Effects of Model Drift on Ship Detection Models

Namita Agarwal[1,a], Anh Vu Vo[1,b], Michela Bertolotto[1,c], Alan Barnett[2], Ahmed Khalid[2] and Merry Globin[2]
[1]*School of Computer Science, University College Dublin, Dublin, Ireland*
[2]*Dell Technologies, Cork, Ireland*

Keywords: Object Detection, Remote Sensing Images, LEVIR-Ship Dataset, YOLO Models, Model Drift.

Abstract: The rapid and accurate detection of ships within the wide sea area is essential for maritime applications. Many machine learning (ML) based object detection models have been investigated to detect ships in remote sensing imagery in previous research. Despite the availability of large-scale training datasets, the performance of object detection models can decrease significantly when the statistical properties of input images vary according to, for example, weather conditions. This is known as model drift. The occurrence of ML model drift degrades the object detection accuracy and this reduction in accuracy can produce skewed outputs such as, incorrectly classified images or inaccurate semantic tagging, thus making the detection task vulnerable to malicious attacks. The majority of existing approaches that deal with model drift relate to time series. While there is some work on model drift for imagery data and in the context of object detection, the problem has not been extensively investigated for object detection tasks in remote sensing images, especially with large-scale image datasets. In this paper, the effects of model drift on the detection of ships from satellite imagery data are investigated. Firstly, a YOLOv5 ship detection model is trained and validated using a publicly available dataset. Subsequently, the performance of the model is validated against images subjected to artificial blurriness, which is used in this research as a form of synthetic concept drift. The reduction of the model's performance according to increasing levels of blurriness demonstrates the effect of model drift. Specifically, the average precision of the model dropped by more than 74% when the images were blurred at the maximum level with a 11×11 Gaussian kernel size. More importantly, as the level of blurriness increased, the mean confidence score of the detections decreased up to 20.8% and the number of detections also reduced. Since the confidence scores and the number of detections are independent of ground truth data, such information has the potential to be utilised to detect model drift in future research.

1 INTRODUCTION

With the expansion in maritime transportation applications, automatic ship detection has been a promising research topic (Zhao et al., 2019). Automatic ship detection in remote sensing (RS) images refers to finding ships and locating them in the images automatically (Hashmani et al., 2019). Nowadays, various convolutional neural networks (CNNs)-based deep learning (DL) techniques have been used in RS ship detection for their ability to rapidly and accurately detect ships (Chen et al., 2022). ML challenges such as the Airbus Ship Detection Challenge[1] demonstrated that despite the significant success achieved, detection of ships from medium resolution images is a non-trivial problem. The accuracy and robustness of most ship detection models can be challenged when being faced with images captured in different scenarios, i.e., rough sea, calm sea, thick cloud, thin cloud etc. (Bayram et al., 2022). As supervised ML models rely on a data snapshot available at the training time, the models can become ineffective when the statistical properties of the data vary according to circumstances, for example, abnormal weather conditions, oil spills in the area, under sea gas pipeline explosion, etc. (Mehmood et al., 2023). Such a deterioration in ML model performance is known as model drift, and its prompt identification is crucial for detection and tracking of ships engaging in potentially illegal activities (Zhang et al., 2022).

[a] https://orcid.org/0009-0005-5168-7951
[b] https://orcid.org/0000-0002-6471-4905
[c] https://orcid.org/0000-0003-0122-7656
[1] https://www.kaggle.com/c/airbus-ship-detection

Agarwal, N., Vo, A., Bertolotto, M., Barnett, A., Khalid, A. and Globin, M.
Effects of Model Drift on Ship Detection Models.
DOI: 10.5220/0012443600003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 2: VISAPP*, pages 750-754
ISBN: 978-989-758-679-8; ISSN: 2184-4321

In this paper, we investigate the effects of model drift on ship detection models trained using the LEVIR-ship dataset by (Chen et al., 2022). The dataset comprises a total of 3896 medium resolution (MR) satellite images captured in real world conditions. Particularly, we introduce artificial blurriness, as a form of controlled change, into the imagery data and observe the changes in the accuracy and confidence level of the model trained with the relatively large scale dataset. Furthermore, we measure the effect on model performance degradation in terms of the total number of detection calculated.

2 STATE OF THE ART

In recent years, various deep learning-based object detection models, especially convolutional neural networks (CNNs) have been proposed to detect ships in satellite images (Chen et al., 2022; Xu et al., 2022; Chen et al., 2021; Wu et al., 2021; Chen, 2018). Object detection models based on CNN are generally divided into two categories: one-stage networks [e.g., you only look once (YOLO), single-shot detector (SSD)] (Chen et al., 2021) and two-stage networks [e.g., R-CNN, Fast R-CNN] (Wu et al., 2021). For example, (Xu et al., 2022) proposed an one-stage based low-resolution marine object (LMO) detection YOLO model (LMO-YOLO) for ship detection in low-resolution images. In another work, (Wu et al., 2021) proposed a two-stage based instance segmentation assisted ship detection network (ISASDNet) for ship detection in synthetic aperture radar (SAR) images. A good number of studies use different YOLO architecture-based models such as ImYOLOv3 (Chen et al., 2021) (i.e., Improved YOLOv3), DRENet (Chen et al., 2022) (i.e., YOLOv5 with degraded reconstruction enhancement feature), LMO (Xu et al., 2022) (i.e., YOLOv4 with a multi-scale dilated convolution module). Other studies use different versions of YOLO models, such as YOLOv3, YOLOv4, YOLOv5 directly, in detecting ships for different satellite images ranging from low-resolution RS images to high-resolution RS images, due to their fast and accurate detection capabilities (Huang et al., 2023; Wu et al., 2021).

Furthermore, in the last few decades, various researchers have created large-scale remote sensing datasets (HRSC 2016 (Liu et al., 2016), NWPU-VHR-10 (Zhang et al., 2019), HRRSD (Zhang et al., 2019), DOTA (Ding et al., 2022), DIOR (Li et al., 2020), AI-TOD (Wang et al., 2021)), and made them publicly available to promote object detection tasks for maritime applications. These datasets are either collected from Google Earth or the focus has not been on ship detection with a small number of images containing ships. With respect to model drift, multiple approaches have been introduced over the recent years to analyse the nature of changes in the data, automatically detect model drift and ultimately make the ML model more resilient during its deployment (Gama et al., 2014; Frías-Blanco et al., 2015; Raab et al., 2020; Webb et al., 2018; Mehmood et al., 2023).

3 EXPERIMENT DESIGN

To observe model drift effects, a number of experiments were conducted by training a YOLOv5 ship detection model using original, non-degraded images from the LEVIR-ship dataset (Chen et al., 2022). Subsequently, the model was validated on images subjected to artificial Gaussian blurriness at different levels (i.e., degraded images). The experiment was conducted for five Gaussian blurriness levels determined by square Gaussian kernel sizes of (0, 0), (5, 5), (7, 7), (9, 9), (11, 11). Here, (0, 0) indicates no blurring effect, (5, 5) indicates a slight blurring effect, (7, 7) indicates a moderate blurring effect, (9, 9) indicates a substantial blurring effect, and (11, 11) indicates a more pronounced blurring effect. Every model in this research was trained for 300 epochs. Longer training appeared to reduce only training loss and not validation loss, which could lead to overfitting. All experiments were performed using a NVIDIA A16 GPU available on DELL high computing virtual machine.

4 RESULTS AND DISCUSSION

Table 1 shows the average precision (AP50) of the ship detection model at epoch 300th at different blurriness levels. The mean average precision (mAP50) is a popular metric representing object detection model performance. Note that mAP is calculated by averaging AP over different classes. As we have only one class (i.e., ship), the term AP is more appropriate for this work. The metric corresponds to the area under the precision-recall curve calculated at an Intersection over Union (IoU) of 50%, thereby capturing both the precision and recall of the model in question. The results showed in Table 1 demonstrate that the ship detection model which was trained using clear images performed less well on blurred images. As the level of blurriness increased, the AP50 score of the model reduced accordingly. The reduction in AP50 was as high as as 74.6% when the level of blurriness reached

Table 1: Ship detection model performance at different levels of blurriness; the model was trained using clear images.

AP50				
Without blur	**Blur (5, 5)**	**Blur (7, 7)**	**Blur (9, 9)**	**Blur (11, 11)**
0.777	0.561	0.417	0.304	0.197

(a) (b) (c) (d)

Figure 1: Samples of the detected results from test dataset.

Table 2: YOLOv5 model confidence score values with 0.25 threshold, with testing on both non-degraded and degraded test images at 300th epoch.

Confidence scores			
Test dataset	**Mean**	**Median**	**Variance**
Without blur	0.606	0.659	0.023
blur (5,5)	0.606	0.662	0.023
blur (7,7)	0.579	0.620	0.022
blur (9,9)	0.535	0.561	0.020
blur (11,11)	0.480	0.480	0.017

to the highest level [i.e., kernel size of (11×11)]. That significant reduction in the AP50 scores demonstrates that the artificial blurriness introduced in the imagery negatively impacted the model performance. In other words, it demonstrated the model drift phenomenon.

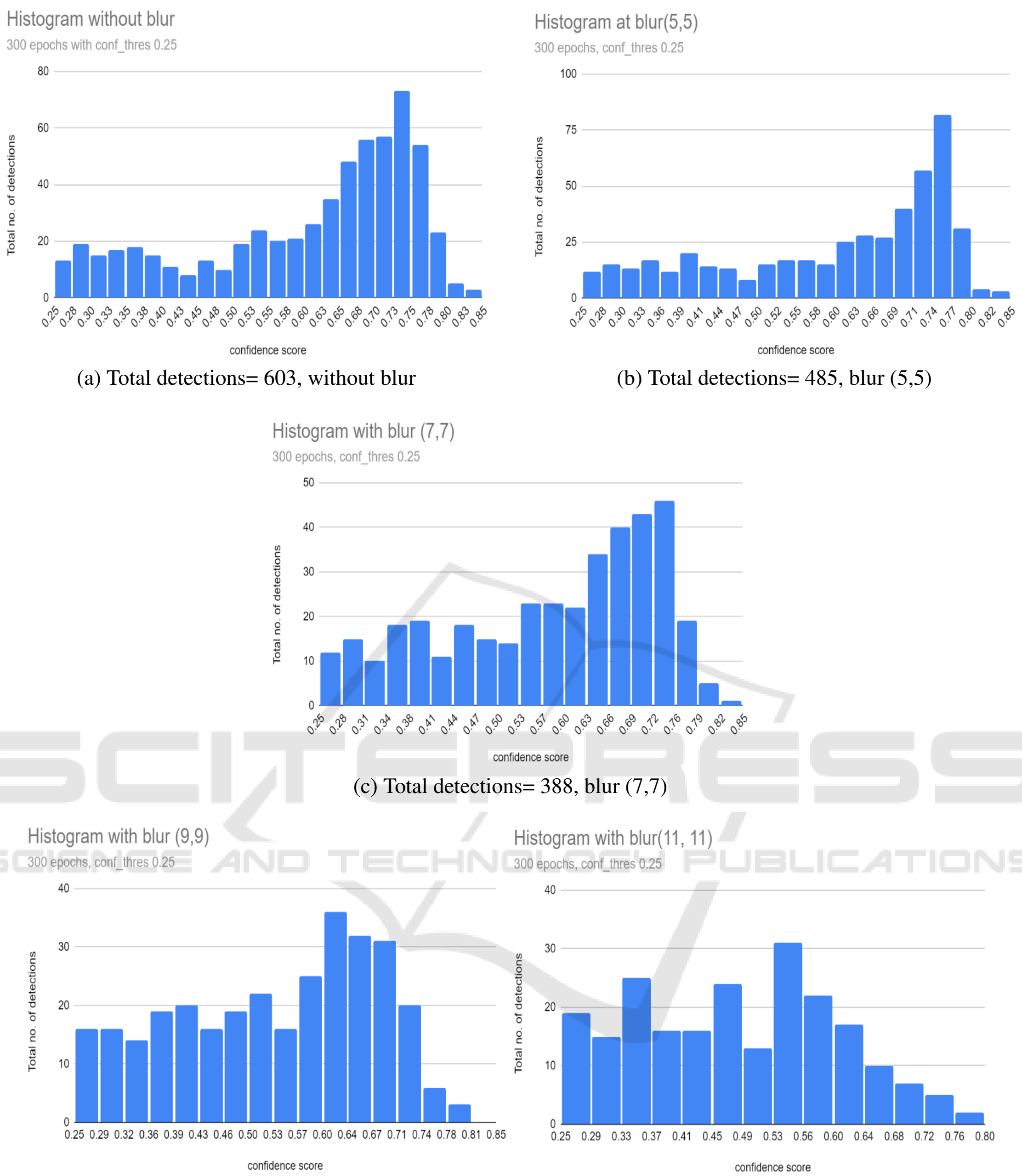
Figure 1 illustrates some detected results from the test dataset, where each detected ship is enclosed within the red bounding box with its corresponding confidence score. This figure depicts that the YOLOv5 model can provide good detection performance under different weather conditions despite being very small size of ships in MR images. The confidence score is measured as the product of the probability of a ship being present in the predicted bounding box and the IoU between the predicted and ground truth bounding boxes. The confidence score ranges from 0 to 1, with 1 indicating the highest degree of confidence in the detection. In this paper, confidence scores are calculated based on a 25% confidence threshold. This shows that the model considers an object as detected only if it has a confidence level of at least 25%, and any detected objects falling below this threshold are disregarded (Yadav et al., 2022).

Furthermore, Table 2 shows the confidence scores when the model was tested under different Gaussian blurriness levels. With an exception for the first level of blurriness [i.e. (5×5)], the confidence scores at all other levels reduced according to the increasing level of image blurriness. The maximum level of reduction in the median confidence score was 27% at the Gaussian kernel size of (11×11). Such a reduction is sufficiently significant to indicate a potential to investigate the confidence score as a proxy to detect model drift. The histograms in Figure 2 confirm and provide a more comprehensive observation of the changes in the confidence scores. In addition to the shift in the overall confidence scores, it is observed that the number of detections reduced as the level of blurriness increased. At the highest level of blurriness, there were only 222 detections compared to 603 detections achieved in the clear image set. Abnormal changes in the number of detections might be another valuable information for detecting concept drift.

5 CONCLUSIONS

This paper presents a preliminary research on concept drift phenomena for ship detection in remote sensing images. A widely known YOLOv5 model was employed as ship detection model in this paper and the model performance was validated against images subjected to artificial blurriness of different limits. The results show that the object detection model degrades its performance (in terms of both accuracy and confidence level) due to model drift. These results demonstrated the effects of model drift in a simple, manually controlled context. In addition, the results provide initial evidence suggesting the potential use of the confidence scores and the number of detections for automatic detection of model drift.

This paper is the first preliminary study of the impact of model drift phenomena on the performance of



(a) Total detections= 603, without blur

(b) Total detections= 485, blur (5,5)

(c) Total detections= 388, blur (7,7)

(d) Total detections= 311, blur (9,9)

(e) Total detections= 222, blur (11,11)

Figure 2: Histogram of confidence scores with 0.25 threshold for testing on both non-degraded and degraded test images at 300$^{\text{th}}$ epoch.

the ship detection model, and does not fully develop a model drift detection method. That potential can be investigated in future research. Another potential direction is to investigate how data augmentation can alleviate the effects of model drift in ship detection models.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the CAMEO (Creating an Architecture for Manipulating Earth Observation Data) project team for their support during our research. They also acknowledge the financial support awarded to the CAMEO project under the Disruptive Technology Innovation Fund (DTIF), an initiative of the Irish Government's Department of Enterprise Trade and Employment (DETE).

REFERENCES

Bayram, F., Ahmed, B. S., and Kassler, A. (2022). From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems*, 245:108632.

Chen, J., Chen, K., Chen, H., Zou, Z., and Shi, Z. (2022). A degraded reconstruction enhancement-based method for tiny ship detection in remote sensing images with a new large-scale dataset. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14.

Chen, L., Shi, W., and Deng, D. (2021). Improved yolov3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images. *Remote Sensing*, 13(4).

Chen, Y. (2018). Airbus ship detection-traditional v . s . convolutional neural network approach.

Ding, J., Xue, N., Xia, G.-S., Bai, X., Yang, W., Yang, M. Y., Belongie, S., Luo, J., Datcu, M., Pelillo, M., and Zhang, L. (2022). Object detection in aerial images: A large-scale benchmark and challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7778–7796.

Frías-Blanco, I., Campo-Ávila, J. d., Ramos-Jiménez, G., Morales-Bueno, R., Ortiz-Díaz, A., and Caballero-Mota, Y. (2015). Online and non-parametric drift detection methods based on hoeffding's bounds. *IEEE Transactions on Knowledge and Data Engineering*, 27(3):810–823.

Gama, J. a., Žliobaitundefined, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Comput. Surv.*, 46(4).

Hashmani, M., Syed, M., Alhussian, H., Rehman, M., and Budiman, A. (2019). Accuracy performance degradation in image classification models due to concept drift. *International Journal of Advanced Computer Science and Applications*, 10.

Huang, Z., Jiang, X., Wu, F., Fu, Y., Zhang, Y., Fu, T., and Pei, J. (2023). An improved method for ship target detection based on yolov4. *Applied Sciences*, 13(3).

Li, K., Wan, G., Cheng, G., Meng, L., and Han, J. (2020). Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159:296–307.

Liu, Z., Wang, H., Weng, L., and Yang, Y. (2016). Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. *IEEE Geoscience and Remote Sensing Letters*, 13(8):1074–1078.

Mehmood, H., Khalid, A., Kostakos, P., Gilman, E., and Pirttikangas, S. (2023). A novel edge architecture and solution for detecting concept drift in smart environments. *SSRN Electronic Journal*.

Raab, C., Heusinger, M., and Schleif, F.-M. (2020). Reactive soft prototype computing for concept drift streams. *Neurocomputing*, 416:340–351.

Wang, J., Yang, W., Guo, H., Zhang, R., and Xia, G.-S. (2021). Tiny object detection in aerial images. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3791–3798.

Webb, G., Lee, L., and Goethals, B. e. a. (2018). Analyzing concept drift and shift from sample data. *Data Min Knowl Disc*, 32:1179–1199.

Wu, Z., Hou, B., Ren, B., Ren, Z., Wang, S., and Jiao, L. (2021). A deep detection network based on interaction of instance segmentation and object detection for sar images. *Remote Sensing*, 13(13).

Xu, Q., Li, Y., and Shi, Z. (2022). Lmo-yolo: A ship detection model for low-resolution optical satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:4117–4131.

Yadav, P. K., Thomasson, J. A., Searcy, S. W., Hardin, R. G., Braga-Neto, U., Popescu, S. C., Martin, D. E., Rodriguez, R., Meza, K., Enciso, J., Diaz, J. S., and Wang, T. (2022). Assessing the performance of yolov5 algorithm for detecting volunteer cotton plants in corn fields at three different growth stages. *Artificial Intelligence in Agriculture*, 6:292–303.

Zhang, L., Li, C., and Sun, H. (2022). Object detection/tracking toward underwater photographs by remotely operated vehicles (rovs). *Future Generation Computer Systems*, 126:163–168.

Zhang, Y., Yuan, Y., Feng, Y., and Lu, X. (2019). Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):5535–5548.

Zhao, Z.-Q., Zheng, P., Xu, S.-T., and Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232.