Proceedings of SPIE, Volume 11543, 20 September 2020, Artificial Intelligence and Machine Learning in Defense Applications II; Article number 1154309
https://doi.org/10.1117/12.2573774

Deep Learning for Automatic Target Recognition with Real and Synthetic Infrared Maritime Imagery

Samuel T. Westlake[a], Timothy N. Volonakis[b], James Jackman[c], David B. James[a], and Andy Sherriff[b]

[a]Centre for Electronic Warfare Information and Cyber, Cranfield University, Defence Academy of the United Kingdom, Shrivenham, SN6 8LA
[b]MBDA UK Ltd, PO Box 5, Filton, Bristol, BS34 7QW
[c]University of Oxford, Institute of Biomedical Engineering, Old Road Campus, Oxford, OX3 7DQ

ABSTRACT

Supervised deep learning algorithms are re-defining the state-of-the-art for object detection and classification. However, training these algorithms requires extensive datasets that are typically expensive and time-consuming to collect. In the field of defence and security, this can become impractical when data is of a sensitive nature, such as infrared imagery of military vessels. Consequently, algorithm development and training are often conducted in synthetic environments, but this brings into question the generalisability of the solution to real world data.

In this paper we investigate training deep learning algorithms for infrared automatic target recognition without using real-world infrared data. A large synthetic dataset of infrared images of maritime vessels in the long wave infrared waveband was generated using target-missile engagement simulation software and ten high-fidelity computer-aided design models. Multiple approaches to training a YOLOv3 architecture were explored and subsequently evaluated using a video sequence of real-world infrared data. Experiments demonstrated that supplementing the training data with a small sample of semi-labelled pseudo-IR imagery caused a marked improvement in performance. Despite the absence of real infrared training data, high average precision and recall scores of 99% and 93% respectively were achieved on our real-world test data. To further the development and benchmarking of automatic target recognition algorithms this paper also contributes our dataset of photo-realistic synthetic infrared images.

Keywords: Automatic target recognition, deep learning, infrared, anti-ship, synthetic, maritime, dataset

1. INTRODUCTION

Most infrared (IR) anti-ship automatic target recognition (ATR) algorithms are designed around classical computer vision concepts, such as adaptive thresholding and hand-crafted feature extraction.[1–6] However, in many comparable domains, the state-of-the-art has been redefined by deep convolutional neural network (DCNN)-based algorithms.[7–11] The application of these algorithms to anti-ship ATR has the potential to improve robustness in adverse conditions and yield significant gains in target recognition and identification performance. These developments are crucial to emerging systems, facilitating the automatic prioritisation of high value targets and contributing neutral and friendly shipping avoidance capabilities.[12]

DCNNs are heavily reliant on increasingly large collections of data and, in many applications, they require large benchmark datasets.[13–15] Consisting of thousands—and sometimes even millions—of annotated examples, these datasets have practically eliminated the time-consuming task of data collection and annotation in their respective domains. In addition, such benchmarking promotes the direct comparison of algorithms, enabling rapid identification of suitable approaches for further development. However, within the field of IR anti-ship ATR, no large benchmark datasets are available, causing the development of new algorithms to be hampered

Further author information: (Send correspondence to Samuel T. Westlake.)
Samuel T. Westlake: E-mail: s.t.westlake@cranfield.ac.uk

Artificial Intelligence and Machine Learning in Defense Applications II, edited by Judith Dijk, Proc. of SPIE Vol. 11543, 1154309 · © 2020 SPIE
CCC code: 0277-786X/20/$21 · doi: 10.1117/12.2573774

by the cumbersome task of generating bespoke datasets. Furthermore, such datasets typically contain too few examples to train deep neural networks or facilitate robust validation, and are rarely made openly available, further hindering the direct comparison of algorithms.

For future IR anti-ship ATR algorithms to fully leverage state-of-the-art techniques, new approaches to both dataset generation and algorithm training are required—and thus, in this paper, we explore both. In Section 3, we present a new synthetic IR anti-ship-focussed dataset and detail its design, generation and online augmentation. While in Section 4, this dataset is used to investigate several approaches to training the YOLOv3 object detection algorithm in the absence of any real-world IR imagery. The resultant performance is presented in Section 5, its associated implications and recommendations are discussed in Section 6, while Section 7 concludes the paper.

2. RELATED WORK

The principal deficiency of many ship-focussed datasets is their size, with many datasets consisting of fewer than 200 IR images.[16–18] Such datasets are not suited to training DCNNs as, for complex tasks like target recognition, training examples must be sufficiently numerous and diverse to comprehensively represent expected test conditions. For comparison, the COCO object detection dataset, which includes 80 categories, contains *ca.* 120,000 training examples,[13] and there are over 1 million labelled examples in the ImageNet classification and localisation dataset.[15] Consequently, deep learning algorithms trained with such small quantities of data cannot be expected to operate reliably under the wide range of conditions that anti-ship missiles are expected to face.

To the best of our knowledge, the largest relevant dataset is the Singapore Maritime Dataset (SMD),[19,20] which contains a mix of fully labelled off- and on-shore, visual-spectrum and NIR video sequences, totalling 31,653 frames. However, the absence of infrared imagery in the long wave infrared band—the waveband of choice for modern IR anti-ship missiles—means this dataset is not wholly suitable for the naval domain. Furthermore, despite its large size, the SMD still lacks the necessary diversity for training complex algorithms that contain millions of learnable parameters.[19] Nevertheless, the SMD constitutes the largest openly-accessible collection of real-world maritime image data, and is thus used as a valuable benchmark in other applications, such as harbour surveillance[21,22] and collision avoidance.[23]

A second deficiency of many datasets relates to their lack of variation in environmental factors, such as atmospheric conditions, sea states, and the presence of background 'clutter'. Broad ranges of possible conditions are rarely represented, with often just a singular set of atmospheric and sea-surface conditions considered.[18,24] Furthermore, background clutter is generally considered a challenging source of false positive detections, especially in littoral environments, yet such objects are rarely depicted in existing datasets. If ATR algorithms are to become truly robust against the huge diversity of possible deployment conditions, we consider it crucial that future datasets include increased combinations of these conditions.

Furthermore, missile seeker algorithms must be capable of detecting and recognising ships of any size or design within a wide envelope of orientations. However, few datasets fully account for this, with some concerned solely with broadside perspectives, while others ignore elevation and only consider a horizontal view.[3,25,26] Moreover, most existing datasets depict just a small sample of different ship classes—typically no more than six.[1,18,27] This simplifies the detection task and hinders development of recognition and identification capabilities, and therefore it is crucial that future datasets include more numerous and varied collections of both military and civilian vessels.

The collection and open distribution of an IR dataset containing a comprehensive selection of military vessels is unlikely. In response to this, there have been several attempts to generate datasets of synthetically generated imagery. An early notable example used five wireframe CAD models to generate *ca.* 41,000 silhouettes at incremental elevations and azimuths.[1] However, as binary images, these are geared towards the task of ship classification only. Improving on this concept, later work used sophisticated missile target engagement software to generate realistic long wave IR imagery of military vessels, which were then used to train a neural network classifier for target discrimination.[27] These approaches demonstrated the potential of synthetic data for enabling the use of machine learning algorithms for IR ATR. It also brought several advantages into focus, such as the reduced time cost, the collection of comprehensive metadata, and crucially, the opportunity to freely depict military vessels.

Despite these advantages, however, there remain several fundamental challenges concerning the use of synthetic data for anti-ship ATR. For example, existing synthetic datasets are yet to account for the wide range of possible external factors, such as atmospheric conditions, sea states and background clutter. Also, despite the potential to depict any number of ship designs, creation of the these CAD models is a time-consuming and skilled task, and so current synthetic datasets remain just as limited as their real-world counterparts in this respect.[1,27,28] Furthermore, ship thermal signatures change dramatically with external and onboard conditions, though this also is yet to be accounted for, as invariant ship surface temperatures are currently assumed.[27] While synthetic data generation is an attractive solution to some of the many data-related challenges in IR ATR, significant improvements are required if such data is to enable the effective training of robust detection algorithms.

In our view, overcoming the restricted availability of real-world IR imagery and limited realism of synthetic datasets will likely require a hybrid dataset which draws on the advantages of each. Consequently, this paper details an improved approach to synthetic IR dataset generation and presents our own high-quality and openly available dataset for the development of future machine learning IR anti-ship ATR algorithms. We also demonstrate the use of this data for training complex high-performance deep learning object detection algorithm which we evaluate using a sequence of real-world IR imagery.

3. SYNTHETIC DATASET

This section describes the generation of our synthetic IR dataset: CAD model selection and design, thermal property assignment and image generation. This section also describes the sea, sky and background augmentation process, used to increase image complexity and diversity during algorithm training. This dataset has been made openly available.[29]

3.1 Selection and design of ship models

Ten classes of ship were selected across four type designations, as summarised in Table 1. Three types of military vessel were included: corvette, frigate and destroyer; each represented by three different ship classes, and a single civilian vessel, the MV Armorique passenger ferry. These classes of vessels were selected to represent a variety of different designs and all are currently in operation with a range of different nations. Visualised in Appendix A, the ships were modelled using 3D CAD software and were designed to represent their real-world counterparts as accurately as possible.

Table 1. Classes of ships selected for the dataset (displacement refers to standatd displacement).[30]

Class	Type	Commissioned	Length (m)	Displacement (tonnes)
Ada	Corvette	2011	99.0	1,524
Independence	Corvette	2010	128.5	3,188
Visby	Corvette	2009	72.7	630
Alvaro de Bazan	Frigate	2002	146.4	6,250
Jiangkai II (Type 054A)	Frigate	2008	134.0	3,556
Oliver Hazard Perry	Frigate	1977	135.6	2,794
Akizuki	Destroyer	2012	151.0	5,050
Sejong Daewang (KDX-III)	Destroyer	2008	165.9	7,600
Zumwalt	Destroyer	2016	186.0	15,995
Armorique	Ferry	2008	168.0	29,500

3.2 Assignment of thermal properties

Before the IR appearance of these ships could be simulated, it was necessary to apply thermal properties, including temperature and emissivity, to the surfaces of each model. To account for the continuously varying nature of ship thermal signatures, nine versions of each CAD model were created, with each to be prescribed a unique set of surface temperatures.

A selection of real-world ship imagery was collected using a Tau™2 long wave IR camera to inform the design of an algorithm that could generate diverse and realistic thermal signatures. For a given ship model, a mean temperature, μ and standard deviation, σ were drawn from uniform distributions $U(2, 6)$ and $U(5, 20)$ respectively. These were used to define a Normal distribution from which the temperature values for each surface were drawn. If a given surface was an exhaust funnel, its temperature value was elevated by an amount drawn from a $N(40, 10^2)$. The temperature of antenna and radome surfaces were also elevated in 50% of cases, by an amount drawn from $N(U(3, 10), U(0, 4)^2)$. All surfaces were prescribed an emissivity value of 0.97, with exception of glass surfaces, such as windows, which were given an emissivity value of 0.85.

Through this procedure, nine dictionaries of thermal signatures were generated for each ship and applied to their corresponding CAD model. This resulted in nine unique versions of each vessel, as illustrated in Fig. 1 with the Akizuki class destroyer used in this study.

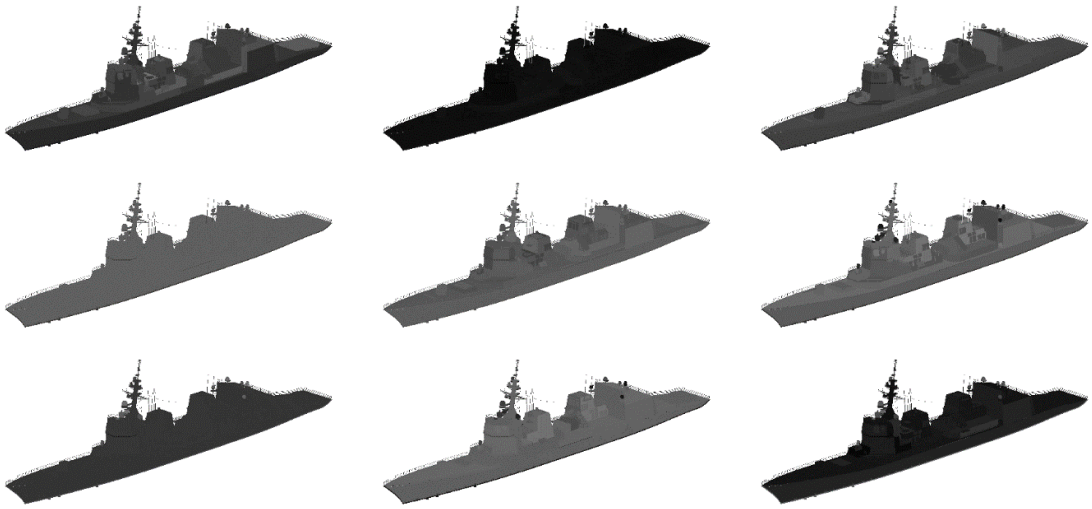


Figure 1. Illustration of the different thermal appearances generated for the Akizuki class destroyer.

3.3 IR image generation

Thermal imagery of the CAD models were generated using CounterSim, a target-missile engagement simulator developed by Chemring Countermeasures Ltd. A virtual long wave thermal imager was defined with a temperature range of 0–100 °C and resolution of 1024×512. Using this virtual camera, thermal imagery for each ship was generated at each ship azimuth in {0, 1, ..., 359}°, each camera pitch in {0, -10, -20}°, and each camera-ship distance in {1000, 1111 1250, 1429, 1666, 2000, 2500, 3333, 5000, 10,000} m. These range values were selected to give a linear reduction in the perceived height of a given target.

During image generation, the ocean was modelled as a flat surface with temperatures drawn from $U(5, 20)$, background sky was modelled as a constant with temperatures drawn from $U(5, 25)$, and atmospheric transmission was modelled using the moderate resolution atmospheric transmission (MODTRAN4) atmospheric model.[31] These arbitrary values were selected to cover a wide range of feasible temperature values for both the sea surface and sky temperatures values, with constant flat surfaces assumed to facilitate augmentation at a later stage. In total, 972,000 images were generated, along with a further 108,000 binary masks for use in online data augmentation and for semantic segmentation.

3.4 Online image augmentation

To increase variation within our synthetic data, a three-stage stochastic online data augmentation pipeline was designed to enable the addition of different sea and sky-states, and background clutter. Each of these steps relied on collections of pre-processed images that were stochastically superimposed into a given image at runtime.

For sky-state augmentation, images that include the sky were collected from various online sources and pre-processed by setting pixels that corresponded to blue sky to zero and converting to grayscale. At runtime, for a given synthetic image, a cloudy image was selected, randomly resized and cropped to shape, with its pixel intensity scaled according to Equation (1), where $\tilde{I}$ is the normalised real-world image and c is the average pixel value of the sky region in the synthetic image. Values of 2 and 50 were chosen for the bounds a and b respectively, to provide a feasible variety of pixel rages, and therefore temperature ranges. The cloudy image was then superimposed above the sea-sky horizon line of the synthetic image, using the synthetic image's corresponding binary mask to preserve target pixels.

$$I_i, j = \tilde{I}_i, j \times U(a, b) + c, \tag{1}$$

The addition of background clutter followed a similar procedure. Images depicting plausible background scenery and objects such as oil platforms, wind turbines, icebergs, small islands, and built-up coastlines were collected from various online sources. Similar to before, these images were pre-processed by the removal of background pixels and conversion to grayscale. At runtime, background images were selected at random and the same process of resizing, cropping and pixel scaling was applied before the scene was then superimposed along the sea-sky horizon line of the given synthetic image. Pixel scaling of the cluttered images was conducted in accordance with Equation (1); however in this case a value of 0 was used for c and the bounds a and b were selected depending on the nature of the cluttered scene. Values of 20 and 80 respectively were used in the case of images that depicted human-made structures, values of 0 and 2 were used in the case of icebergs, and values of 15 and 60 were used in the remainder of cases. These value pairs were selected arbitrarily to correspond feasibly with the nature of the background clutter.

Finally, for sea-state augmentation, two collections of images of ocean surfaces were collected. The first of these include images taken from near sea-level and the second consisted of images taken from an elevated perspective. At runtime, if the sea-sky horizon was visible in the given synthetic image, a real-world image taken from the first collection was chosen, otherwise an image from the second collection was used. The same pre-processing was applied, with Equation (1) used to rescale pixels; however in this case values of 5 and 30 were used for the bounds a and b respectively, and c corresponded to the average pixel intensity of sea region in the synthetic image. There is also scope for the addition of sensor noise at this stage, and the effect of these augmentation processes for both horizontal and elevated images can be seen in Fig. 2 and Fig. 3 respectively.

4. EXPERIMENTS

4.1 Detection Algorithm

The YOLOv3 algorithm[32] was selected due its high accuracy score, as achieved on the COCO benchmark dataset. Selection of YOLOv3 was also influenced by its potential to generalise well across domains, as indicated by the ability of its predecessor[33] to recognise people in art[34,35] despite being trained on trained on the VOC 2007 dataset.[36] Moreover, YOLOv3 is an efficient single-shot algorithm capable of real-time inference and several of its variants have been further optimised for speed.[33,37]

The architecture of YOLOv3 is fully-convolutional, consisting of a backbone feature detector followed by a Pyramid Feature Network-style architecture[38] that enables inference across three scales. The backbone feature detector is the Darknet-53 DCNN which, through its use of 1×1 convolutional layers and residual blocks, is capable of achieving accuracies similar to the much larger ResNet-152 model,[32,39] but at twice the inference speed. The input to Darknet-53 is down-sampled by a factor of 32, whereby the output and two skip connections at $1/16$ and $1/8$ scale are concatenated with downstream feature maps before the final outputs are regressed.

The outputs of YOLOv3 relate to non-overlapping 8-, 16- and 32-pixel cells. For each of these cells, predictions are inferred containing bounding box coordinates, classification predictions, and an "objectness" score, which

Figure 2. Examples of sea-level synthetic images before and after sky, sea and background state augmentation. (Contrast enhanced for illustrative purposes.)

Figure 3. Examples of high-elevation synthetic images before and after sky, sea and background state augmentation (Contrast enhanced for illustrative purposes.)

relates to the algorithm's confidence that the prediction corresponds to a real object. This output structure was modified to include capacity for the prediction of both object type and object class, in accordance with recognition and identification paradigm that is commonplace within the field of ATR.

4.2 Algorithm training

Three approaches to model training were evaluated, in order to test whether synthetic imagery can in fact be used as an effective substitute for real-world IR data. In the baseline experiment, training relied solely on our synthetic dataset, while in subsequent experiments, the training data was augmented with the addition of semi-labelled visual-spectrum and pseudo-IR imagery.

4.2.1 Baseline

Our synthetic dataset was the sole source of training data during the control experiment, with sea, sky and clutter augmentation being applied as described in Section 3. Further augmentation included random shifts and rotations and, to simulate noise that real systems typically incur, sensor noise was added in the form of random Gaussian blur, motion blur and fixed pattern. After this all training images were resized to $576 \times 288 \times 1$.

Model weights were optimised using the Adam method of gradient descent[40] and YOLOv3 loss function.[32,33] The model was trained for 50 epochs of 128,000 randomly selected training instances, with a batch size of 16 and a learn rate of 1e-4, which was gradually reduced by a factor of 0.01 using the cosine learning rate decay.[41]

4.2.2 Inclusion of semi-labelled visual-spectrum images

In order to improve generalisability to real-world imagery, the model was re-trained with the addition of 8,343 images in the visual spectrum, collected from various online sources. These images were collected programmatically, using the ten class names of the ships considered in this study as search terms, and thus type and class labels for each image were designated automatically. However, to avoid the labour-intensive task of manual bounding box annotation, bounding box labels were omitted, and thus these images are referred to as semi-labelled.

An image classifier was added to the YOLOv3 architecture at the output of backbone feature detector, consisting of global average pooling[42] followed by a fully-connected layer. The output layer of this classifier contained 14 output neurons; four of which related to classification of ship type, and the remaining ten being for the classification of ship class. The resultant type and class predictions from this classifier were scored independently using cross-entropy loss. During training, the quantity of semi-labelled data was artificially inflated by duplication to provide a sampling ratio of *ca.* 5 synthetic images to every semi-labelled image.

4.2.3 Inclusion of semi-labelled pseudo-IR images

With the aim of maximising the impact of semi-labelled visual-spectrum data, the third approach used a set of data transforms for the conversion of visual-spectrum data to pseudo-IR imagery. This was done by the application of multiple stochastically linear transforms to the pixel intensity of each image.

4.3 Evaluation

Algorithm evaluation was conducted with a sequence of real-world imagery, collected by using a Tau™2 IR camera, sensitive in the long wave infrared waveband. This sequence depicts a navy frigate traveling under its own power in a littoral environment with a wide range of dense background clutter, including: a navigation buoy, and both rocky and build up shorelines. Metrics used for evaluation were precision, recall and intersection over union (IoU), each of which ignored ship type and class predictions.

5. RESULTS

Regarding the training data, across each of the three experiment the YOLOv3 algorithm achieved average precision, recall and IoU scores that exceeded 0.98, 0.92 and 0.86 respectively. After evaluation with the IR validation sequence, the best performance was achieved by the model trained with a mix of synthetic and semi-labelled pseudo-IR imagery. This version achieved peak values of 0.991, 0.932 and 0.872 respectively for precision, recall and average IoU.

Fig. 4 illustrates the F-scores achieved by each of the three sets of training conditions. When trained with both synthetic and semi-labelled pseudo-IR data, the algorithm achieved a peak validation F-score of 0.96. Yet training with synthetic data only, or with synthetic data supplemented with semi-labelled visual-spectrum data, this score peaked at just 0.42 and 0.50 respectively. This shows that, despite being insufficient for training the YOL0v3 algorithm on its own, the addition of semi-labelled data facilitated a marked increase in performance. However, this was only possible after such data was tailored for the IR domain and its variance maximised.

Fig. 5 illustrates algorithm inference on some examples of our synthetic training data, with ground truth boxes drawn in green and predicted target detections drawn in blue with accompanying type and class predictions. Fig. 6 and Fig. 7 show a selection of the semi-labelled visual-spectrum and pseudo-IR training images respectively, with inferred target drawn in blue and ship type and class predictions written in green. Semi-labelled images

were not accompanied by bounding box labels, meaning the detection algorithm was not explicitly shown how to correctly annotate these examples but was able to do so regardless. This demonstrated the models ability to generalise and indicated that useful features from these real-world images were being learned and recognised by the algorithm.

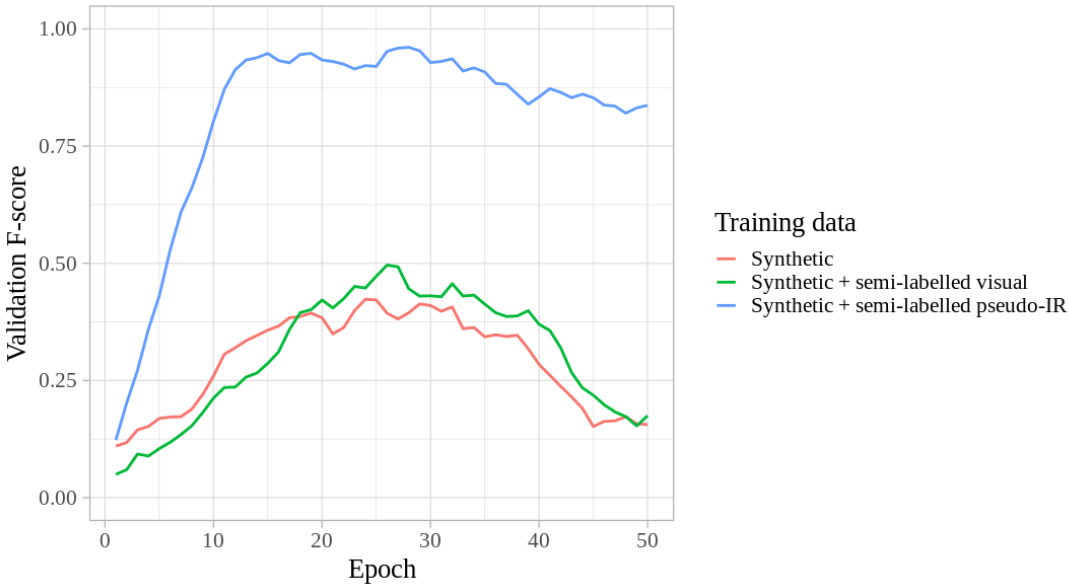


Figure 4. Algorithm validation F-scores.

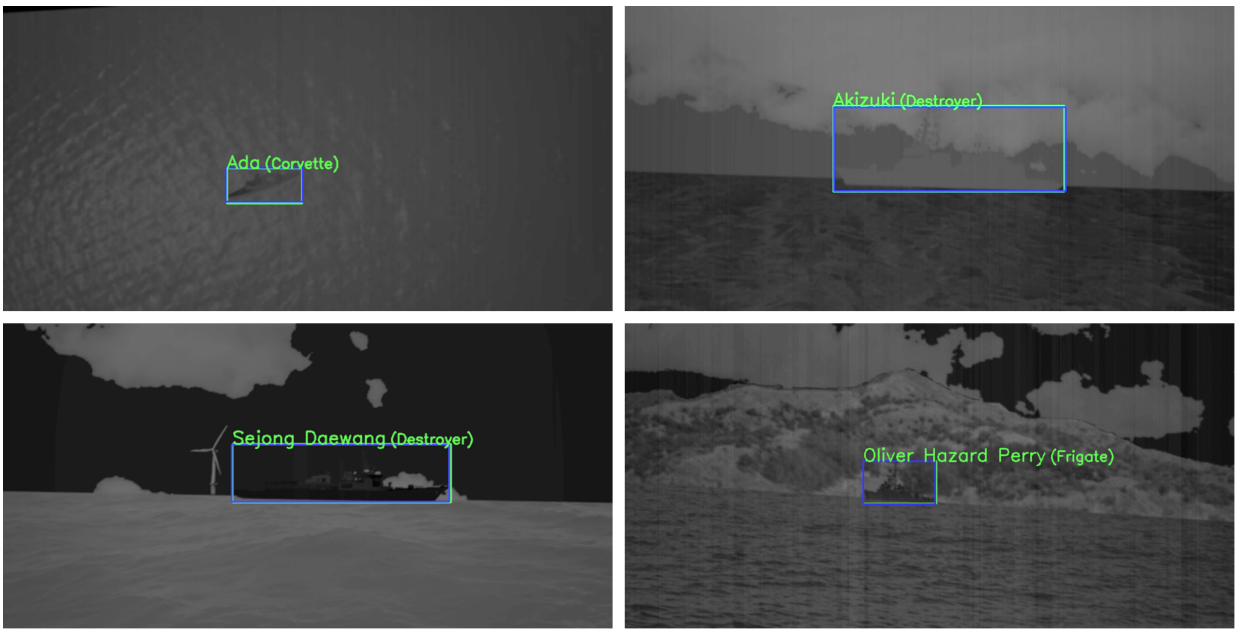


Figure 5. Examples of algorithm outputs on synthetic data during training. (Contrast enhanced for illustrative purposes, green boxes are ground truths, blue boxes and text are algorithm predictions).

Figure 6. Examples of algorithm outputs on semi-labelled data during training (blue boxes and text are algorithm predictions). The depicted boxes were inferred by the algorithm, despite no ground truth bounding boxes being provided for these images.

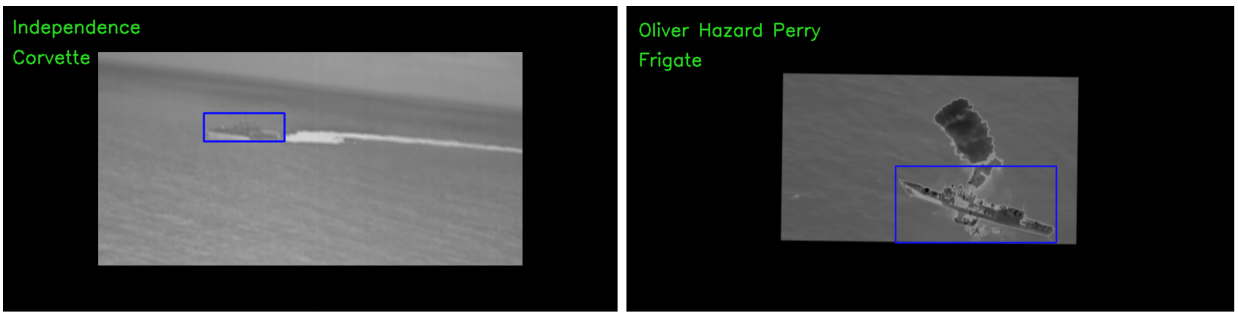


Figure 7. Examples of algorithm outputs on semi-labelled data during training (blue boxes and text are algorithm predictions). The depicted boxes were inferred by the algorithm, despite no ground truth bounding boxes being provided for these images.

6. DISCUSSION

In this paper, an end-to-end deep learning object detector was successfully trained for IR anti-ship ATR with synthetic data as the sole source of training examples. Achieving this required the resolution of several issues regarding the generation of synthetic data. Problems regarding the limited diversity of sky and sea states and background clutter were resolved by the development of an augmentation pipeline that substantially increased the complexity of our synthetic imagery. Secondly, more realistic and varied ship appearances were enabled by increasing the standard of detail in the target CAD models and the development of an approach for the automatic generation of target thermal signatures. These improvements resulted in a deep learning algorithm that was capable of generalising in the domain of real-world IR imagery.

A marked improvement in validation performance was achieved by supplementing our synthetic training data with a relatively small quantity of semi-labelled pseudo-IR imagery. This data appeared to enable the learning of new features that were applicable to the IR validation sequence but not already present in our synthetic dataset, which had the effect of improving ability to generalise in the real-world IR domain. Importantly, this demonstrates that the task of image annotation is not always necessary, and that relevant image data can offer benefits despite a lack of annotation.

That said, the use of semi-labelled visual spectrum data had little effect on algorithm validation performance. The fact that the algorithm was able to successfully detect and recognise targets in these images indicates that features were learned from this data. Though it is apparent that these new features offered little benefit when applied in the IR domain, thus indicating that these visual spectrum images were not sufficiently representative of IR data.

The discrepancy between training and test scores when the algorithm was trained with our synthetic data only indicates that there remains room for improvement of this synthetic data. This synthetic data could be made more representative of the test domain by the use of real IR imagery, as opposed to grayscale visual spectrum

data, in the image augmentation pipeline. Additionally, real-world conditions could be better represented by the inclusion of foreground clutter, such as bow waves and other objects, which are known to cause the partial occlusion of targets. Future iterations of this dataset would also benefit from the inclusion of more atmospheric conditions, including rain, fog and solar glint, and the addition of even more classes of vessel.

7. CONCLUSION

Deep learning algorithms have the potential to redefine the state-of-the-art in the field of IR anti-ship ATR, however they require large and diverse collections of annotated training data. This paper describes how such data can be generated synthetically, using a series of high-fidelity CAD models, each with a range of unique thermal appearances, and an online image augmentation pipeline. Moreover, we trained a modified version of the YOLOv3 algorithm under various conditions and identified the use of pseudo-IR imagery as an effective method to improve the algorithm's ability to generalise.

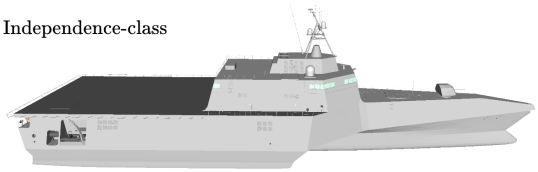
Furthermore, this paper presents our new dataset of realistic longwave IR imagery for the training development of anti-ship ATR algorithms. This dataset contains 972,000 annotated examples, ten different ship classes and uses online augmentation of different sky- and sea-states and background clutter to maximise its complexity and diversity.

APPENDIX A. CAD MODELS

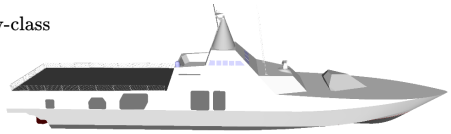
Corvettes

Ada-class

Independence-class

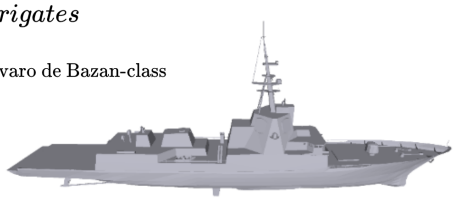


Visby-class

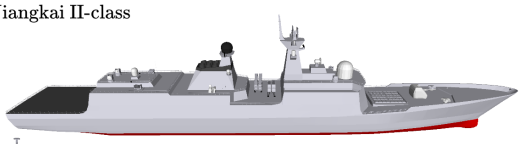


Frigates

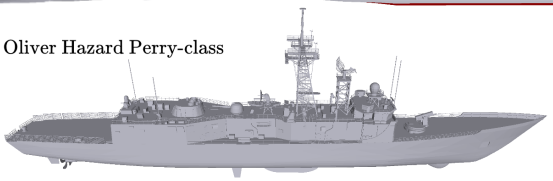
Alvaro de Bazan-class



Jiangkai II-class

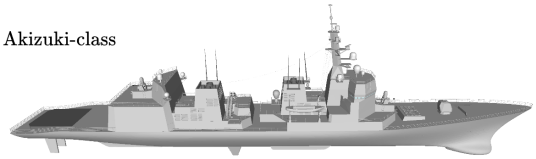


Oliver Hazard Perry-class

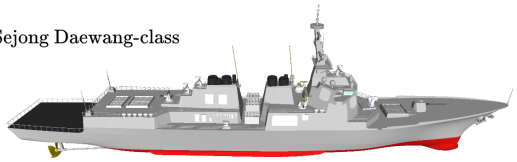


Destroyers

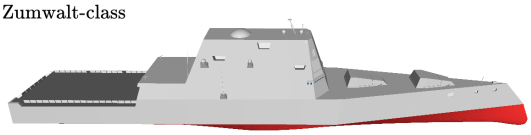
Akizuki-class



Sejong Daewang-class

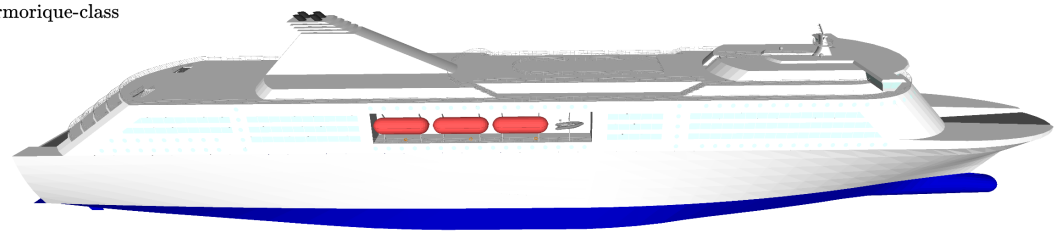


Zumwalt-class



Ferries

Armorique-class



REFERENCES

[1] Alves, J., Herman, J., and Rowe, N. C., "Robust recognition of ship types from an infrared silhouette," tech. rep., Naval Postgraduate School, CA (2004).

[2] Bai, X., Liu, M., Wang, T., Chen, Z., Wang, P., and Zhang, Y., "Feature based fuzzy inference system for segmentation of low-contrast infrared ship images," *Applied Soft Computing* **46**, 128–142 (2016).

[3] Gray, G. J., Aouf, N., Richardson, M. A., Butters, B., Walmsley, R., and Nicholls, E., "Feature-based tracking algorithms for imaging infrared anti-ship missiles," in [*Technologies for Optical Countermeasures VIII*], **8187**, 81870T, International Society for Optics and Photonics (2011).

[4] Hu, M.-K., "Visual pattern recognition by moment invariants," *IRE transactions on information theory* **8**(2), 179–187 (1962).

[5] Lan, J. and Wan, L., "Automatic ship target classification based on aerial images," in [*2008 International Conference on Optical Instruments and Technology: Optical Systems and Optoelectronic Instruments*], **7156**, 715612, International Society for Optics and Photonics (2009).

[6] Tremblay, C. and Valin, P., "Experiments on individual classifiers and on fusion of a set of classifiers," in [*Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat. No. 02EX5997)*], **1**, 272–277, IEEE (2002).

[7] Badrinarayanan, V., Kendall, A., and Cipolla, R., "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence* **39**(12), 2481–2495 (2017).

[8] Girshick, R., "Fast r-cnn," in [*Proceedings of the IEEE international conference on computer vision*], 1440–1448 (2015).

[9] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).

[10] Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J., "Path aggregation network for instance segmentation," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 8759–8768 (2018).

[11] Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Zhang, Z., Lin, H., Sun, Y., He, T., Mueller, J., Manmatha, R., et al., "Resnest: Split-attention networks," *arXiv preprint arXiv:2004.08955* (2020).

[12] "Long range anti-ship missile (lrasm)." https://www.darpa.mil/about-us/long-range-anti-ship-missile. (Accessed on 07/14/2020).

[13] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L., "Microsoft coco: Common objects in context," in [*European conference on computer vision*], 740–755, Springer (2014).

[14] Milan, A., Leal-Taixé, L., Reid, I., Roth, S., and Schindler, K., "Mot16: A benchmark for multi-object tracking," *arXiv preprint arXiv:1603.00831* (2016).

[15] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., "Imagenet large scale visual recognition challenge," *International journal of computer vision* **115**(3), 211–252 (2015).

[16] Zhaoying, L., Fugen, Z., and Xiangzhi, B., "Infrared ship target segmentation based on region and shape features," in [*2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*], 1–4, IEEE (2013).

[17] Tao, W., Jin, H., and Liu, J., "Unified mean shift segmentation and graph region merging algorithm for infrared ship target segmentation," *Optical Engineering* **46**(12), 127002 (2007).

[18] Withagen, P. J., Schutte, K., Vossepoel, A. M., and Breuers, M. G., "Automatic classification of ships from infrared (flir) images," in [*Signal Processing, Sensor Fusion, and Target Recognition VIII*], **3720**, 180–187, International Society for Optics and Photonics (1999).

[19] Moosbauer, S., Konig, D., Jakel, J., and Teutsch, M., "A benchmark for deep learning based object detection in maritime environments," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*], (2019).

[20] Prasad, D. K., Rajan, D., Rachmawati, L., Rajabally, E., and Quek, C., "Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey," *IEEE Transactions on Intelligent Transportation Systems* **18**(8), 1993–2016 (2017).

[21] Andersson, M., Johansson, R., Stenborg, K.-G., Forsgren, R., Cane, T., Taberski, G., Patino, L., and Ferryman, J., "The ipatch system for maritime surveillance and piracy threat classification," in [*2016 European Intelligence and Security Informatics Conference (EISIC)*], 200–200, IEEE (2016).

[22] Palmieri, F. A., Castaldo, F., and Marino, G., "Harbour surveillance with cameras calibrated with ais data," in [*2013 IEEE Aerospace Conference*], 1–8, IEEE (2013).

[23] Campbell, S., Naeem, W., and Irwin, G. W., "A review on improving the autonomy of unmanned surface vehicles through intelligent collision avoidance manoeuvres," *Annual Reviews in Control* **36**(2), 267–283 (2012).

[24] Greer, G., *Advanced Algorithms and Countermeasures for Imaging Infrared Anti-Ship Missiles*, PhD thesis, Cranfield University (2013).

[25] Bizer, M. J., "A picture-descriptor extractions program using ship silhouettes," tech. rep., Naval Postgraduate School, CA (1989).

[26] Herman, J., "Target identification algorithm for the an/aas-44v forward looking infrared (flir)," tech. rep., Naval Postgraduate School, CA (2000).

[27] Gray, G., Aouf, N., Richardson, M., Butters, B., Walmsley, R., and Nicholls, E., "Feature-based recognition approaches for infrared anti-ship missile seekers," *The Imaging Science Journal* **60**(6), 305–320 (2012).

[28] Fernandez, H., de Seixas, J., Neves, S., and Souza Filho, J., "Combining morphological mapping and principal curves for ship classification," in [*International Symposium on Signals, Circuits and Systems, 2005. ISSCS 2005.*], **2**, 605–608, IEEE (2005).

[29] Westlake, S., "Irships dataset." https://doi.org/10.17862/cranfield.rd.12800324 (2020).

[30] [*Jane's Fighting Ships*] (2019).

[31] Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Matthew, M. W., Adler-Golden, S. M., Chetwynd Jr, J. H., Richtsmeier, S. C., Pukall, B., et al., "Modtran4 radiative transfer modeling for atmospheric correction," in [*Optical spectroscopic techniques and instrumentation for atmospheric and space research III*], **3756**, 348–353, International Society for Optics and Photonics (1999).

[32] Redmon, J. and Farhadi, A., "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767* (2018).

[33] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You only look once: Unified, real-time object detection," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 779–788 (2016).

[34] Cai, H., Wu, Q., Corradi, T., and Hall, P., "The cross-depiction problem: Computer vision algorithms for recognising objects in artwork and in photographs," *arXiv preprint arXiv:1505.00110* (2015).

[35] Ginosar, S., Haas, D., Brown, T., and Malik, J., "Detecting people in cubist art," in [*European Conference on Computer Vision*], 101–116, Springer (2014).

[36] Everingham, M. and Winn, J., "The pascal visual object classes challenge 2007 (voc2007) development kit," *University of Leeds, Tech. Rep* (2007).

[37] Huang, R., Pedoeem, J., and Chen, C., "Yolo-lite: a real-time object detection algorithm optimized for non-gpu computers," in [*2018 IEEE International Conference on Big Data (Big Data)*], 2503–2510, IEEE (2018).

[38] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S., "Feature pyramid networks for object detection," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 2117–2125 (2017).

[39] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 770–778 (2016).

[40] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).

[41] Loshchilov, I. and Hutter, F., "Sgdr: stochastic gradient descent with restarts. corr abs/1608.03983 (2016)," *arXiv preprint arXiv:1608.03983* (2016).

[42] Lin, M., Chen, Q., and Yan, S., "Network in network," *arXiv preprint arXiv:1312.4400* (2013).