Identification of Ships in Satellite Images

Peder Heiselberg, Hasse B. Pedersen, Kristian A. Sørensen, and Henning Heiselberg

***Abstract*—Satellite imagery has become a fundamental part for maritime monitoring and safety. Correctly estimating a ship's identity is a vital tool. We present a method based on facial recognition for identifying ships in satellite images. A large ship dataset is constructed from Sentinel-2 multispectral images and annotated by matching to the automatic identification system. Our dataset contains 7000 unique ships, for which a total of 16 000 images are acquired.The method uses a convolutional neural network to extract a feature vector from the ship images and embed it on a hypersphere. Distances between ships can then be calculated via the embedding vectors. The network is trained using a triplet loss function, such that minimum distances are achieved for identical ships and maximum distances to different ships. Comparing a ship image to a reference set of ship images yields a set of distances. Ranking the distances provides a list of the most similar ships. The method correctly identifies a ship on average 60% of the time as the first in the list. Larger ships are easier to identify than small ships, where the image resolution is a limitation.**

***Index Terms*—Automatic identification system (AIS), convolutional neural network (CNN), dark ships, multispectral images, satellite images, ship identification, triplet.**

I. Introduction

LARGER ships are by law required to identify themselves by transmitting automatic identification system (AIS) messages. AIS contains a timestamp, position, class, size, and identity in the form of a unique maritime mobile service identity (MMSI) number. However, the AIS transponder can be disabled and signals lost, or accidentally turned OFF. Without continuous transmission, the vessel goes "dark." Dark ships can be in distress or engaged in illicit activities, such as illegal fishing, smuggling, oil spills, trespassing, or piracy. Spoofing by modifying the transmitted MMSI or position as misinformation is also an increasing problem.

Dark ships are visible in satellite images. Satellite imagery has therefore become a valuable asset for increasing maritime domain awareness. Synthetic aperture radar (SAR) sensors are often preferred since they can operate both day and night and penetrate cloud covers. Multispectral images offer both visible and near-infrared images often with higher resolution, which provide clearer differentiation between vessels based on their shapes, colors, and structural characteristics. A multispectral image of a ship may be seen as a fingerprint of the ship [1].

In the literature, detection, classification, identification, and recognition are often used in a wide sense. For the purpose of this study, the following definitions are used: Detection aims at localizing a ship and discriminating it from other objects, such as offshore structures. Classification provides the detected vessel's type, e.g., *container*, *tanker*, *fishing*, or *military*. Identification establishes the unique identity of the ship, such as its MMSI number. In this study, we propose a methodology that can identify a ship from its Sentinel-2 multispectral satellite image.

Detection and classification in optical satellite images has been studied in numerous works [2], [3]. The AIS ship type is entered manually into the transmitter. This gives rise to a large within-class variation, which make classification more complicated and challenging [4]. Moreover, each defined class may have several subtypes and different sizes. Heiselberg and Heiselberg [5] used a principal component analysis to reduce the dimensions of ships, icebergs, wakes, etc. Each class formed a cluster, which allowed classification by distance in the reduced space dimension.

In recent years, deep neural networks have been used to extract such a reduced dimension feature vector embedding from images. The similarity between two images then directly corresponds to the distance between embedding vectors. There is extensive literature on contrastive learning, face recognition, and person reidentification [6], [7], [8]. We suggest applying these methods to ships by comparing an input ship image against a reference set of ship images (Fig. 1 contains examples of ship images). In this sense, the objective is to rank the input image to the reference set by similarity. The highest similarity is then ideally obtained between identical ships.

Ship identification has, to the best of the authors' knowledge, not been attempted before. It was proposed in [1] to build a spectral library of Sentinel-2 multispectral ship signatures for this purpose. This was met with skepticism [2] due to the large number of ships worldwide. Yet, a similar argument could be made against face recognition. In this article, we present the first study on ship identification in satellite images.

II. Data

The two Sentinel-2 satellites record multispectral optical imagery in the visible, near infrared, and short-wave infrared spectra. Moreover, the large catalog spanning back to 2015 is freely available. The analysis focused on the Level-1 C (L1C) scenes in all 13 spectral bands, that have resolution varying from 10 to 60 m.

Manuscript received 11 October 2023; revised 11 January 2024; accepted 13 February 2024. Date of publication 22 February 2024; date of current version 13 March 2024. *(Corresponding author: Peder Heiselberg.)*

Peder Heiselberg is with the Department for Geodesy and Earth Observation, National Space Institute, Technical University of Denmark, 2800 Kongens Lyngby, Denmark (e-mail: ph@space.dtu.dk).

Hasse B. Pedersen, Kristian A. Sørensen, and Henning Heiselberg are with the Center for Security, National Space Institute of Denmark, Technical University of Denmark, 2800 Kongens Lyngby, Denmark (e-mail: hbype@space.dtu.dk; kaaso@space.dtu.dk; hh@dtu.dk).

Digital Object Identifier 10.1109/JSTARS.2024.3368508

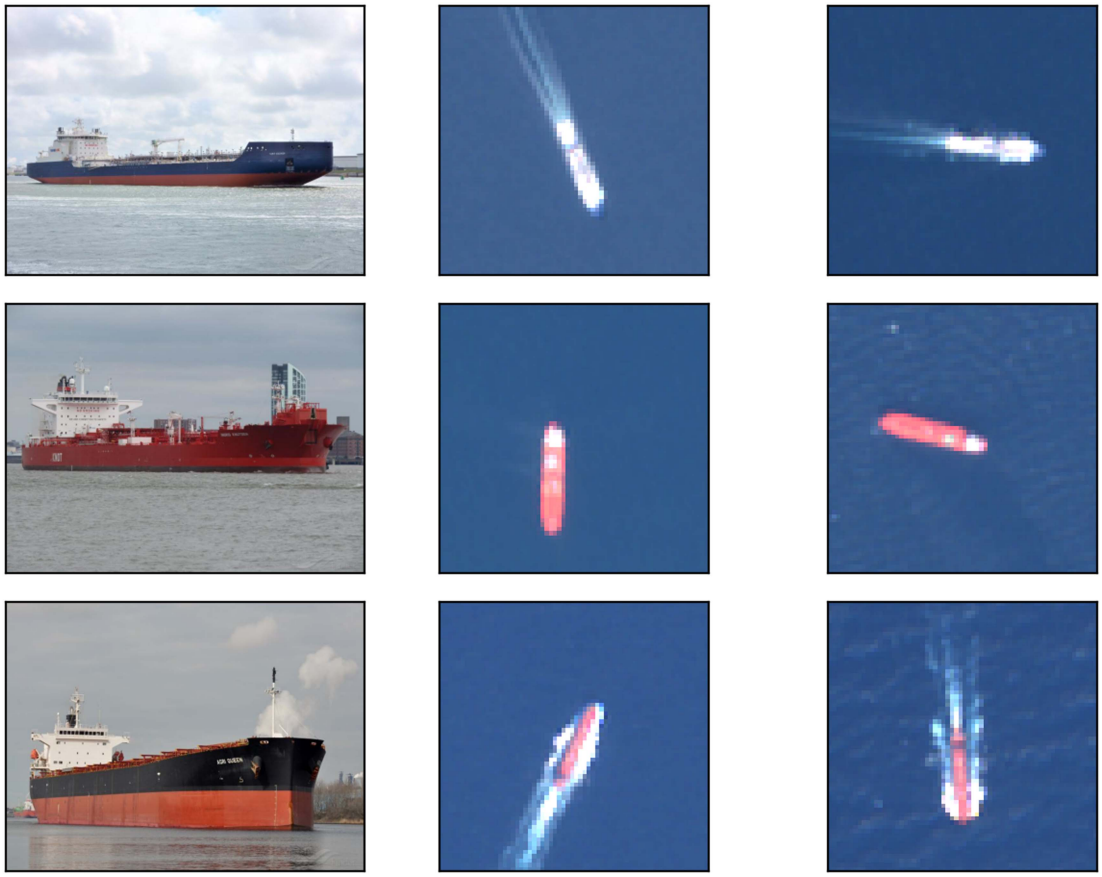

Fig. 1. Three different ships with two Sentinel-2 satellite images, each from a different time and place.

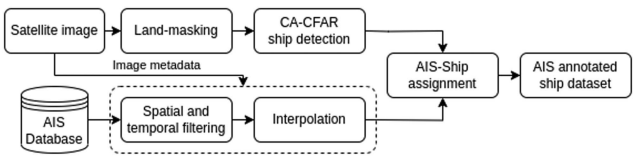

Fig. 2. Illustration of the annotation process.

Three areas were chosen with high maritime traffic: The seas around Denmark, the English Channel, and the Strait of Gibraltar. In total, 25 000 scenes were acquired from the period of 2015 to 2022. However, most scenes were discarded due to cloudy conditions or few ships.

An automated process of ship detection and AIS correlation was employed (see Fig. 2). AIS signals were acquired in a two-hour window for each scene. The AIS signals were then aggregated based on the MMSI number and interpolated to the recording time of the scene. A cell averaged constant false alarm rate [9] detection algorithm was applied to the high resolution near-infrared band to find ships in the Sentinel-2 scenes. The interpolated AIS tracks were then assigned to a detected ship if possible. Through this algorithm, a list of AIS annotated ships were acquired. A 64×64 pixel image in 13 multispectral bands were cropped and saved around each ship. Repeating this process for all Sentinel-2 scenes yielded a total of 7000 unique ships. Each ship was on average found a couple of times, yielding 16 000 AIS annotated ship images.[1] For a more detailed description of the detection and annotation process, we refer to [10] and [11]. Six examples of annotated ship images are presented in Fig. 1.

III. Methodology

To correctly identify a ship based on an image, features that distinguish the ship from others must be extracted. The ship type, length, and width could be considered such features. However, the same ship can use different types and different ships can use the same type. Many ships are also of the same size or built to the same specifications. It is, thus, necessary to use more features. Extracting features was traditionally done using handcrafted local image descriptors, e.g., [12]. Modern methods exclusively leverage deep learning for feature extraction, capitalizing on the advantages offered [13].

The method employed in this work follows recent techniques developed for face recognition. In [6], a unified embedding was constructed from face images and used to train a convolutional neural network (CNN) for face verification, recognition, and clustering. Analogously, we use a CNN to extract an embedding vector representation of the ship to calculate similarities

[1]Data is available online at https://doi.org/10.5281/zenodo.10473184

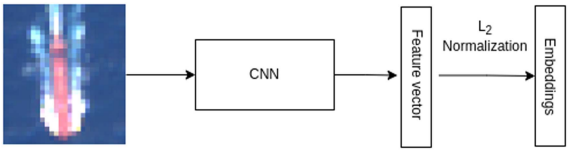

Fig. 3. Illustration of the embedding process for a single ship image.

between ships. Fig. 3 shows the embedding processes for a ship image.

A. Ship Embeddings

The function $f(x) \rightarrow R^d$ embeds the multispectral ship image x into a d-dimensional feature vector. L_2 normalization was subsequently applied to the feature vector, constraining it to a d-dimensional hypersphere, $||f(x)||_2 = 1$. The similarity between two images was then computed as the Euclidean distance between the L_2 normalized embedding vectors

$$D_{x,y} = ||f(x) - f(y)||_2. \tag{1}$$

The function f was learned by a CNN using a triplet loss. A distance matrix between ship images is presented in Fig. 4.

B. Triplet Loss

A triplet contains an anchor (a), a positive (p) and a negative (n) ship image, i.e., two images of the same ship (a, p), and one image of another ship (n). In Fig. 1, positives can be seen on the horizontal axis, and negatives on the vertical axis. This gives the similarities $D_{a,p}$ and (dissimilarity) $D_{a,n}$. Both range between [0, 2]. The loss for a triplet was chosen as in [7]; a sum over triplet similarities

$$\mathcal{L}_3 = \sum_{i \in T} [m + D_{a_i,p_i} - D_{a_i,n_i}]_+ \tag{2}$$

where $[z]_+ = \max(z, 0)$ is the hinge loss and m the margin parameter. T was a carefully chosen set of triplets, as will be discussed in Section III-C. $\mathcal{L}_3$ is minimal when positive ship images are projected closer to the anchor image than the margin, m and vice versa for the negative ship images. If optimized over the entire data set for a sufficient duration, all positive pairs will eventually be pulled together and negative pairs pushed away, referred to as pull–push. The loss rewards forming separate ship clusters with radius m, separated by distances exceeding m on the hypersphere. Each cluster should contain ships with the same identity.

C. Triplet Mining

The number of possible triplets grows rapidly with the size of the dataset. It is, thus, advantageous to mine difficult triplets. Triplet mining can be done offline or online during training. Offline mining is done in sequence to training. After a pass through the dataset, similarities can be calculated and subsequently used to sample the hard images. Online mining is performed during training within the sampled batch of images. The former is associated with a significant increase in training time, thus, we opted for online mining. Even within a batch of P unique ships with K images of each ship, the number of possible triplets is large. Learning to separate, e.g., a large red from a small white ship is trivial, rendering many triplet combinations unnecessary. Considering all possible triplets within a batch can, thus, be considered "easy" as it includes many easily separated examples.

D. Selecting Hard Triplets

A batch of P unique ships were randomly sampled. For each ship, K images of the same ship were then randomly sampled. In cases where there are fewer than K images of the same ship, we sampled with replacement, e.g., we used augmentation to replicate images. The CNN then computed an embedding for each image, and the similarity was calculated between all ship embeddings. Thus, each ship image has $(K-1)$ positive and $(P-1) \cdot K$ negative pairs. However, these contain numerous easy triplets. The hardest triplets were then selected within the batch by taking the positive pair with the lowest similarity and negative pair with the highest similarity. This resulted in a subset of hard triplets for which the loss [see (2)] was calculated. We consider the ships in Fig. 4 to make up semihard triplets. They are all similar size and, thus, require additional information to discriminate. In Fig. 5, we present hard examples. These ships are both the same size and color.

E. Training

The dataset was divided into two parts, for training and for testing, with an 80% split. The division was made randomly based on the identity, such that a ship occurring in the training set did not appear in the test set. This meant that the CNN had not seen any of the ships in the test dataset.

As baseline, we used the ResNet18 [14] CNN. It has a well-known architecture and allowed for relatively fast training due to its size. Any neural network is capable of embedding multispectral images. A few other CNNs were also tested in Section III-G. Our default settings were $m = 0.2$, $K = 2$, and $d = 128$. d was the number of outputs from the CNN. In Section III-G different settings are experimented with. We used the AdamW optimizer with default parameters [15] and a batch size of $P = 64$ ships. During training and testing, we augmented the images by random 90° rotations and flipping. In total, 20% of the training dataset was further set aside for validation. After each epoch (pass through the training dataset), the loss was calculated for the validation dataset. The model corresponding to the minimum validation loss was saved and used for evaluation. We allowed 50 epochs without improvement of the validation loss before terminating the training. Training a single model usually took about 200 epochs, corresponding to 24 h, using an Intel Core i7-10710 U CPU with base frequency of 1.10 231 GHz. The training flow was implemented using the python autograd library PyTorch [16].

F. Evaluation

We calculated the embeddings for all ship images in the test dataset. The embeddings were averaged over five random

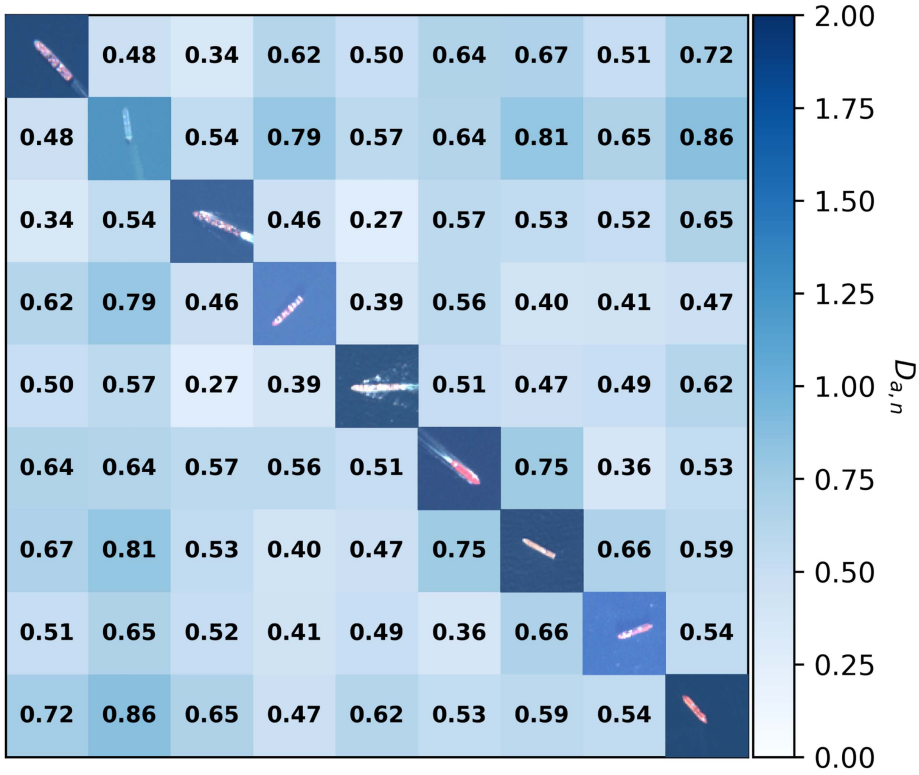

Fig. 4. Distances between images of different ships with similar size.

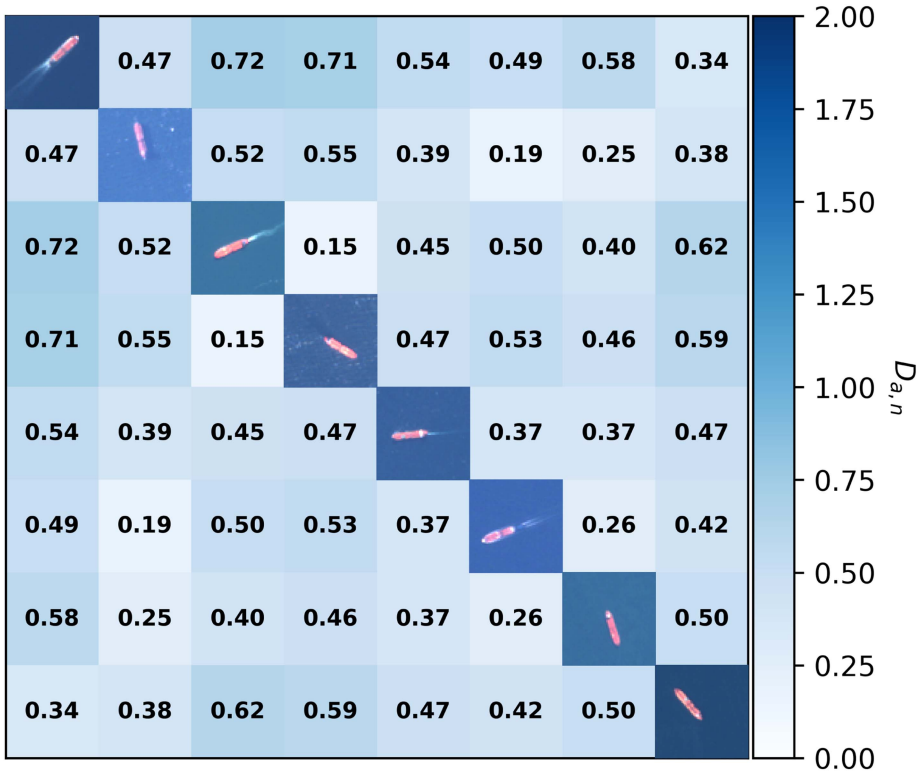

Fig. 5. Distances between images of different ships that are very similar (hard examples). The bottom most is also the bottom most of Fig. 4.

TABLE I
ACCURACIES FOR DIFFERENT MINING STRATEGIES AND LOSS FUNCTIONS

	Hard	Easy	NT-Xent
mAP	**38.0**	32.9	38.5
Rank 1	**51.1**	44.4	50.5

Metrics presented in percentages. Bold face numbers refer to the baseline model.

TABLE II
ACCURACIES FOR DIFFERENT CNN MODELS (SEE TABLE I FOR TABLE EXPLANATION)

Model	ResNet18	ResNet50	ResNet101
mAP	**38.0**	40.4	34.3
Rank1	**51.1**	52.5	47.3

Bold face numbers refer to the baseline model.

augmentations of the image. A trick that also improved the test accuracies of [7]. For a given ship image, we computed the similarity to all other ship images in the test dataset (without replacement). This yielded a query sorted by similarity, which could contain many instances of the same ship (positives) and many negatives.

The Rank1 metric is the proportion of queries where the highest similarity is obtained for an image of the same ship, e.g., the first item in the sorted query was positive. Likewise, Rank$N$ is the proportion of queries where a positive is within the top N most similar ships, etc. By definition, Rank$N$ increase with N.

The Rank$N$ metric does not account for queries where two or more images of the same ships are found in the ranked list. We, therefore, also calculated the mean average precision (mAP) [17], which has become a standard measure in recent years. The average precision (AP) is defined as the area under the precision versus recall curve for a query. For a sorted query, it is

$$\text{AP} = \frac{1}{K}\sum_{k=1}^{K}\frac{k}{N_k}\,. \tag{3}$$

Here K was the number of positives in the query, which are ranked as numbers $N_k, k = 1, .., K$ in the ranked list. For example, a query with $K = 3$ positives found in the ranking list at positions 2, 4, and 6, would result in an AP $= 1/3(1/2 + 2/4 + 3/6) = 0.5$. Likewise, AP $= 0.5$ if they are ranked as 1, 4, and at infinity. Finally, mAP is the mean of AP over all queries. In most cases mAP $\leq$ Rank1 [17] as also found in Section III-G.

G. Experiments

Several experiments were performed. We divided these into ablation and comparative. In the ablation experiments, we investigated, the use of a different batching strategy, loss function, CNN architecture, and altering the input. This is followed by comparative experiments varying the parameters $d, m,$ and K. The baseline (bold face numbers in Tables I, II, IV–VI) used the settings described in Section III-E and the triplet loss with hard batching. We varied only one setting upon the baseline per experiment, e.g., changing the embedding size d as in Table V.

TABLE III
EFFECT OF DATA REDUCTION ON ACCURACIES (SEE TABLE I FOR TABLE EXPLANATION)

Reduction	None	Uint8	RGB	RBGN
mAP	**38.0**	32.3	30.2	36.1
Rank1	**51.1**	45.4	44.6	50.1

Bold face numbers refer to the baseline model.

TABLE IV
ACCURACIES FOR DIFFERENT CHOICES OF POSITIVE SHIPS K (SEE TABLE I FOR TABLE EXPLANATION)

K	2	4	8
mAp	**38.0**	38.8	36.9
Rank1	**51.1**	51.7	49.6

Bold face numbers refer to the baseline model.

TABLE V
ACCURACIES FOR DIFFERENT CHOICES OF FEATURE VECTOR HYPERSPHERE DIMENSION, d (SEE TABLE I FOR TABLE EXPLANATION)

d	2	16	32	64	128	256	512
mAP	1.5	38.7	36.9	38.3	**38.0**	37.2	37.1
Rank1	2.9	50.0	49.0	51.5	**51.1**	49.8	49.8

Bold face numbers refer to the baseline model.

TABLE VI
ACCURACIES FOR DIFFERENT CHOICES OF CLUSTER MARGIN, m (SEE TABLE I FOR TABLE EXPLANATION)

m	0.01	0.05	0.1	0.2	0.3	0.4	0.5	0.6
mAP	38.3	37.5	37.6	**38.0**	35.4	36.1	34.5	36.7
Rank1	49.6	50.0	50.0	**51.1**	48.7	48.9	47.8	49.9

Bold face numbers refer to the baseline model.

1) Ablation: There are different approaches and loss function for learning suitable embeddings besides the hard batch strategy described above. Instead of only the hardest triplets, computing the loss over all possible triplets in the batch was mentioned in Section III-C. This included easy triplets, hence, we reference it as "Easy." When $K = 2$, the normalized temperature-scaled cross entropy (NT-Xent) can be calculated as [8]

$$\mathcal{L}_{\text{NT-Xent}} = -\log\frac{\exp(\text{sim}_{a,p}/\tau)}{\sum_{i=1}^{2(P-1)}\exp(\text{sim}_{a,i}/\tau)} \tag{4}$$

where $\text{sim}_{a,x} = 1 - D^2_{a,x}/2$, was the cosine similarity between embeddings and the temperature scale was chosen as $\tau = 0.1$. This loss is similar to applying the softmax function and computing negative log likelihood for classification neural networks. The results are shown in Table I and compared with the hard batching. The two strategies performed approximately equally well, but better than easy batching. A similar case was observed in Table II when using larger CNN's.

The Sentinel-2 satellite images have 13 bands in 16-bit unsigned integer format, which is a large amount of data to store and process. We, therefore, investigated the effectiveness of the methodology on reduced data. Three types of data reductions were implemented and tested during training and evaluation, and the results are listed in Table III. Uint8 converts the images

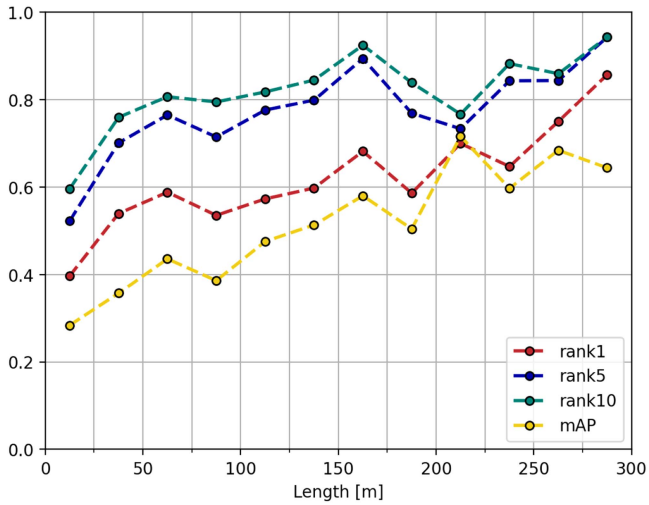

Fig. 6. mAP and rank metrics by ship length. Bins are given by vertical lines.

from 16 to 8-bit. RGB use only the red, green, and blue channels (spectral bands 2, 3, and 4). RGBN also includes near-infrared channel (band 8), which are the four high resolution bands with a pixel size of 10 m. The first two led to a significant reduction in both mAP and Rank1, whereas RGBN did not. This indicates that the four high resolution images are most important and necessary. The lower resolution images do not improve ship identification significantly. Furthermore, the 16-bit dynamic range contains import information on the ship reflection intensities.

2) Comparative: Varying the number of positive images K was done by selecting multiples of K such that $P \cdot K = 128$ remained unchanged within a batch. Increasing the number of positive images from 2 to 4 increased the performance slightly (see Table IV). Yet, further increasing did not lead to an increase in performance.

We also investigated the number of embeddings d and the choice of margin m. The proposed methodology was robust in regard to both parameters. When d were chosen sufficiently large and m adequately small (see Tables V and VI) no significant improvement or loss was achieved.

IV. Results

In this section, we evaluate the baseline approach described in Section III-G in depth. We break down the evaluation by ship sizes and show examples of distances (1) between ships. Furthermore, an analysis of the spectral bands was carried out.

A. Size Dependence

The most significant factor in identification was the ship size. Ships are easily separated by size. Consequently, large ships were easily identified among smaller ships and vice versa. The ship dimensions were, thus, encoded in the embeddings. Fig. 6 shows the mAP and Rank metrics by ship length. The entire test dataset was still used as reference set. Small ships were expectedly more difficult to identify. With an average ship aspect ratio of 7, a 140 m long ship is only two pixels wide in the 10 m resolution images. Despite this, the proposed method queries the correct ship about 60% of the time, or on average within the 2-3 most similar images. For larger ships, the mAP and Rank1 approach 70% and 80% respectively, and higher rankings even better.

B. Ship Similarities

In Figs. 4, 5, and 7 we show the distance matrices between ships. The distance was calculated as in (1) between the ship embeddings. On the diagonal, the respective ship is shown. The relative distance from one ship to another is a measure of the similarity between the ships. Fig. 4 shows the distance matrix for nine different ships with similar size. The respective ship for the row/column was placed on the diagonal. Not surprisingly, ships that looked similar to the human eye also had higher similarity. The examples of Fig. 4 can be considered of "medium" difficulty. For the bottom most ship of Fig. 4 we also provide a comparison to "hard" and identical ships. Fig. 5 shows the distances between eight very similar ships with the same size. Fig. 7 shows the distances between six images of the same ship. The bottom most ship of Fig. 4 was placed on the bottom row of Fig. 5 and used as reference in Fig. 7. Despite the hard examples of Fig. 5, the algorithm can still differentiate the ships. The distances for the ships with the same identity (Fig. 7) are smaller than the hard ships (Fig. 5).

C. Ship Types

The AIS provided ship type is often manually input. Our dataset contained about 100 different types. This results in a large span of ship classes. The classes often have within class overlap, e.g., *tanker*, *chemical tanker*, and *bitumen tanker*. Identical classes, *fishing ship*, *fishing vessel*, and *fishing boat* may be referenced differently, and spelling mistakes are common. It is, therefore, difficult to classify ships using the AIS type. Therefore, classes are often selected beforehand and the different types assigned. The embeddings, calculated by this method, can provide an insight into the underlying ship classes, as shown in Fig. 8. It can be seen that ships are naturally clustered, not only by size and color. Fig. 9 shows three input ships annotated with green ground truth AIS type and the two most similar ships of the reference set. The embeddings, by comparison, allow rare types such as *refrigerated cargo ship* to be expressed. Despite the middle example retrieving a different type, it is still a meaningful one. This introduces the use of the embeddings for ship classification in future work.

V. Discussion

The size of a ship plays an important role in identification (see Fig. 6). This is a natural consequence of the satellite image resolution. The Sentinel-2 images used in this study have at most 10 m pixel resolution. Consequently, the identification measures are, on average, mAP $= 40\%$ and Rank1 $= 60\%$. For large ships longer than 200 m it is higher, but lower for small ships shorter than 50 m. Smaller ships, the size of only a few pixels, were

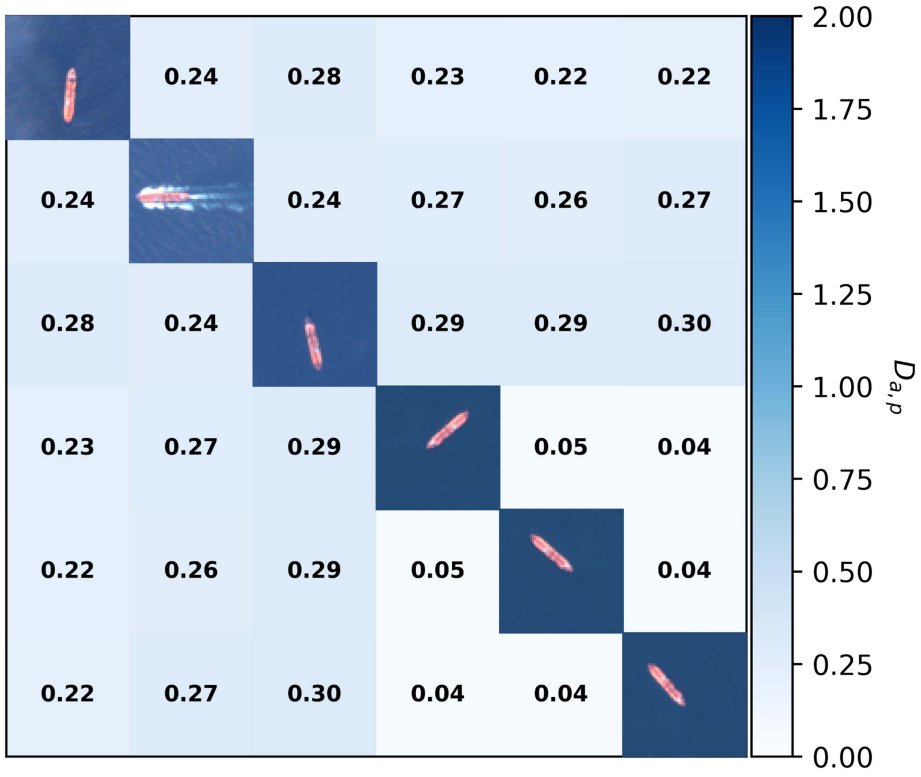

Fig. 7. Distances between different images of the same ship. The ship is the bottom most ship of Figs. 4 and 5.

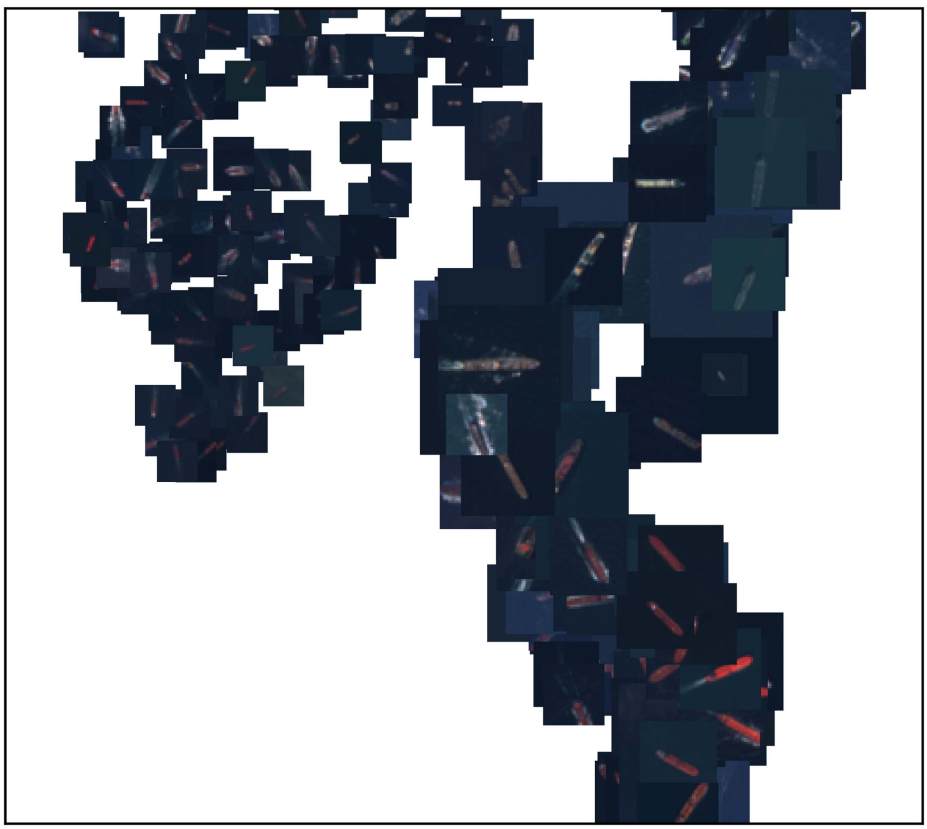

Fig. 8. Small section of the Barnes-Hut t-SNE [18] of the learned embeddings. The images were cropped proportional to ship length.

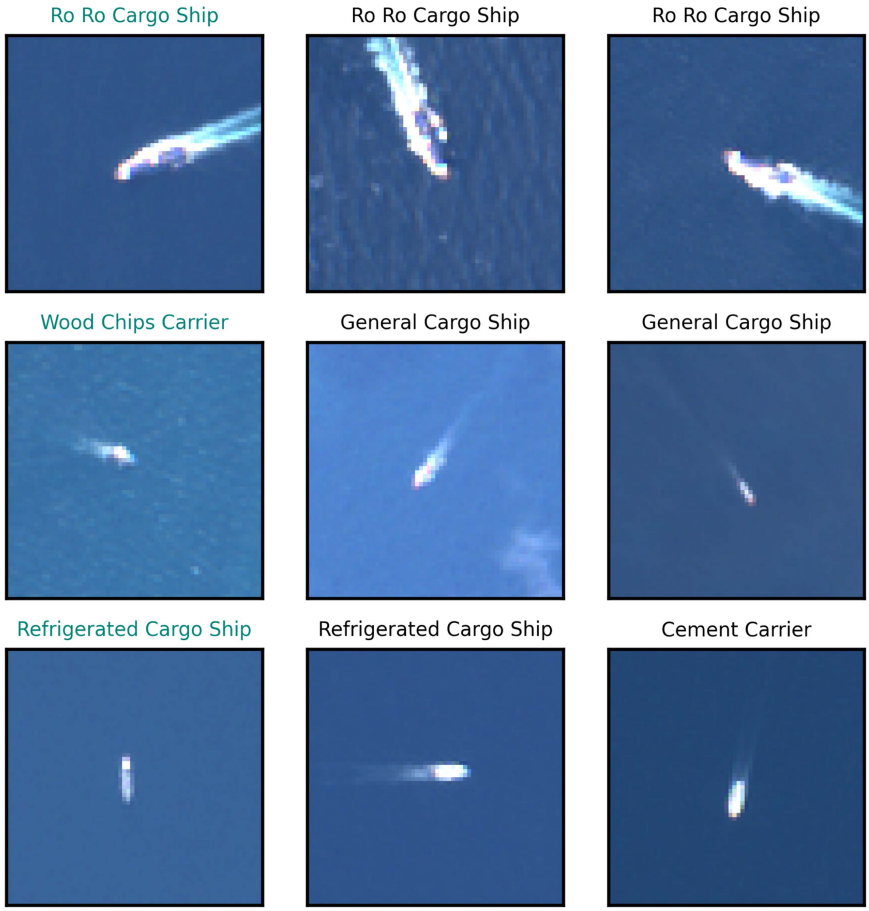

Fig. 9. Example of classification using similarities for three ships. The first column contains the input ship and ground truth AIS type. The following two images are the most similar ships in the reference set.

expectedly harder to identify. This indicates that satellite images with higher spatial resolution will improve identification. Higher order rankings also yielded significant better identification, e.g., Rank5 $\simeq 80\%$. It can in part be due to the size of the dataset. The method has a harder time differentiating between very similar ships (see Fig. 5). For some ships, there may only be a few very similar ships in the dataset. All are, therefore, easily included in the top rankings.

Newly constructed ships of the same model are often identical (twin ships). These are likely impossible to tell apart. In Fig. 5, we showed the possibility of identification even for very similar ships. Yet, extending the gallery will lead to a reduction in precision, as also found in [7].

In Section III-G we evaluated different hyperparameter configurations and CNNs. Some settings led to small improvements only. However, the effect of all parameters, including the size of the CNN, also depends on the dataset. Different configurations may be more or less optimal with a larger dataset. Our proposed methodology is robust to selection of hyperparameters. Tables V and VI show little change as long as m is chosen small and d large. In contrast, we noticed a sharp drop in accuracy of the NT-Xent algorithm for $\tau = 1$ during initial experiments. However, NT-Xent has a familiar formulation known from classification. Moreover, the methodology may also benefit from larger batch sizes and more aggressive augmentation [8].

The four high resolution Sentinel-2 MSI bands with 10 m pixel resolution are the most important for identification (see Table III). Disregarding the lower resolution bands did not result in a significant decline in accuracy. The near-infrared band in particular is important to include alongside the red, green, blue bands. It is also commonly used for ship detection [1].

One of the main obstacles in conducting this study was creating a balanced dataset. Multiple images of the same ship are easily acquired for smaller fishing vessels and ferries, which go back and forth between local harbors. On the other hand, tankers, dry-bulk, and container ships are rarely imaged twice. These ships can sail globally, whereas our study focused on predetermined regions. Moreover, varying cloud cover further limits the probability of observation. Consequently, we had to include thousands of Sentinel-2 scenes to get a big and diverse enough dataset. Automation was necessary to ensure enough training data for fast scalability and transferability, limited only by storage. AIS was used to automatically label the detected ships. However, AIS can be erroneous and, while false alarms were kept to a minimum by the annotation procedure, few may exist in the dataset. A poorly imaged ship may be ranked low in the query, leading to a skew of the mAP.

The dataset was split based on ship identity. Thus, no ship in the test dataset had the same identity as a ship in the training dataset. In an operational scenario, this would not be the case.

We expect the precision of the methodology would be higher in this scenario. However, this dataset split allowed for an unbiased benchmark. If the ship images were split random, the accuracy for a ship would then be dependent on the number of training images of that ship. There would then be an optimal split for each ship. The precision for a ship would then be dependent on how its images were split.

VI. Conclusion

Monitoring the oceans with satellites is an important part of increasing maritime safety and reducing negative environmental impacts. Dark ships that do not transmit an AIS signal can be in distress or engage in illicit activities. Detection of dark ships is an area of active research. Yet, due to the size of the oceans and limited naval patrolling, detection is not always enough. Correct identification of dark ships provides the next level of monitoring. It also allows detection of spoofing, e.g., ships transmitting the wrong MMSI.

In this study, we proposed a methodology for identifying ships in satellite imagery. The ships are correctly identified, about 60% of the time and up to 80% for large ships. The very high probability of on average $\sim 80\%$ for Rank5 and higher demonstrates that our proposed ship identification method is promising. Correct identification of ships is an important measure to counteract illegal activities.

This is the first study of ship identification from satellite images. We have identified several ways of further improving ship identification. Fig. 8 shows that the embeddings can be separated by the ship size. During training, it would therefore be beneficial to introduce a batching strategy based on the ship size. It can be implemented by first sampling a single random ship instead of P. Then, sampling the remaining $P-1$ with a probability based on the size of the initial ship. The entire batch would thus be composed of harder examples.

Reducing the number of possible candidates is key to improve identification. This can be achieved by limiting the reference set to smaller regions, e.g., the North Sea, Baltic Sea, Arctic, Black Sea, etc. as also suggested in [2]. It will also counteract the reduction in precision when the gallery is extended [7].

Correctly identifying a ship may also be improved using information about its prior whereabouts. It introduces a velocity and spatial component, e.g., the ship's range. The spatio-temporal information and memory could be encoded into a model to accomplish this.

Identification is not limited to optical satellite images. SAR satellites have traditionally been used for maritime monitoring. The methodology is not specific to optical images. It could also be applied for ships in SAR satellite images, utilizing the extensive SAR knowledge available.

Maritime monitoring using satellites is a growing field in the scientific community. Numerous satellites are launched and ship datasets published. AIS data is also increasingly made available. For instance, the European Space Agency is set to launch the third satellite in the Sentinel-1 constellation, with an AIS receiver.

We hope this article spurs further interest in the field of ship identification from satellites.

Acknowledgment

The authors would like to thank the Danish Defense for support, in particular, the Joint Arctic Command (JAC) and the Danish Defense Acquisition and Logistics Organization (DALO).

References

[1] H. Heiselberg, "A direct and fast methodology for ship recognition in sentinel-2 multispectral imagery," *Remote Sens.*, vol. 8, no. 12, 2016, Art. no. 1033.
[2] U. Kanjir, H. Greidanus, and K. Oštir, "Vessel detection and classification from spaceborne optical images: A literature survey," *Remote Sens. Environ.*, vol. 207, pp. 1–26, 2018.
[3] L. Bo, X. Xiaoyang, W. Xingxing, and T. Wenting, "Ship detection and classification from optical remote sensing images: A survey," *Chin. J. Aeronaut.*, vol. 34, no. 3, pp. 145–163, 2021.
[4] L. Huang, W. Li, C. Chen, F. Zhang, and H. Lang, "Multiple features learning for ship classification in optical imagery," *Multimedia Tools Appl.*, vol. 77, pp. 13 363–13 389, 2018.
[5] P. Heiselberg and H. Heiselberg, "Ship-iceberg discrimination in sentinel-2 multispectral imagery by supervised classification," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1156.
[6] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 815–823.
[7] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, *arXiv:1703.07737*.
[8] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.
[9] M. A. Richards et al., *Fundamentals of Radar Signal Processing*. New York, USA: McGraw-hill, 2005, vol. 1.
[10] P. Heiselberg, K. A. Sørensen, H. Heiselberg, and O. B. Andersen, "SAR ship–iceberg discrimination in arctic conditions using deep learning," *Remote Sens.*, vol. 14, no. 9, 2022, Art. no. 2236.
[11] P. Heiselberg, K. Sørensen, and H. Heiselberg, "Ship velocity estimation in SAR images using multitask deep learning," *Remote Sens. Environ.*, vol. 288, 2023, Art. no. 113492.
[12] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, vol. 2, pp. 1150–1157.
[13] O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf. Brit. Mach. Vis. Assoc.*, 2015.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[15] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
[16] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 8026–8037.
[17] C. D. Manning, *An Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[18] L. V. D. Maaten, "Accelerating t-Sne using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, 2014.

Peder Heiselberg received the bachelor's degree in physics and the master's degree in geophysics from the University of Copenhagen, Copenhagen, Denmark, in 2014 and 2017, respectively. He is currently working toward the Ph.D. degree in space based maritime monitoring using deep learning with the Technical University of Denmark, Kongens Lyngby, Denmark.

His research interests include machine learning, ship detection in synthetic aperture radar (SAR) and optical images, and AIS datafusion.

Hasse B. Pedersen received the bachelor's degree in earth and space physics and engineering in 2022 from the Technical University of Denmark, Kongens Lyngby, Denmark, where he is currently working toward the master's degree in earth and space physics and engineering.

Kristian A. Sørensen received the bachelor's and master's degrees in earth and space physics and engineering in 2018 and 2021, respectively, from the Technical University of Denmark, Kongens Lyngby, Denmark, where he is currently working toward the Ph.D. degree in maritime monitoring from space using deep learning.

His research interest focuses on ship trajectory prediction using deep learning.

Henning Heiselberg received the B.S., M.S., and Ph.D. degrees in physics from the University of Aarhus, Aarhus, Denmark, in 1982, 1984, and 1987, respectively.

He is currently the Head with Center for Security, National Space Institute, Technical University of Denmark, Kongens Lyngby, Denmark. He has authored numerous journal papers and scientific books. His current research interests include remote sensing, Arctic and space surveillance, and artificial intelligence.