

Self-Supervised Vessel Detection from Low Resolution Satellite Imagery

A. Farr^{*} and I. Petrunin[†]

*Cranfield University School of Aerospace Transport and Manufacturing,
Cranfield, Central Bedfordshire, United Kingdom.*

G. Kakareko[‡]

Spire Global, Inc., Boulder, CO, United States.

J. Cappaert[§]

Spire Global, Inc., Luxembourg.

Maritime surveillance is a growing field of application for optical imaging satellites. In this paper, we aim to increase the practical effectiveness of vessel detection algorithms by developing a computationally inexpensive method that can discern vessels in low resolution optical satellite imagery. This creates an opportunity to host the algorithm on-board small, low-cost nanosatellites, generating feasibility for large surveillance constellations. The presented algorithm, which is based upon a deep convolutional autoencoder utilized as an anomaly detector, was found to achieve a detection precision of 98%, and a recall of 79%. Data used to test and train the algorithm was produced by augmenting ESA Sentinel-2 imagery, reducing the resolution to mimic that of a nanosatellite's. The results presented here are compared with comparable papers in the field, where it is demonstrated that the proposed algorithm is capable of outperforming classical techniques in terms of detection precision. For an equivalent resolution, the proposed algorithm provides a 34% increase in precision over the Constant False Alarm Rate (CFAR) technique.

I. Introduction

THE interest of monitoring and assessing activities at sea is growing. In order to ensure that the safety, security, environment, and economy of the Earth's oceans are upheld, organisations such as the International Maritime Organisation employ a variety of remote sensing techniques. Areas of specific concern for maritime surveillance include the likes of illegal fishing, piracy operations, routine traffic management, and military surveillance.

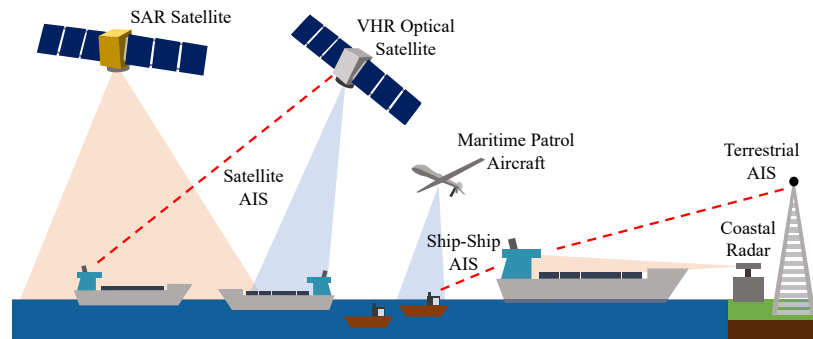


Fig. 1 Current technologies present in maritime surveillance.

^{*}Research Student, Cranfield University Cranfield School of Aerospace Transport and Manufacturing, Cranfield, United Kingdom.

[†]Lecturer, Cranfield University Cranfield School of Aerospace Transport and Manufacturing, Cranfield, United Kingdom.

[‡]Head of Machine Learning, Spire Global, Inc. Boulder, CO, United States.

[§]CTO & Co-founder, Space Program, Spire Global, Inc., Luxembourg

Maritime surveillance is performed using a vast ecosystem of sensors (Fig. 1). The lack of ability for conventional terrestrial systems to monitor activities in remote regions of ocean has created a need for satellite based monitoring platforms. The predominant satellite remote sensing system at present is SAR (Synthetic Aperture Radar) [1]. SAR has a number of advantages over its optical counterpart; including a heavily reduced influence from weather, illumination, and sea conditions [1]. Despite this however, SAR is not a viable platform more widespread maritime surveillance. SAR suffers from high levels of intrinsic noise, which greatly impacts classification ability [2]. Smaller vessels, which tend to be made of wood or fiberglass, or those which employ stealth technologies, make certain targets difficult to detect [1]. Additionally, at the time of Kanjir et al.'s publication in 2017, there existed around a dozen SAR satellites in orbit [1]. This limited number results in a poor spatiotemporal coverage of the Earth's oceans, rendering SAR data insufficient for near-constant maritime monitoring.

Satellite optical imaging is currently a secondary platform for satellite maritime surveillance. However, due to the increasing number of earth observation satellites in-service (over 900 as of August 2021 [3]), and the drawbacks of SAR, optical imaging is becoming an increasingly critical part of effective maritime domain awareness. At present, almost all existing literature concerns itself with a category of optical data known as Very High Resolution (VHR) imagery. In this resolution range (defined by Kanjir et al. as images with a Ground Sample Distance (GSD) $< 5\text{m}$ [1]), smaller vessels can be clearly elicited, and characteristics such as shape, size, direction and class can be readily determined. High resolution optical platforms have several drawbacks. Their resolutions result in massive amounts of data that need to be transmitted and pre-processed, which is slow and computationally expensive. Additionally, VHR satellites often only perform image acquisition on-demand, with a relatively small coverage areas [4]. To overcome these issues, and to make optical imagery viable for surveillance, industry is steering toward nanosatellite constellations fitted with medium resolution sensors [5]. These systems allow for effective, near real-time surveillance of Earth's oceans, with suitable spatiotemporal coverage. Traditionally, these factors often came with a compromise of reduced on-board computing ability and poor sensor resolution, however recent advances in the qualification of COTS components for space has seen complex algorithms being introduced onto more capable flight hardware [4, 6].

The aims of this paper are two-fold. First, we aim to increase the practical effectiveness of vessel detection algorithms by creating a computationally inexpensive method that can discern vessels in medium resolution ($\sim 50\text{ m}$ GSD) optical imagery. Second, we aim to utilize the field of autoencoder-based anomaly detection methods, that will support an improvement in detection precision and operational efficiency over existing techniques.

The remaining sections of this paper are composed as follows. First in Section II, a brief overview of existing comparable algorithms are provided. This is followed in Section III by an introduction and justification of the proposed vessel detection strategy, which utilizes the concept of convolutional autoencoders to elicit vessel candidates at sea. The method is then tested, and the results of the algorithm compared with existing papers in the field in Section IV. Section V concludes the paper and introduces avenues of further work based off of the findings of this paper.

II. Related Works

A. Vessel Detection Methods Based on Low Resolution Satellite Imagery

As previously mentioned, the vast majority of papers in the field of optical vessel detection focus exclusively on high resolution imagery [1]. As a result techniques developed within the field, such as shape and space, or texture-based methods, often make use of the large amount of features present within the data. As seen in Fig. 2, lower resolution images do not contain this rich feature information. Whilst the tug boats are clearly discernible from the tanker vessel in higher resolutions, they become increasingly hard to categorize, or even detect, as GSD increases, due to a reducing amount of features to infer from. This lack of features has required authors to turn to other mechanisms of detection in an attempt to produce desirable performances.

Ghosh et al. are one such group of authors who operate on medium resolution (42m GSD) optical imagery [4]. They first gather suitable images by using the raw image data from the Resourcesat-2 satellite, implementing a number of pre-processing steps including the reduction of noise and residual strips. Once they have suitable images, they perform candidate detection through a statistical method known as Constant False Alarm Rate (CFAR) [4].

CFAR is an anomaly detection method typically found in SAR systems. It often outperforms traditional threshold based methods, as it uses the context of a smaller area to detect potential targets, rather than averaging over a whole image [8]. Reviewing the literature of authors who operated on higher resolution images, very few of them utilize CFAR techniques. However, it appears to be the dominant method for optical images which exceed a resolution of 40m GSD, with other papers including those by Liu et al. [8, 9] utilizing this strategy. Liu et al. first praise deep learning for its

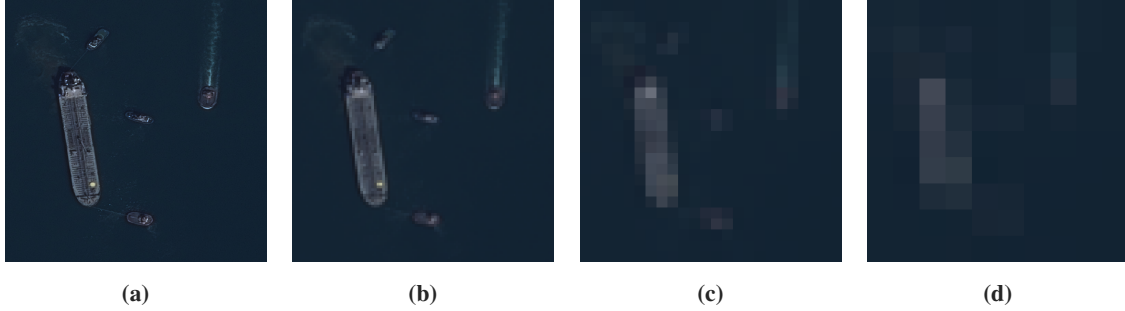


Fig. 2 A comparison between GSDs. Image source: ESA [7].
(a) 0.7m, (b) 5m, (c) 20m (d) 50m

abilities when operating on high resolution imagery, but go on to state that they may not be feasible for lower resolutions due to a lack of available features[8]. The use of deep learning techniques for low resolution vessel detection was not investigated in detail in their study however, which invites an investigation of their claims in this paper.

1. Comparable Performances

In order to assist in the evaluation of the proposed algorithm's performance, authors who have used medium resolution imagery (>20 m GSD) have had their performance metrics assessed. Note that existing papers in the field of low-resolution vessel detection are extremely limited. Additionally, some of the papers which do use low resolutions like one of Liu et al.'s [8], do not provide an evaluation of results; an issue common in optical vessel detection, as noted by Kanjir et al. in their literature review of the field [1].

Ding et al.'s paper use a saliency based method to achieve a very high level of precision [10]. Despite this result, it is difficult to compare the performance of their algorithm with others, given the differences in test methodology and lack of test metrics provided. They test their algorithm on simulated vessel images when quoting the algorithms performance [10], unlike the others, who use real captured images of vessels at sea [4, 9]. Ghosh et al. [4] provide more information when evaluating their performance, although the number of test images used is not mentioned. Instead of simulating images or augmenting a pre-built dataset, they use a human operator to identify vessel targets within test images, and set this number as ground truth (N_{real}). This could have resulted in some missed candidates and false alarms within an image, particularly with smaller vessels in lower resolutions, where a human operator may make a mistake.

Liu et al.'s [9] algorithm appears to have the worst performance out of the three authors; which reflects the fact that they used the lowest resolution. They produced their test set through acquiring 5 frames each from 2 different regions of the South China Sea. The first region alone provides a precision average of 65.71% and a recall of 88.52% [9], which outperforms Ghosh et al.'s algorithm. Liu et al's second region affects the performance due to increased environmental disturbances. They state the conditions of the sea were worse in the second region during testing times. Liu et al. utilize AIS signals of the South China Sea to provide a ground truth for their algorithm [9]. As their satellite platform (GF-4) is based in GEO, they can quite easily match known vessel positions with ones that have been detected or missed. They state that their algorithm mainly missed vessels in the 50-70 m class [9]. Smaller vessels, which do not have AIS turned on, will likely not have been detected or considered in their performance review.

The massive variation in the construction of test sets, the information on criteria for which true vessel numbers are attained, and the sea environments which are used in testing make a comparison of results extremely difficult. Various authors in literature recommend the usage of third party test sets in order to standardize results [1, 11], although there does not yet exist one of such for low resolution detection problems.

B. Deep Learning in Vessel Detection

Deep learning is still being used to a limited extent in vessel detection problems. Nevertheless, initial studies into applying deep-learning techniques to vessel detection have yielded promising results. The field of deep-learning can be broadly divided into categories based on the type of data it operates on. Supervised learning, where training data is labeled based on class, or unsupervised, where the algorithms only receive unlabeled training data to operate on. There are no corresponding outputs that the algorithm aims to map to, which results in unsupervised learning often being the approach adopted to discover new or unexpected features and trends in data. Unsupervised learning also requires far

less items to train on, increasing its feasibility to be hosted on-orbit.

A common method of the unsupervised group of techniques is the autoencoder; such a method has already been used in higher resolution and SAR imagery, where it has demonstrated impressive performance [12–14]. Autoencoders take an input, in this case an image of the sea, which perhaps contains a vessel, and extracts the most representative features of that image in a compressed form into what is known as a latent space. This compression is performed using a series of hidden encoding layers, which aim to retain characteristic features surrounding the input. The ability for the autoencoder to retain only vital features is ensured through an optimization process, where it attempts to reconstruct the original input from the latent code, while minimizing the difference between the input and output through a suitable loss function. By training the deep-learning based autoencoder on solely split image tiles of the sea, the autoencoder ‘learns’ how to map nominal images into the latent space. When fed with an image that contains a vessel, the representation of the image within latent space is considerably different than the nominal image, allowing suspects to be elicited [12]. Tang et al. utilize Stacked Denoising Autoencoders (SDAs) for feature extraction in their wavelet-based detection strategy [13], whereas Ferreira and Silveria utilize a deep variational autoencoder to implement anomaly detection on SAR imagery [12]. Whilst Ferreira and Silveria’s paper [12] does not utilize optical imagery as an input, it serves to prove that autoencoders are capable of operating on a variety of different types of data, provided they are trained and initialized correctly. The SAR imagery they utilized ranged extensively in resolution, suggesting that autoencoders are capable of operating within low feature detection problems.

C. On-Orbit Processing

On-orbit processing is an emerging field in the setting of deep-learning based earth observation problems. Initial implementations, such as the CloudScout neural network on-board ESA’s Φ -Sat 1, have yielded promising results [6, 15]. Implementing neural networks on-orbit provides several advantages to small satellite operations, including a heavily reduced communications bandwidth [16], improved scalability, and the capability for each satellite to be tuned and re-trained with its captured raw data [17], tailoring an algorithm to a specific end device. Despite these impressive benefits, such systems also have several key challenges that must be overcome.

Most existing On-Board Computers (OBCs) built for satellite operations were not designed with machine learning tasks in mind, and as a result lack the computational resources to operate them effectively [17]. This is also true for other aspects, e.g. of the satellite bus, including increased demands for the power and thermal subsystems [17]. Due to the infancy of applying machine learning technologies on-orbit, sourcing suitable datasets for training and testing algorithms prior to launch is extremely difficult [16, 17]. Additionally, the space environment itself poses a huge challenge in itself for sourcing capable flight hardware. Due to a desirable requirement to use Graphics Processing Units (GPUs) in favor of Central Processing Units (CPUs), ML capable hardware models can expect to be almost three times as likely to fail than traditional ones [16]. Detailed discussions and mitigation actions to these concerns can be found in the works of Furano et al. [17] and Di Mascio et al. [16].

III. Autoencoder-Based Vessel Detection

The impressive performances that authors adopting deep-learning based approaches were able to obtain in both high resolution and SAR imagery [12, 13] invites the development of a novel detection algorithm using such methods for lower resolution optical imagery. The relative infancy of deep-learning methods in the field also offers a unique chance to develop these technologies with on-orbit experimentation in-mind, producing a robust and novel algorithm that can contribute to effective maritime surveillance.

As the algorithm is intended for use on-orbit, supervised methods of machine learning (such as image classification) are highly undesirable. Re-training a supervised algorithm on-orbit requires vast amounts of computational resources and storage space (the Airbus ship detection dataset consists of almost 193,000 JPEG compressed images, taking up over 27 GB of storage). Furthermore, there are no publicly available medium resolution datasets that reflect the image performance of a typical nanosatellite. Unsupervised methods provide an advantage in terms of computational load, as well as their ability to learn the inherent structure of the data they are trained upon [18].

With this consideration in mind, the vessel detection algorithm was developed as seen in Fig. 3. The proposed approach to achieving such an algorithm is as follows: first, synthetic nanosatellite images were produced by augmenting publicly available satellite imagery. These images then underwent a pre-processing phase to perform land-masking and are then used to generate suitably sized datasets for the purposes of training, testing, and validating the finished deep-learning algorithm.

For the detection phase of the algorithm, an anomaly detection based strategy was adopted based on a convolutional

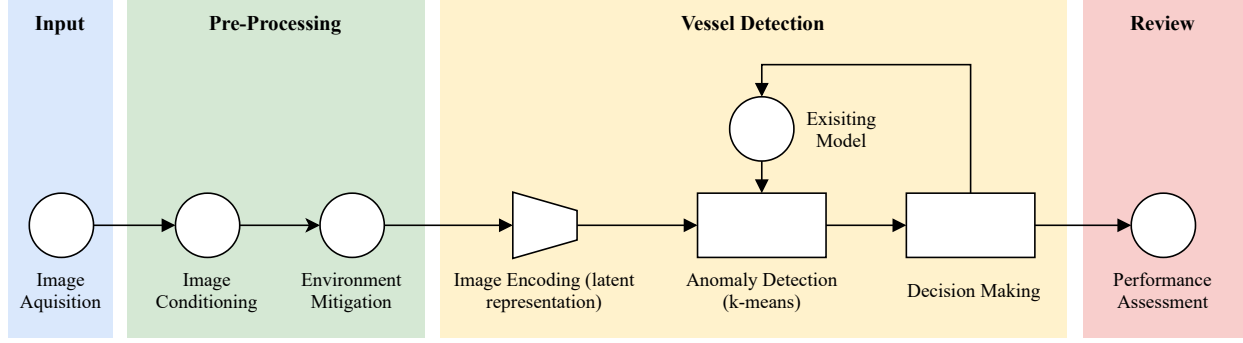


Fig. 3 A block diagram of the vessel detection algorithm for low-resolution satellite imagery

autoencoder and K-Means clustering. The autoencoder is first trained on the synthetic images, and then the encoder extracted from this to produce a latent representation of inputted images. Inputted images are fed through in groups, where K-Means clustering is then applied identify groupings in the data. The number of clusters is set equal to 2, which corresponds to a grouping of nominal images of the sea, and anomalous images, which could potentially contain a vessel. The largest cluster identified is assigned the label of sea, and the smallest cluster is identified as a grouping of potential vessel candidates. K-Means clustering was selected over other unsupervised clustering methods due to its favorably low computational intensity [19] and usage in previous vessel detection work [12]. Finally, a computer vision based discriminator is then applied to determine whether or not a vessel is present within the suspected image. If the image is deemed not to be a vessel, the image is reviewed before being fed into the training dataset for the means of updating the model.

A. Dataset Creation

Before embarking on producing a custom dataset for this paper, existing datasets were considered. Utilizing an existing dataset can both speed up development time and also provide a standardized testing dataset which can assist in performance evaluation [1]. Unfortunately, due to a variety of factors relating to the resolution and source platforms that the datasets were built from, these datasets were not suitable for use in this work. This is an issue seen in other papers in the field [4, 6], which has lead to authors creating their own novel datasets for algorithms; an approach which can also be seen implemented here.

To produce a suitably sized dataset that reflects a sensor representative of one found onboard a satellite, ESA Sentinel-2 imagery was acquired from the Copernicus interface. This imagery is then augmented through a series of preprocessing steps using the GDAL and Rasterio Python libraries to reduce the resolution by a scale factor; corresponding to a GSD conversion from Sentinel-2's 10 m to a typical nanosatellite's 45 m. Sea-land segmentation was then implemented, through means of the European Environment Agency's coastline shape file with a 1 km buffer zone.

A decision was made to split the produced land-masked medium-resolution images into a number of equally sized tiles, a technique which can be seen adopted in other works [20, 21]. Having a large image split into smaller tiles assists in batch-processing, allowing multiple portions of the same image to be operated on simultaneously. The images were split equally into $(n_w, n_h, n_c) = (128, 128, 3)$ tiles, where w, h, c represent the width, height, and channel count respectively. A 10% overlap between images was also implemented to prevent vessel candidates on tile edges from being cropped. This approach can also be seen adopted in the work of Song et al. [22]. Once these tiles were produced, individual tiles were selected for training and testing purposes. The training and validation datasets consists of 8500 sea tiles from 10 Sentinel-2 images, whereas the testing dataset consists of 1600 sea and boat tiles from 5 Sentinel-2 images. These images were labeled, purely for evaluation purposes, to monitor how effectively the anomaly detection algorithm could discern vessels from sea. Some sample images containing vessels can be viewed in Fig 6.

Using a human operator to select images for datasets was seen as undesirable, despite it being the strategy used for the comparable paper produced by Ghosh et al. [4]. Vessels which are considerably smaller than the GSD of 42.9 m will be represented by a slight change of intensity within a single image pixel. These smaller vessels can easily be interpreted as noise within the image to a human operator, but may impact the performance of the algorithm if incorrectly used as training data. To help mitigate this occurrence, sea tiles were produced and carefully inspected under Sentinel 2's native 10 m GSD, if the image was found to contain no vessel, the corresponding 42.9 m GSD tile was selected for training. To

help boost performance, the produced tiles were processed through the Keras library's data generator. The generator was able to double the amount of training data to 17,000 through random rotations, crops, flips, and channel range adjustments. The testing dataset consisted of 1600 image tiles, of that set, 800 were of purely sea tiles and 800 were image tiles that contained a vessel. The size and orientation of vessels present within an image varied greatly, as did the sea state conditions.

B. Anomaly Detection

The autoencoder, an unsupervised learning technique that was seen in the papers discussed in section II, offers promising results in comparable systems, inviting an investigation into their performance on low resolution imagery. An anomaly detection approach is adopted by training an autoencoder on solely images of the sea, where the training dataset can be quickly and efficiently constructed. Images containing vessels can be elicited by observing the image's representation in latent space, images which do not belong to the grouping of sea images are then extracted as suspected candidates, where a computer-vision based length-to-width estimator is applied as a discriminator.

As mentioned, autoencoders can be used to detect the presence of an anomaly through their inherent ability to learn representations of images they are trained on. They use an encoder (ϕ), to map an original input ($x \in \mathbb{R}^d = X$) into a compressed representation in code ($z \in \mathbb{R}^p = \mathcal{F}$), which is also known as the latent space. Often, this space is of a lower dimensionality than the original input's dimensions ($p < d$). The decoder (φ) then takes this latent code and attempts to reconstruct the original input (x^*). During the training process, the input and output are compared using a suitable objective function (f_{obj}), and the autoencoder's weightings are then adjusted via a back-propagation algorithm.

The objective function that was selected for use in this paper is known as the Structural Similarity Index (SSIM), described in Eq. (1).

$$f_{obj} = 1 - SSIM(\mathbf{x}, \mathbf{x}^*) = 1 - \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{x}^*} + c_1)(2\sigma_{\mathbf{x}\mathbf{x}^*} + c_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{x}^*}^2 + c_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{x}^*}^2 + c_2)} \quad (1)$$

SSIM produces a distance measure between an input (\mathbf{x}), and its reconstructed output (\mathbf{x}^*). It does so by accounting for the similarity in luminance, contrast, and structure [23]. The constants c_1 and c_2 , which are used to ensure numerical stability, have been set to 0.01 and 0.03, following the advice of Bergmann et al. [23]. By using such an objective function, the autoencoder attempts to reconstruct image tiles of sea, taking into account the output image's change in luminance, contrast, and structure. When a vessel is introduced into the image, these three parameters change dramatically, producing an anomalous representation within the latent vector. SSIM was found experimentally to offer the greatest performance for this problem over mean-squared error and binary cross-entropy, coinciding with the findings of Bergmann et al. [23]. To the authors knowledge, this objective function has not yet been used in the field of vessel detection, providing a novelty to the field.

C. Convolutional Encoder

The CNN based encoder model was constructed as shown in Fig. 4. First, the created inputs were normalized between the values of 0 and 1 (a technique used to improve convergence time [6]), and were then fed through into the encoder model using the Keras library input layer, hosted within the TensorFlow Python libraries.

Images with three channels (RGB) (n_h, n_w, n_c) = (128,128,3) and single channel (n_h, n_w) = (128,128) were tested as inputs to the CAE due to conflicting claims of performance in literature [10, 24]. As seen in Fig. 4, the encoder model consists of both a feature learning and a dimensionality reduction phase. The feature learning phase consists of three similar groupings of neural network layers. Their filter size sequentially increases between groups in the order of 8, 16, and 32. All three have a kernel size of 3, a stride of 1, with a same-padding approach to detect vessels in edge positions. The outputs of each convolutional layer are regularized through batch normalization, smoothing the training gradient, before being passed into the activation function, which was chosen experimentally to be leaky ReLU, which has been used to improve performance of the networks in [25]. Finally, a dropout layer is incorporated to reduce the chances of overfitting and the equivariance of representations is countered between layer groupings through the means of a maximum pooling layer.

After passing through the three layered groupings, the subsequent output is flattened and reduced using a dense layer to produce a latent vector. Flattening is performed to produce a one-dimensional representation of the learned features, and the dense layer (which applies a matrix vector multiplication to its inputs) is used to reduce the dimensionality of

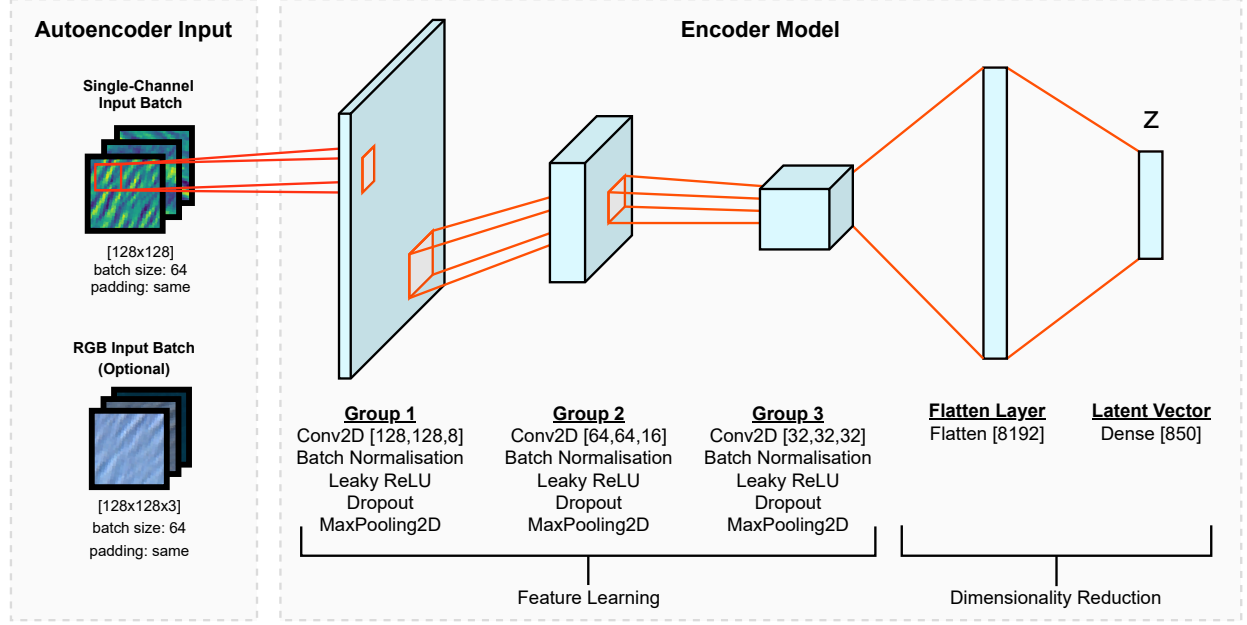


Fig. 4 A simplified diagram of the convolutional encoder. Inputs (X) are first input and passed through three feature-learning layer groups, before being flattened and reduced into a latent vector of size 850.

the flattened group output to the desired latent vector size, which was chosen through experimentation to be 850, as it offered the best separation between clusters within the latent space. This latent vector z , which is the output of the encoder, is used as both the input to the decoder and for means of anomaly detection.

In order to extract and identify potential vessel candidates, the latent representations of a set of images is gathered and the K-Means clustering algorithm performed on them. This process, described earlier in Section III, produces a grouping of potential vessel candidates that are passed onto a discriminator phase. The discriminator, which consists of a series of computer vision techniques, reviews the original source image of the suspect using a luminosity threshold, followed by a length-to-width discriminator of the threshold map. If a vessel is found not to be present within an image, this image is kept for means of re-training the existing model.

D. Convolutional Decoder

Whilst not strictly required for the vessel detection phase, the decoder is still necessary in order to train the autoencoder. It takes the output of the encoder, which is the latent vector z , and produces a reconstructed image through a series of layers. During training, this reconstructed image is compared with the original image through the means of the SSIM objective function, which is then fed to the back-propagation algorithm. The decoder was designed to be symmetrical to the encoder through empirical evaluation, a technique that have been seen used by other researchers in the field [12]. The model was trained for 200 epochs on the shuffled training dataset, with a batch size of 128, using the Adam optimizer and a learning rate of $1 \cdot e^{-4}$.

IV. Results & Discussion

The first part of the algorithm to be reviewed is the candidate detection phase. Two separate autoencoder models were trained, similar in architecture, with only difference in using inputs of single and three channel images respectively. These models were then tested using a set consisting of 800 confirmed vessel images and 1000 images of sea (1800 total).

Figure 5 demonstrates the autoencoders ability to separate nominal sea images and anomalies. For the purposes of visualization, these latent spaces have had their dimensionality reduced (after clustering) from 850 to a 2-dimensional sub-space through the traditional means of Principal Component Analysis (PCA). The result of this is shown in Fig. 5, which exhibits the clustered, 2D representations of the latent space from the single channel trained CAE. It is evident

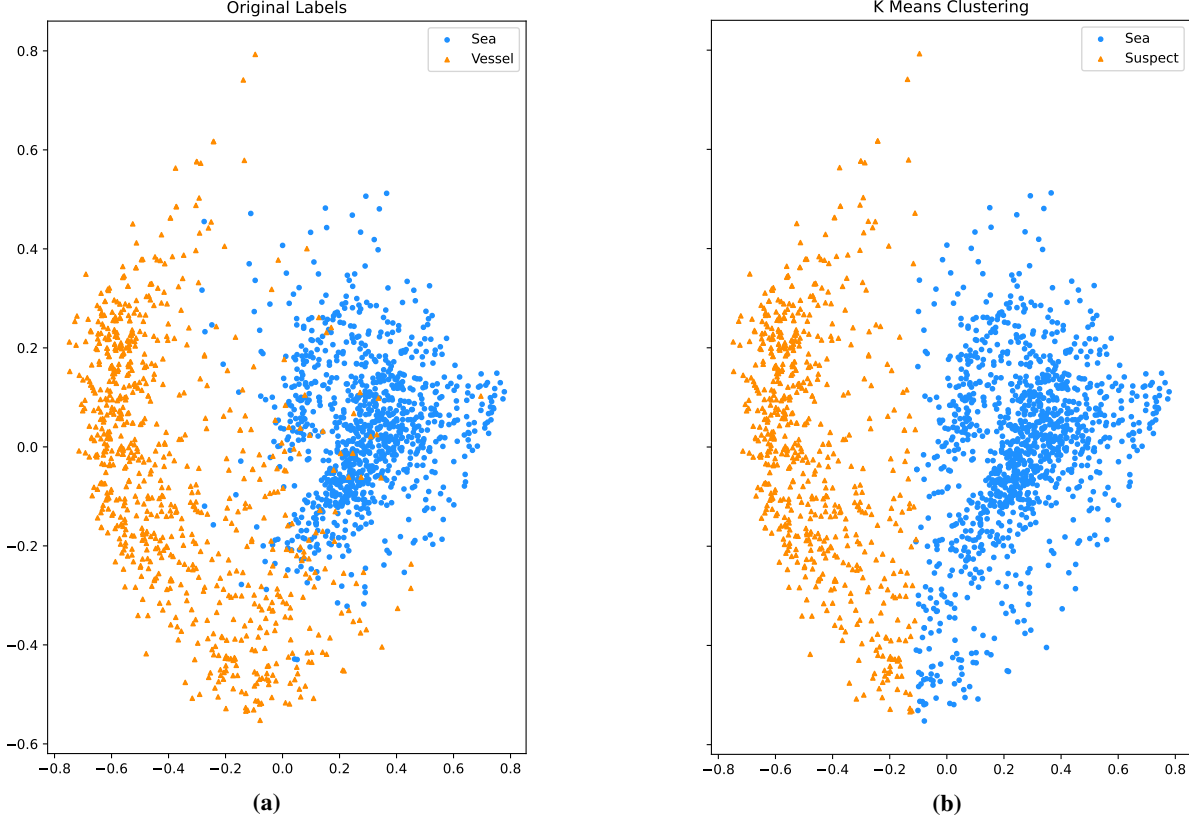


Fig. 5 2D PCA visualizations of the latent vector z for reduced channel CAE inputs. (a) shows the correct labels of the test set, (b) shows the generated labels obtained through K-Means clustering.

from the results of K-Means clustering that the algorithm struggles to separate the classes completely within the latent space. With no bounds to restrict representations, the convolutional autoencoder has mapped images into two distinct, but entangled distributions. As shown in Fig. 5, K-Means has attempted to split latent representations into anomalies (vessels) and sea, but due to the entangled representations, has struggled to separate classes within the common regions. Whilst the clustering ability appears poor in Fig. 5, performances achieved by the algorithm are impressive, view-able in Table 1.

The performance of the two models was assessed using precision and recall, described in Equation 2 - two metrics which are commonly used for assessment of the ML methods performance in cases with imbalanced datasets, such as the works described in section II. T_p and T_n refer to correct classifications, both true and false, of detection. Their counterparts are represented as F_p and F_n . N_{real} refers to the ground truth for vessels present ($N_{real} = T_p + F_n$).

$$Precision = \frac{T_p}{T_n + F_n} \quad Recall = \frac{T_p}{N_{real}} \quad (2)$$

Both precision and recall are important metrics in vessel detection problems, but their weighting of importance can depend on the application. Automated systems, such as routine traffic surveillance, will often favor a high precision over a recall, as missed detections are less important than overwhelming the system with outliers. More sensitive applications, such as military operations or illegal smuggling, will favor recall, as missed detections can prove costly. As a result algorithms are often tuned to their application. The results of the detection phase of the algorithm are shown below in Table 1, alongside the results from the works described in Section II.

Before comparing, the reader is reminded that drawing comparisons between vessel detection algorithms, even ones which utilize similar GSDs, is an extremely difficult task. This issues arises due to the variety witnessed between test sets and input conditions, requirements, and success criteria. Nevertheless, an attempted comparison has been conducted, with these factors considered in an attempt to draw meaningful comparisons.

Table 1 A comparison of results obtained from literature and this report.

Author	Method	GSD (m)	N_{real}	T_p	T_n	F_p	F_n	Precision (%)	Recall (%)
Ding [10]	Saliency	30	-	-	-	-	-	98.95	-
Ghosh [4]	CFAR	42	778	668	110	378	110	64	86
Liu [9]	CFAR	50	885	720	1191	165	165	37.68	81.36
Model	Method	GSD (m)	N_{real}	T_p	T_n	F_p	F_n	Precision (%)	Recall (%)
CAE-3	CAE	42.9	800	357	962	38	443	90.56	44.38
CAE-1	CAE	42.9	800	635	987	13	165	97.99	79.38

When comparing the produced models with Ghosh et al.’s CFAR based approach [4], the merits of the proposed autoencoder models can be seen. Both algorithms use very similar GSDs, and include a similar number of test vessels within images. Both developed deep-learning models are able to vastly outperform the CFAR strategy in terms of precision, but not recall, likely due to the issues discussed with K-Means. Additionally, whilst computational efficiency has been considered throughout the approach in this project, there have been no direct measurements of performance using on-board hardware. Ghosh et al. have implemented their algorithm on small-satellite hardware, which is a merit which should be investigated further with this project.

The single channel (CAE-1) method is able to outperform Liu et al.’s model by some margin. Although it is believed Liu et al. have utilized more challenging test images in their datasets, a factor which was neglected in favor of baseline operation in this paper. As a result, it is difficult to draw conclusions without a shared baseline dataset. The same issue can be said for comparing with Ding et al. [10] who use higher resolution simulated test images, rather than one taken directly from an optical remote sensing satellite. Comparing, the 3-channel RGB method (CAE-3) to the single channel method (CAE-1) in Table 1, it is evident that the CAE-1 outperformed CAE-3. The number of false negatives has been drastically reduced, providing a much more acceptable 79% recall. The model precision has also been increased, reaching nearly 98%.

The findings of this paper align with those of Ding et al. [10], suggesting that for an equivalent model, combining RGB channels into one channel offers a greater performance than operating on them individually. The reasoning for this is likely to be multifaceted. First, the dimensions for the RGB and reduced inputs in the latent space were the same, requiring the same autoencoder architecture to reduce the dimensions of the RGB image three times as much of that of the single channel. Increasing the size of both the latent vector z and convolutional filters for the three-channel model could be investigated in further work to test this; but this would also introduce more nodes into the neural network, reducing computational efficiency and impacting the algorithm’s ability to be hosted on-orbit. Another reason could be that the feature learning performed in the 3-channel scenario was not well constrained, resulting in the autoencoder relying more on contrast and luminance than structural features. This could be addressed by processing each channel separately, an approach that was seen used by other works [21, 26]. Future work in this regard could involve the comparison of other color spaces for vessel detection, such as CIELAB.

A major influence in the performance of the proposed algorithms was the datasets that they were tested and trained on. Despite mitigation attempts which were described earlier to source vessel images of any size, it is believed the test set currently biases toward larger vessels as a result of the manual production of test data. Seen in Fig 6, vessels smaller than the GSD were represented as a luminance change in a single pixel, which provides an extremely difficult detection problem. A single pixel does not provide any shape or texture information, apart from that the object is below the GSD of the image. This factor greatly hampered the results of the discrimination phase, which could only localize vessels when their imprint on an image, either by their size or wake, exceeded 100 m.

Whilst the 3-channel CAE mainly produced false negatives as a result of a vessels size in an image, the single channel CAE produced false negatives due to the target blending into the shape and texture of the images. Figure 6 demonstrates this, where Fig 6b highlights two false negatives of interest. The even to a human operator, identifying a vessel within these images is difficult, as they match both the intensity and texture of the images, likely producing a nominal latent representation due to the SSIM objective function. The selection of test candidates in higher resolution, a technique discussed in section III.A, has resulted in small, almost impossible to see images of vessels being incorporated into the test set.

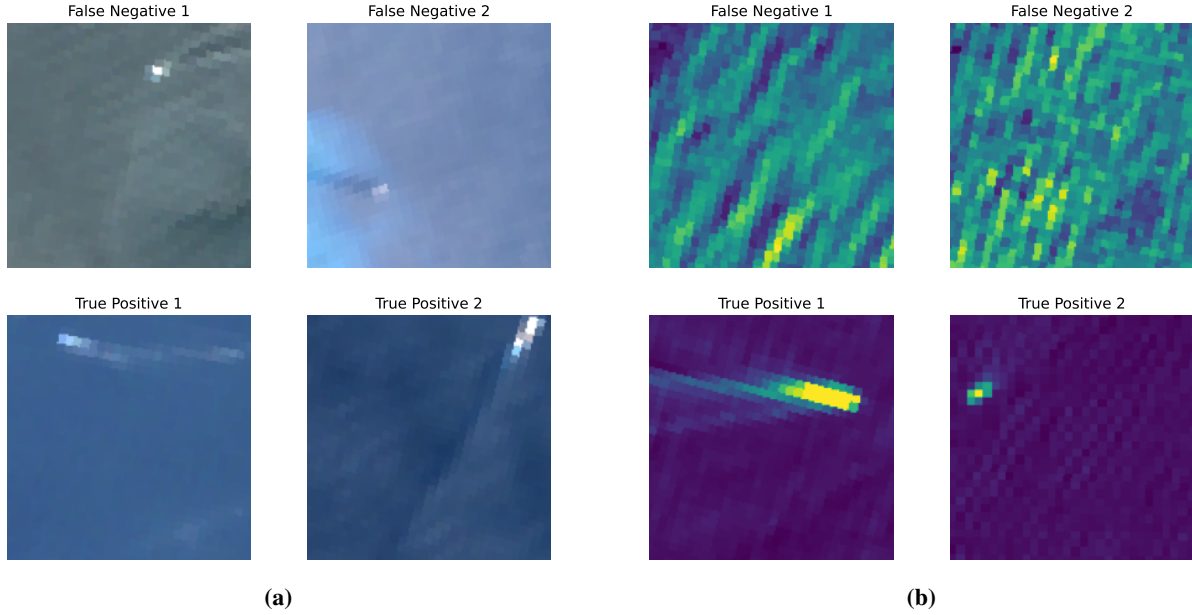


Fig. 6 An overview of selected false negative and true positive samples from both the RGB and single channel autoencoder models. (a) shows samples from the RGB model, (b) shows samples from the single channel model.

V. Conclusion

With the interest of maritime surveillance constantly growing, small nanosatellite systems are being increasingly considered as a suitable means to perform vessel detection. This report consisted of two primary aims. The first was to develop a computationally inexpensive means of performing vessel detection on medium-resolution satellite optical imagery, whilst the second was to utilize the field of machine learning to achieve this.

The developed algorithm was capable of outperforming existing traditional techniques found in literature, however the need to generalize when comparing such different algorithm approaches affected the confidence in this claim. The findings in this paper do generate confidence that machine learning on-orbit techniques can be used to accurately identify vessels from medium resolution optical images, a research problem that had not yet been addressed.

In regards to further work, further categorization could be applied to the test data to witness how vessels of different sizes, shapes, and speeds are represented within the latent space. Further work in regards to candidate detection should be considered, including the use of supervised methods, or a variational autoencoder, which have been shown to provide a disentanglement in their latent space [18]. The self-supervised learning strategy should also be re-attempted, which proved troublesome due to smaller vessels being classified as sea from the discrimination approach.

Based on the conclusions made in this paper, readers should consider this new field of vessel detection as a viable option to consider when developing systems of their own, bearing in mind its benefits in potential spatiotemporal coverage, and limitations with vessels which are smaller than the source image's ground sample distance.

References

- [1] Kanjir, U., Greidanus, H., and Oštir, K., "Vessel detection and classification from spaceborne optical images: A literature survey," *Remote sensing of environment*, Vol. 207, 2018, pp. 1–26.
- [2] Greidanus, H., "Assessing the operability of ship detection from space," *ESA Special Publication*, Vol. 584, 2005, p. 6.
- [3] UCS, "Union of Concerned Scientists (UCS) Satellite Database," *UCSUSA*, 2020. URL <https://www.ucsusa.org/resources/satellite-database>, [Available Online] Accessed 23/06/20.
- [4] Ghosh, S., Konugurthi, P. K., Shankar Rao Singupurapu, G., Patel, S., Tammanagari, T., Rao Desu, M., Thakar, L. K., and Ghara, I., "On-board ship detection for medium resolution optical sensors," *Sensors*, Vol. 21, No. 9, 2021, p. 3062.
- [5] Bannister, N., and Neyland, D., "Maritime domain awareness with commercially accessible electro-optical sensors in space," *International Journal of Remote Sensing*, Vol. 36, No. 1, 2015, pp. 211–243.

- [6] Giuffrida, G., Fanucci, L., Meoni, G., Batič, M., Buckley, L., Dunne, A., Van Dijk, C., Esposito, M., Hefe, J., Vercruyssen, N., Furano, G., Pastena, M., and Aschbacher, J., "The Phi-Sat-1 mission: the first on-board deep neural network demonstrator for satellite earth observation," *IEEE Transactions on Geoscience and Remote Sensing*, 2021, pp. 1–1. <https://doi.org/10.1109/TGRS.2021.3125567>.
- [7] ESA, "Pleiades Mission Data," , 2019. URL <https://earth.esa.int/eogateway/missions/pleiades>, [Available Online] Accessed 19/06/21.
- [8] Liu, Y., Yao, L., Xiong, W., and Zhou, Z., "GF-4 satellite and automatic identification system data fusion for ship tracking," *IEEE Geoscience and Remote Sensing Letters*, Vol. 16, No. 2, 2018, pp. 281–285.
- [9] Liu, Y., Yao, L., Xiong, W., Jing, T., and Zhou, Z., "Ship target tracking based on a low-resolution optical satellite in geostationary orbit," *International Journal of Remote Sensing*, Vol. 39, No. 9, 2018, pp. 2991–3009.
- [10] Ding, Z., Yu, Y., Wang, B., and Zhang, L., "An approach for visual attention based on biquaternion and its application for ship detection in multispectral imagery," *Neurocomputing*, Vol. 76, No. 1, 2012, pp. 9–17.
- [11] Tomljenovic, I., Höfle, B., Tiede, D., and Blaschke, T., "Building extraction from airborne laser scanning data: An analysis of the state of the art," *Remote Sensing*, Vol. 7, No. 4, 2015, pp. 3826–3862.
- [12] Ferreira, N., and Silveira, M., "Ship Detection in SAR Images Using Convolutional Variational Autoencoders," *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2020, pp. 2503–2506.
- [13] Tang, J., Deng, C., Huang, G.-B., and Zhao, B., "Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53, No. 3, 2014, pp. 1174–1185.
- [14] Dolgoplov, A. V., Kazantsev, P. A., Bezuhliy, N. N., Dolgoplov, A., Kazantsev, P., and Bezuhliy, N., "Ship detection in images obtained from the unmanned aerial vehicle (UAV)," *Indian J. Sci. Technol*, Vol. 9, 2017, pp. 1–7.
- [15] Giuffrida, G., Diana, L., de Gioia, F., Benelli, G., Meoni, G., Donati, M., and Fanucci, L., "Cloudscout: a deep neural network for on-board cloud detection on hyperspectral images," *Remote Sensing*, Vol. 12, No. 14, 2020, p. 2205.
- [16] Di Mascio, S., Menicucci, A., Gill, E., Furano, G., and Monteleone, C., "On-board decision making in space with deep neural networks and risc-v vector processors," *Journal of Aerospace Information Systems*, Vol. 18, No. 8, 2021, pp. 553–570.
- [17] Furano, G., Meoni, G., Dunne, A., Moloney, D., Ferlet-Cavrois, V., Tavoularis, A., Byrne, J., Buckley, L., Psarakis, M., Voss, K.-O., et al., "Towards the use of artificial intelligence on the edge in space systems: Challenges and opportunities," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 35, No. 12, 2020, pp. 44–56.
- [18] Murphy, K. P., *Probabilistic Machine Learning: An introduction*, MIT Press, 2022.
- [19] Patel, E., and Kushwaha, D. S., "Clustering cloud workloads: k-means vs gaussian mixture model," *Procedia Computer Science*, Vol. 171, 2020, pp. 158–167.
- [20] Corbane, C., Najman, L., Pecoul, E., Demagistri, L., and Petit, M., "A complete processing chain for ship detection using optical satellite imagery," *International Journal of Remote Sensing*, Vol. 31, No. 22, 2010, pp. 5837–5854.
- [21] Kurekin, A. A., Loveday, B. R., Clements, O., Quartly, G. D., Miller, P. I., Wiafe, G., and Adu Agyekum, K., "Operational monitoring of illegal fishing in Ghana through exploitation of satellite earth observation and AIS data," *Remote Sensing*, Vol. 11, No. 3, 2019, p. 293.
- [22] Song, Z., Sui, H., and Wang, Y., "Automatic ship detection for optical satellite images based on visual attention model and LBP," *2014 IEEE Workshop on Electronics, Computer and Applications*, IEEE, 2014, pp. 722–725.
- [23] Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., and Steger, C., "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," *CoRR*, Vol. abs/1807.02011, 2018. URL <http://arxiv.org/abs/1807.02011>.
- [24] Daniel, B. J., Schaum, A. P., Allman, E. C., Leathers, R. A., and Downes, T. V., "Automatic ship detection from commercial multispectral satellite imagery," *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*, Vol. 8743, International Society for Optics and Photonics, 2013, p. 874312.
- [25] Pedamonti, D., "Comparison of non-linear activation functions for deep neural networks on MNIST classification task," *arXiv preprint arXiv:1804.02763*, 2018.
- [26] Heiselberg, H., "A direct and fast methodology for ship recognition in sentinel-2 multispectral imagery," *Remote Sensing*, Vol. 8, No. 12, 2016, p. 1033.

Self-supervised vessel detection from low resolution satellite imagery

Farr, Alexander J.

2021-12-29

Attribution-NonCommercial 4.0 International

Farr AJ, Petrunin I, Kakareko G, Cappaert J. (2021) Self-supervised vessel detection from low resolution satellite imagery. In: AIAA SciTech 2022 Forum, 3-7 January 2022, San Diego, USA and Virtual Event, Paper number AIAA 2022-2110

<https://doi.org/10.2514/6.2022-2110>

Downloaded from CERES Research Repository, Cranfield University