). Figure 2 depicts the precision-recall curve on the test set for the RetinaNet model. We note how the default score threshold t of 0.5 yields the best possible F1-score.

	Precision	Recall	F1-score	AP
RetinaNet	0.865	0.725	0.795	0.799
Faster R-CNN	0.776	0.684	0.727	0.628

Table 1. Assessment metrics for the vessel detection models on the test set of the WV dataset.

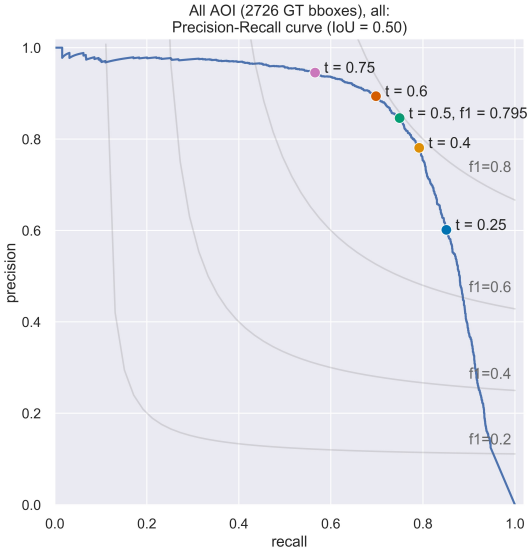
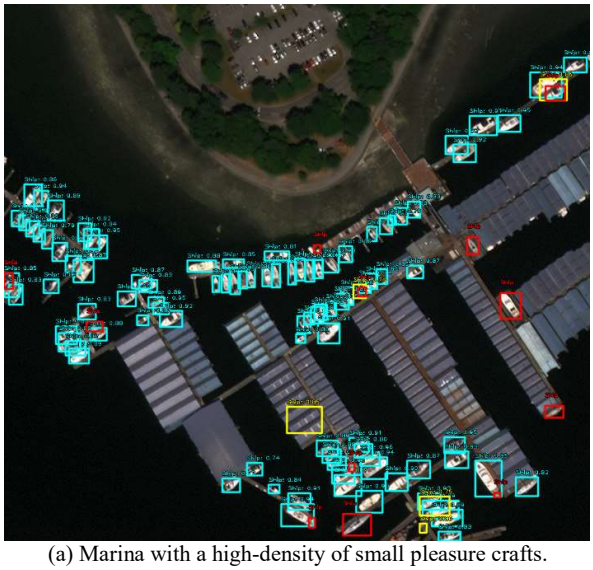
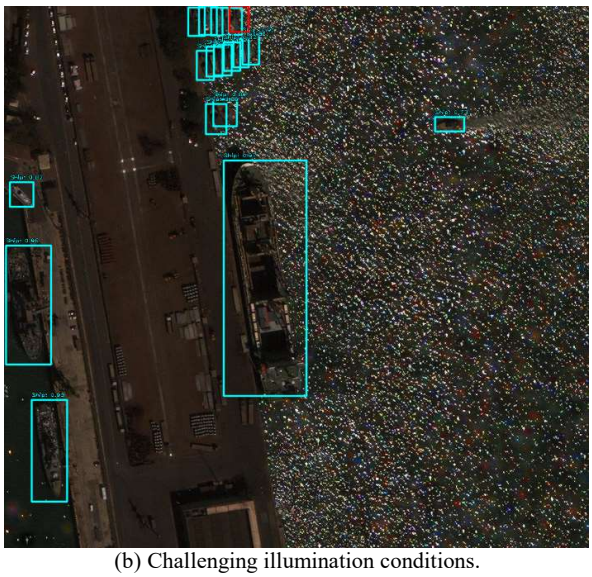


Figure 2. RetinaNet precision-recall curve on the test set of the WV dataset. The colored circles depict possible thresholds t on the prediction score. Isolines for F1-score are drawn in gray.



(a) Marina with a high-density of small pleasure crafts.



(b) Challenging illumination conditions.

Figure 3. RetinaNet detection results in various environments (light blue = TP, yellow = FP, red = FN). Image chips are 1024 x 1024 pixels in size, corresponding to a ~ 350 m side.

Figure 3(a) shows a typical result in the detection of small vessels in tight spaces, whereas Figure 3(b) shows the ability of the model in handling adverse illumination conditions (sun glare).

3.2 Vessel re-identification

3.2.1 Data: To build a suitable large scale dataset for the vessel identification experiments, we used a semi-automated labeling procedure leveraging the RetinaNet vessel detector presented in Sections 2.1 and 3.1 and an AIS association tool. As a first step, since we needed a dataset based on real-world examples of vessels appearing in two or more images, an image selection process was undertaken to identify satellite images where a large number of repeated vessels could be present. This ensured a suitable dataset with a large number of vessels occurring in multiple images (different ports, sea states, lighting conditions, etc.).

Focusing on 29 AOIs over busy ports in Northern/Western Europe, the Mediterranean Sea, the Black Sea, and East Asia, a total of 207 WV-2 and -3 images acquired in the year 2018 were obtained. Subsequently, using the RetinaNet model we developed, a large-scale inference on these images resulted in the detection of more than 224,000 ships. Figure 4 shows an example of the capabilities of our vessel detection approach in the Istanbul AOI (Turkey). Finally, using an in-house AIS association framework, information regarding the detected vessels (and the manually annotated ones, see Section 3.1.1) was gathered from an AIS database offered by ORBCOMM[1]. The automated procedure links the detected vessels to candidate tracks built from series of AIS contacts broadcast by the same ship (same MMSI number). The logic leverages information about vessel length, heading and position.

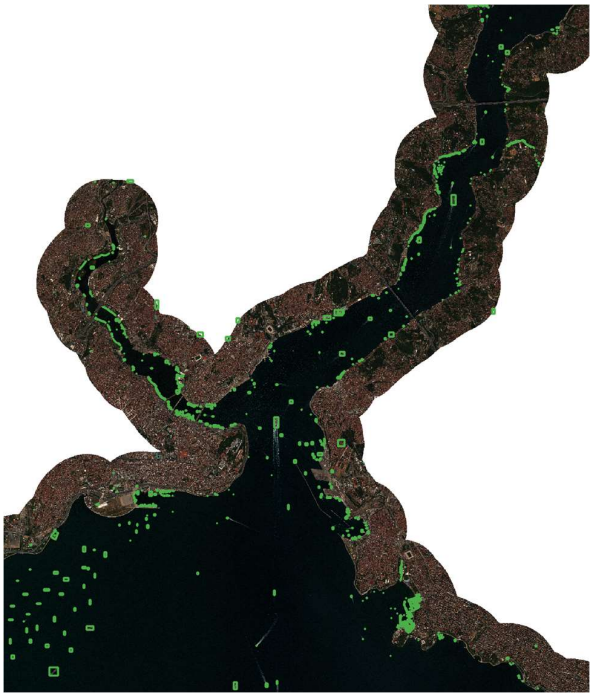


Figure 4. Panoptic view of the vessel detection result in the Istanbul AOI. Bounding boxes for the detected vessels are displayed in green.

[1] https://www.orbcomm.com/en/industries/maritime/satellite-ais

We obtained a total of 2575 unique vessels appearing twice or more across different WV images (3.1 occurrences on average). Although the detection model can locate vessels as small as motorboats (few meters in length), the size of the vessels ultimately included in this re-identification dataset was controlled by the type of vessels generally transmitting AIS signals, which due to International Maritime Organization regulations governing vessel requirements for AIS transponders is statistically more likely to be mid to large vessels (>15m). These repeated vessels were split into a training (70%), validation (10%) and test (20%) sets. We also added a set of 9169 ship occurrences from ships appearing only once across the entire set of images. The goal was to be able to build a more diverse set of negative pairs (different ships) at training time for the Twin network.

3.2.2 Model setup and training: The cropped-out vessel images were resized to a fixed size of 500 × 500 pixels (by resampling) before being fed to the Twin network. To ensure the model learned from useful pairs only, the negative pair was formed only with images whose diagonal (a proxy for vessel length) is within 30% of the anchor image diagonal. Indeed, vessels with drastically different lengths can be ruled out upfront, as they cannot represent the same individual vessel. In constructing the negative pair, to ensure diversity, vessels with single image occurrences were used 50% percent of the time. After several experiments involving different architectures and hyperparameter values, the best performing network featured a ResNet-50 backbone with pre-trained weights from ImageNet and a final embedding size n of 100. We used a margin value m of 5 and a threshold set at 2.5 (positive pair if distance $d <$ threshold). Various image augmentations were applied during the training: random changes to image brightness and contrast, horizontal and vertical flips, 360° rotations, and slight re-scaling of the images. The validation set was used to select the best model based on the overall accuracy in the binary classification of the pairs based on the distance threshold. PyTorch (Paszke et al., 2019) was used for the implementation.
As shown in Figure 5, the best model was obtained at epoch 21 with a binary validation accuracy of 82.5%.

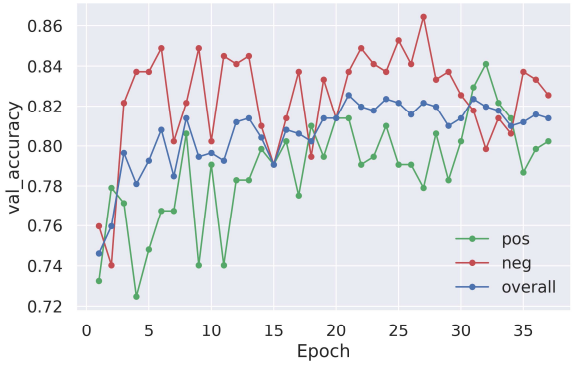


Figure 5. Accuracies on the validation set for the Twin network with a ResNet-50 backbone and ImageNet pre-trained weights. Overall accuracy in blue, class-specific accuracies (recall) for the positive and negative pairs in green and red, respectively.

To visualize the network results, Figure 6 shows examples for four image-pairs, corresponding to the four possible binary classification outcomes. Figure 6(a) features an image pair correctly identified as belonging to the same vessel (True Positive, distance score d of 0.70 < 2.5): the same tanker is observed in open-waters and while docked. Figure 6(b) shows an image pair correctly identified as representing different vessels (True Negative): two different tankers resulting in a high distance (5.78 > 2.5). In Figure 6(c), the model incorrectly predicted two images as being of the same vessel (False Positive), while in Figure 6(d) it failed to identify two images of the same vessel (False Negative, with a score of 2.94 close, yet not below the threshold of 2.5).

(a) True Positive.

(b) True Negative.

(c) False Positive.

(d) False Negative.

Figure 6. Examples of vessel identification results for our Twin network model from the validation set (best epoch). The identification score d (Euclidian distance between the two images) is show on the top left.

3.2.3 Assessment in a real-world re-identification scenario: The objective of the absolute identification of each vessel (as done for person identification), is highly ambitious due to the presence of standard models for many vessel types and the changes in appearance a vessel can undergo over its lifetime. A more realistic objective is that of re-identifying a target vessel among a finite set of candidates by ranking their relevance, instead of returning an absolute match.

To this end, we evaluate our Twin model by using the independent test set of 515 vessels (each vessel with two or more image occurrences) by setting up a vessel re-identification scenario. Starting from a Vessel of Interest (VOI) observed/detected in given large-scale remote sensing acquisition (one image occurrence for a test set vessel, i.e., the probe image), we assess the model's capability in returning a ranked list of relevant results from a set of previously observed vessels available in a database (all the other test set vessels). The following procedure is applied:

1) For each individual image in the test set (across all vessels):
   a) Set the vessel of the current image as VOI.
   b) Select all the other available images for this VOI (previous "looks" for the VOI), excluding the original one.
   c) Select all images of the other vessels in the test set with the condition that their length is within 30% of the original image (this represents the candidate ships in the database, each with a series of previous "looks").
   d) Obtain the embeddings for the original image of the current VOI and all images selected in step b) and c).
   e) Compute the average Euclidean distance between the original image and each vessel (average of the distances to all image occurrences for a given vessel).
   f) Sort these distances from smallest to largest and compute the rank of the VOI among all candidates (ideally the VOI should rank first).
2) Based on the retrieved ranks for each individual image in the test set, compute the overall metrics:
   - top-*k* accuracy
   - Mean Reciprocal Rank (MRR) (Voorhees and others, 1999)

The Top-*k* Accuracy metric (range in [0, 1], with best at 1 = 100%) for various values of *k* conveys the model's ability to rank, on average, the target vessel within the top-*k* of searched vessels, based on the sorted score associated with each instance (the distance in our case). This metric, if plotted for all values of k, results in a Cumulative Matching Characteristic (CMC) curve (Farenzena et al., 2010).

MRR (range in [0, 1], with best at 1) is a measure of the model's average ability in returning a ranked list in which the VOI is as close as possible to the 1st rank. It is computed as

$$\mathrm{MRR} = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{\mathrm{rank}_i} \quad , \qquad (2)$$

where rank_i is the rank of the relevant result in the *i*-th query and $|Q|$ is the total number of queries (the number of vessel images in the test set, in our case).

Following the procedure outlined above, we evaluated the model's ability to return a sorted list of relevant vessels from a database (the 515 vessels in the test set) when observing one instance of a given vessel. The model showed a very promising retrieval performance with a top-1 accuracy of 38.7% and top-10 accuracy of 76.5%, meaning that the correct vessel was returned in the 1st position (out of 515) of the ranked list in more than 1/3 of the cases and within the first 10 positions in more than 3/4 of the cases. To provide the overall picture, the CMC plot of Figure 7 outlines the model's performance in terms of top-*k* accuracy for values of *k* between 1 and 50. The associated MRR was 0.512, indicating a strong performance in returning the relevant vessel high in the ranking (~ rank 2 on average).

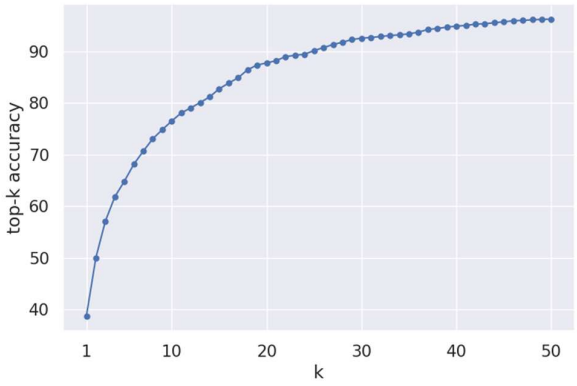


Figure 7. CMC plot of top-*k* accuracies for all *k* in [1, 2, …, 50] on the test set counting 515 vessels.

Figure 8 shows an example of the closest matches returned by the model for a vessel of the test set (cargo vessel with MMSI number 357358000). Our Twin network was in this case able to return the correct vessel identity at rank number 1, with an average distance to the probe image of 1.07. The top-5 ranked vessels are all visually similar (large red cargo vessels), confirming the ability of our model to appropriately rank satellite images of ships based on their appearance. The performance of the model is affected by factors such as vessel type (some vessel classes present less variations in shape/color than others) and vessel size in relation to the satellite image resolution (with fewer pixels representing a vessel, it is more difficult to isolate distinctive features for each individual).

4. CONCLUSIONS

In this work we demonstrated the opportunities for maritime surveillance offered by VHR optical imagery when analyzed via state-of-the-art deep learning methods. The powerful CNN-based methods applied herein allow users or interested parties to automate time-consuming image interpretation tasks such as localizing vessels and identifying them. Flagging only the most relevant samples for an operator to review immensely reduces the cognitive load and analysis time. Our Twin neural network trained on a real-world dataset with 2500+ unique vessels was able to effectively rank candidate vessels from an independent test set, with top-1 and top-10 accuracies of 38.7% and 76.5%, respectively.

Initial promising results were obtained in assessing our model in such a ranking scenario at inference time, however more effort is needed to devise real-life scenarios simulating challenges faced in operation (e.g., re-identification of vessels based on a fixed set of candidates observed in previously acquired satellite images).

To further advance the vessel identification capabilities to support vessel tracking and re-identification, we plan on moving away from treating the problem as a binary classification problem (positive vs. negative pair). Explicitly considering the ranking of the candidate samples during training (Cakir et al., 2019) would allow us to improve the relevance of the return list of candidates. Adopting a self-supervised learning approach to better utilize the large volumes of unlabeled data available in remote sensing is also a promising avenue of research. Other types of loss function could be tested, as proposed in the deep metric learning framework (Hoffer and Ailon, 2015). Additionally, models

Figure 8. Top-5 ranked vessels (each row shows the available images of each candidate) as retrieved by the Twin network for a specific probe image in the test set (vessel with MMSI 357358000). The green checkmark indicates the rank of the corresponding correct MMSI.

considering the separate subparts of the object of interest could also be explored (Zhang et al., 2020).

Ultimately, the goal is to fit the developed techniques into the broader context of vessel tracking in a tip-and-cue scenario. This will involve satellite tasking to execute targeted image collections following the predicted track of a given vessel. To ensure greater coverage, multiple imaging platforms should be involved, stressing the importance of developing models that are sensor-agnostic.

ACKNOWLEDGEMENTS

The authors would like to thank Jean-Pierre Ardouin (Defense Research and Development Canada) for his insights and the support during all the phases of this work. The authors acknowledge the Government of Canada and the Defence Innovation Research Program (DIRP) for supporting this project. Ken Wong and Leigh Martin-Boyd are acknowledged for the project management.

REFERENCES

Cakir, F., He, K., Xia, X., Kulis, B., Sclaroff, S., 2019. Deep Metric Learning to Rank, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA, pp. 1861–1870. https://doi.org/10.1109/CVPR.2019.00196

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A., 2020. Unsupervised learning of visual features by contrasting cluster assignments. arXiv preprint arXiv:2006.09882.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations, in: International Conference on Machine Learning. PMLR, pp. 1597–1607.

Chu, R., Sun, Y., Li, Y., Liu, Z., Zhang, C., Wei, Y., 2019. Vehicle Re-Identification With Viewpoint-Aware Metric Learning, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, Seoul, South Korea, pp. 8281–8290. https://doi.org/10.1109/ICCV.2019.00837

Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M., 2010. Person re-identification by symmetry-driven accumulation of local features, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2360–2367.

Gaiser, H., de Vries, M., 2019. Keras RetinaNet. Fizyr.

Gundogdu, E., Solmaz, B., Yücesoy, V., Koç, A., 2017. MARVEL: A Large-Scale Image Dataset for Maritime Vessels, in: Lai, S.-H., Lepetit, V., Nishino, K., Sato, Y. (Eds.), Computer Vision – ACCV 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 165–180. https://doi.org/10.1007/978-3-319-54193-8_11

Heiselberg, H., 2016. A Direct and Fast Methodology for Ship Recognition in Sentinel-2 Multispectral Imagery. Remote Sensing 8, 1033. https://doi.org/10.3390/rs8121033

Hoffer, E., Ailon, N., 2015. Deep metric learning using triplet network, in: International Workshop on Similarity-Based Pattern Recognition. Springer, pp. 84–92.

Ji, X., Henriques, J.F., Vedaldi, A., 2019. Invariant information clustering for unsupervised image classification and segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9865–9874.

Kanjir, U., Greidanus, H., Oštir, K., 2018. Vessel detection and classification from spaceborne optical images: A literature survey. Remote Sensing of Environment 207, 1–26. https://doi.org/10.1016/j.rse.2017.12.033

Koch, G., 2015. Siamese Neural Networks for One-Shot Image Recognition (MSc). University of Toronto, Toronto.

Lam, D., Kuzma, R., McGee, K., Dooley, S., Laielli, M., Klaric, M., Bulatov, Y., McCord, B., 2018. xView: Objects in Context in Overhead Imagery. arXiv:1802.07856 [cs].

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F. d\textquotesingle, Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 8024–8035.

Qiao, D., Liu, G., Dong, F., Jiang, S.-X., Dai, L., 2020. Marine Vessel Re-Identification: A Large-Scale Dataset and Global-and-Local Fusion-Based Discriminative Feature Learning. IEEE Access 8, 27744–27756. https://doi.org/10.1109/ACCESS.2020.2969231

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497 [cs]

Stasolla, M., Mallorqui, J.J., Margarit, G., Santamaria, C., Walker, N., 2016. A comparative study of operational vessel detectors for maritime surveillance using satellite-borne synthetic aperture radar. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 2687–2701.

Tournadre, J., 2014. Anthropogenic pressure on the open ocean: The growth of ship traffic revealed by altimeter data analysis. Geophysical Research Letters 41, 7924–7932.

Tunaley, J., 2004. Algorithms for ship detection and tracking using satellite imagery, in: IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1804–1807.

Voorhees, E.M., others, 1999. The TREC-8 question answering track report, in: Trec. pp. 77–82.

Weinberger, K.Q., Saul, L.K., 2009. Distance metric learning for large margin nearest neighbor classification. Journal of machine learning research 10.

Wu, Y., others, 2016. Tensorpack.

Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., Guo, Z., 2018. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. Remote Sensing 10, 132.

Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., Hoi, S.C.H., 2021. Deep Learning for Person Re-identification: A Survey and Outlook. IEEE Trans Pattern Anal Mach Intell PP. https://doi.org/10.1109/TPAMI.2021.3054775

Zhang, S., Wu, R., Xu, K., Wang, J., Sun, W., 2019. R-CNN-Based Ship Detection from High Resolution Remote Sensing Imagery. Remote Sensing 11, 631. https://doi.org/10.3390/rs11060631

Zhang, X., Zhang, R., Cao, J., Gong, D., You, M., Shen, C., 2020. Part-Guided Attention Learning for Vehicle Instance Retrieval. IEEE Transactions on Intelligent Transportation Systems 1–13. https://doi.org/10.1109/TITS.2020.3030301

Zheng, Z., Ruan, T., Wei, Y., Yang, Y., 2019. VehicleNet: Learning Robust Feature Representation for Vehicle Re-identification., in: CVPR Workshops. p. 3.

Zhong, Z., Zheng, L., Luo, Z., Li, S., Yang, Y., 2019. Invariance Matters: Exemplar Memory for Domain Adaptive Person Re-Identification, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Long Beach, CA, USA, pp. 598–607. https://doi.org/10.1109/CVPR.2019.00069

Zhu, Z., Diao, W., Chen, K., Zhao, L., Yan, Z., Zhang, W., Xu, G., Sun, X., 2020. DiamondNet: Ship Detection In Remote Sensing Images By Extracting And Clustering Keypoints In A Diamond, in: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. Presented at the XXIV ISPRS Congress, Commission II (Volume V-2-2020) - 2020 edition, Copernicus GmbH, pp. 625–632. https://doi.org/10.5194/isprs-annals-V-2-2020-625-2020