Journal of Engineering Science and Technology
Vol. 20, No. 1 (2025) 165 - 180

CROSS-DOMAIN ANALYSIS OF YOLOV8 AND FASTER R-CNN MODELS FOR ENHANCED PRECISION IN MARITIME OBJECT DETECTION

ZULKIFLI Z. ABIDIN[1,2,*],
MUHAMMAD A. NORAZARUDDIN[1,2], TENGKU A. T. ANUAR[1,2]

[1]Department of Mechatronics, Kulliyah of Engineering, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia
[2]Centre for Unmanned Technologies (CUTe), International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia
*Corresponding Author: zzulkifli@iium.edu.my

Abstract

Recent advancements in machine vision, particularly through Convolutional Neural Networks (CNNs), have significantly enhanced object detection in maritime environments. This study focuses on the performance of different object detection algorithms and its respective variants: YOLOv8 and Faster-RCNN across two specialised maritime datasets: the SeaShip and Sea Maritime Dataset (SMD). Through comprehensive intra-domain and cross-domain evaluations, we analysed the models' precision, recall, and mean Average Precision (mAP) metrics over 50 training epochs. Notably, the YOLOv8x variant demonstrated exceptional adaptability to the SMD dataset, achieving high precision and recall rates with scores of 98.3% and 96.1%respectively. Whilst the YOLOv8m variant was more effective on the SeaShip dataset. The Faster R-CNN X101-FPN model variant shared similar metrics to the intra-domain evaluations for the YOLO comparisons, however showed significant improvement for cross-domain evaluation, noticeable for the SMD model, outperforming its YOLO counterpart with an improvement of 47.9% for the mAP(50) score. Likewise, the SeaShip model had an improvement of 4.48% for the same metric This paper highlights the challenges of applying machine vision in maritime settings due to environmental variability and dataset specificity. The cross-domain analysis revealed significant performance degradation when models were applied outside their training dataset, emphasising the need for robust domain adaptation strategies. Our findings underscore the importance of selecting appropriate object detection algorithms tailored to specific dataset characteristics to optimise object detection performance in diverse maritime environments.

Keywords: Convolutional neural network, Domain adaptation, Faster-RCNN, Machine vision, Maritime object detection, YOLOv8.

1. Introduction

Machine vision, particularly using CNNs, has increasingly become integral to various fields, including maritime applications. Such cases include ship navigation and security [1, 2], highlighting the importance that machine vision plays in maritime industry. Furthermore, the creation of specialised datasets like the Multi-Category Large-Scale Dataset for Maritime Object Detection (MCMOD) [3] has propelled advancements in object detection research specifically tailored for maritime environments. Such advancements are crucial for navigating the challenges posed by light conditions, horizon effects, and fog, which contribute to the complex environmental conditions [4], thereby a constraint training dataset may inhibit a model's functionality due to the impacts of domain shifting. Therefore, the impact of a robust object detection model is critical, especially in cases where machine vision is needed to enhance navigational systems [5].

In real-world applications, the performance of models trained on labelled data often degrades when tested in new environments or with data from different sources. Lie et al. highlighted for domain adaptation that two algorithms exist: feature space adaptation and classifier adaptation. However, both have drawbacks as feature space adaption is limited to the minimising source-target domain differences whilst the latter requires pre-labelling from the target domain which is unsuitable for limited datasets [6]. Previous works have also tested the concept of two-stage detectors such as Faster R-CNN in successfully minimising the impact of domain adaptation [7, 8]. The drawback of Faster R-CNN is the underperformance in inference speed [9], however Domain Adaptive YOLOv3 models have also been implemented in order match Faster R-CNN adaptation capabilities without minimal inference speed lost by incorporating Regressive Image Alignment followed by Multi-Scale Image Alignment [10]. As such this study will reflect the cross-domain evaluation between one and two detectors in maritime scenarios.

Object Detection can be generalised into two categories: regression/classification and region proposal based such as YOLO and R-CNN models respectively [11, 12]. Both rely on the concept CNNs which can visualise a series of vertical layered stages that incorporate convolutional and pooling layers that aid in the extraction and retaining of features from the input [13]. Within this study the effectiveness of each type of detection algorithm will be evaluated when a single model is trained on a constrained dataset limited to a particular scene, then cross-evaluated with a testing set with different environmental factors. The objective of this study is to highlight the effectiveness of a model's ability to adapt for detection in different situational backgrounds when the initial model is trained from a limited dataset.

2. Materials and Methods

2.1. Model selection

In this study, the YOLOv8, released and developed by Ultralytics in January 2024, is recognized for its effectiveness and scalability in object detection tasks. The architecture allowed for the training of several model detection variations that differ by the model parameters. To determine the trained model to represent the domain required for comparison, each dataset was used and trained under each model variant. These variants were nano(n), small(s), medium(m), large(l), and extra-large(x)

YOLOv8 models variants. Figure 1 provides a detailed illustration of the structural configurations of the YOLOv8 variants as adapted from Terven et al. [14].

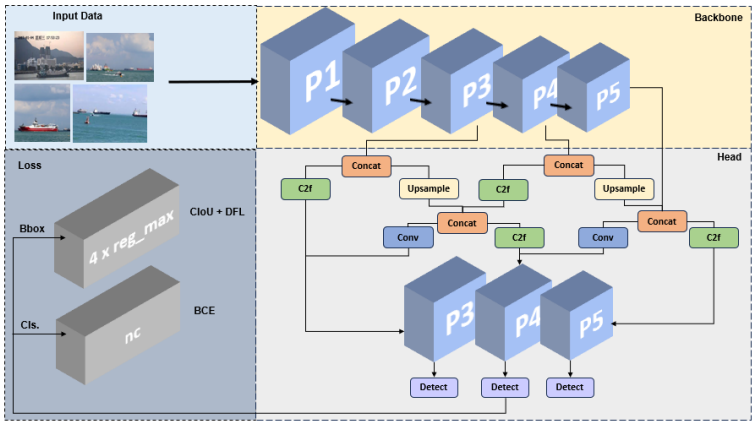


Fig. 1. Architecture of YOLOv8 detection models (P5) - YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x.

These variants range from compact models designed for operational speed to larger models optimised for accuracy. This broad spectrum allows for a comprehensive analysis of domain adaptation across different model capacities. Environmental variables for the captures varied in the time of day consisting of sunrise, midday, afternoon, evening, after sunset and the occurring weather that varied between haze, rain. SeasShip datasets were captured through monitoring cameras in a deployed coastline video surveillance system that ensured diverse data collection by having varied illumination, background, and occlusion captures [15].

Developing and implementing the Faster R-CNN model required the use of Detectron2, a library developed by Facebook AI Research (FAIR) in 2019 to aid in the computer visual research using easier implementation of object detection algorithms [16]. As Faster R-CNN is recognised as the successor to its predecessor Fast R-CNN and R-CNN due to its increased efficiency and speed, the development and implementation of RPN based model will inherit the Faster R-CNN architecture [17]. Detectron2 model library offers multiple model variants, the selection of which was determined by comparing the baseline parameters from its source page as well as comparing its performance metrics when tested on the "MS COCO Validation 2017" set as performed by Oksuz et al. [18]. Hence it was recognised that the variant "X101-FPN" was proven to be more effective as it has seen academic implementation [19]. For these reasons, it was decided that the X101-FPN variant was implemented directly for training the RPN model for the Faster R-CNN model.

2.2. Dataset description

Two online available datasets were chosen to evaluate domain adaptation in object detection models: the SeaShip and the SMD. The criteria of which both had to show were maritime elements that were localised to specific region, however, were distinct to one another, in order to better clearly demonstrate the domain shift effect. The datasets featured high-definition photographs of maritime locations of onshore locations that contained annotated labels of several categories of ships as shown in Fig. 2.

(a)

(b)

Fig. 2. Maritime dataset (a) SeaShip and (b) SMD [5, 20].

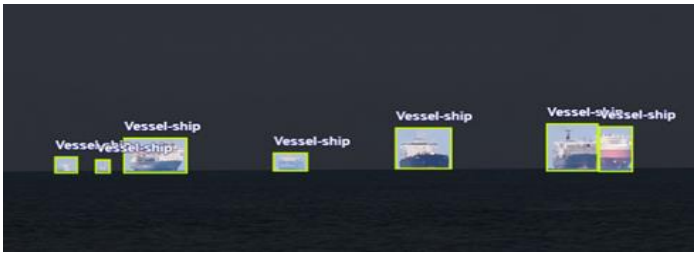
The SMD Data consisted of several onshore videos and on-board videos that were acquired by placing Canon 70D cameras on fixed platforms and moving vessels respectively enabling high definition (1080X1920) captures. Moreover, Near InfraRed (NIR) videos were also captured and annotated using the Canon 70D camera with hot mirror removed and MID-Opt BP800 Near-IR Bandpass filter. The environmental conditions were consistent throughout and focused on a clear day coastline. The SeaShip dataset consisted of high-definition images from coastal surveillance systems, annotated with various ship categories such as ore carriers, container vessels, and passenger ships [20]. A full list of each class name and the number of occurrences is presented in Table 1 for the datasets. The image set was recorded from several viewpoints at Hengqin Island, Zhuhai city, China at varying points of time, limited to 6:00 to 20:00. The varying locations and time were selected to simulate the varying conditions between both datasets.

Table 1. Maritime dataset ship categories.

Dataset	Classes	Instances	Percentage
SeaShip Dataset	Bulk cargo carrier	1406	19.85
	Ore carrier	1504	21.24
	Fishing boat	1986	28.05
	Container ship	742	10.48
	General cargo ship	1043	14.73
	Passenger ship	399	5.63
	Total	**7080**	
Sea Maritime Dataset (SMD)	Boat	382	0.79
	Buoy	750	1.55

Ferry	2237	4.62
Flying bird-plane	231	0.47
Kayak	865	1.79
Speed boat	2227	4.60
Sailboat	616	1.27
Vessel-Ship	34597	71.60
Other	6412	13.27
Total	**48317**	

Annotations across both datasets were standardised to "sea_obstacle" with an ID of 0 for consistency. Prior to training, the naming schema for each identifying class ID differed as it was sourced online and hence had different class names, hence for a fair and homogenous evaluation and model training, each annotated image was relabelled to have the same ID and name. This involved creating a script that edited all the annotation files and changing the ID to 0 and reference name to "sea_obstacle". The pre and post processing results of the data can be viewed in Figs. 3(a) and 3(b) respectively.



(a)

(b)

Fig. 3. Original label compared to modified class groups (a) original data set annotation (b) modified dataset.

Each variant of the YOLOv8 model underwent training over 50 epochs with an initial learning rate of 0.001. To enhance model robustness, standard data augmentation techniques such as random scaling and cropping were applied. The training aimed to optimise the models for precision and recall, essential metrics for evaluating object detection systems.

2.3. Evaluation metrics

A comprehensive evaluation framework was implemented and can be visualised in Fig. 4, incorporating an array of performance metrics to scrutinise the efficacy of each YOLOv8 model variant. These metrics comprised the mean Average Precision (mAP) at specific IoU thresholds: $mAP(50)$ and $mAP(0.50-0.95)$, alongside precision (P) and recall (R). Precision, essential in domains where false positives have considerable implications, is calculated as the proportion of true positive detections (TP) to all positive detections, i.e., TP plus false positives (FP), as shown in Eq. (1). Recall, critical in scenarios where omissions of true positives are unacceptable, is quantified as the ratio of TP to the sum of TP and false negatives (FN), as encapsulated in Eq. (2). Recall measures the model's capability to identify all relevant instances and is crucial in scenarios where missing a true detection could be detrimental [21].

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

A predefined Intersection over Union (IoU) threshold is juxtaposed with the ground truth to determine if a detection qualifies as a TP or FP, adjudicated by the IoU computation as denoted by $IoU(b, b_{gt})$ in Eq. (3) [22]. The performance metric $mAP(50)$, specified in Eq. (4) [23], evaluates the precision of object detection by calculating the average precision through the integration of the precision-recall curve between recall values 0 and 1, ensuring a minimum IoU 50% overlap with the ground truth bounding boxes. The mean value is then calculated by summarising the precision-recall and dividing by the number of classes, C. The metric $mAP(0.50-0.95)$, articulated in Eq. (5), extrapolates and accumulates mAP calculations across IoU thresholds from 0.50 to 0.95 in increments of 0.05 over the number of IoU thresholds, in this case 10. Thus, offering a more granular assessment of the model's detection proficiency across a continuum of precision thresholds.

$$IoU(b, b_{gt}) = \frac{b \cap b_{gt}}{b \cup b_{gt}} \tag{3}$$

$$mAP(50) = \frac{1}{|C|}\sum_{i=1}^{C}\int_{0}^{1} P(r)dr \tag{4}$$

$$mAP(0.50-0.95) = \frac{1}{10}\sum_{t=0.5}^{0.95} mAP(t) \tag{5}$$

2.4. Evaluation strategy

2.4.1. Intra-domain evaluation

The intra-domain evaluation serves as the initial step in assessing the performance of each model variant within the confines of the dataset on which it was trained as shown in Fig. 4. This evaluation is crucial for establishing baseline performance

metrics, which include P, R, and the mAP. By evaluating each model on its training dataset, we can determine its optimal performance metrics in a controlled environment. This phase allows for a thorough analysis of the model's capabilities and limitations without the added complexity of domain variability.

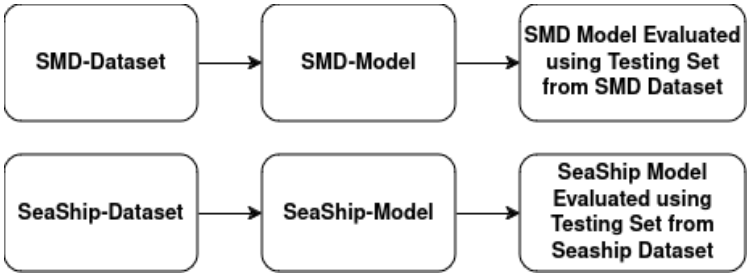


Fig. 4. Intra-domain evaluation.

2.4.2. Cross-domain evaluation

For the cross-domain evaluation, models trained on the SeaShip dataset were tested on the SMD, and vice versa as shown in Fig. 5. This process is designed to assess the effectiveness of domain adaptation strategies by evaluating how well a model trained in one visual domain performs when applied to a different, yet related, visual domain.

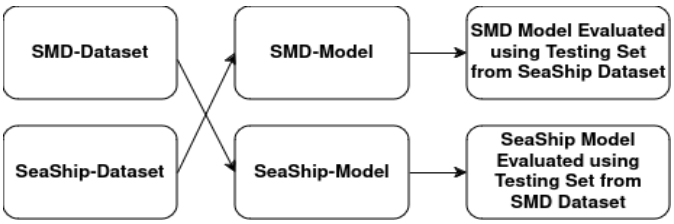


Fig. 5. Cross-domain adaptation.

The primary metric for evaluating adaptation effectiveness is the change in mAP from the training domain (intra-domain) to the new domain (cross-domain). The cross-domain performance degradation is quantified as the difference in mAP between the two datasets represented in Eq. (6).

$$\Delta mAP = mAP_{intra-domain} - mAP_{cross-domain} \tag{6}$$

The mAP scores from the intra-domain evaluation establish a benchmark for performance. The performance degradation, a critical metric in this assessment, is determined by the difference in mAP scores between the two domains. A smaller ΔmAP indicates a more successful domain adaptation by the model. This method of cross-testing not only quantifies the effectiveness of adaptation but also provides insights into how model parameters and architectural choices impact performance when transitioning between domains.

This cross-domain evaluation strategy is crucial for understanding the dynamics of domain adaptation in practical applications, where models must perform

effectively across varied settings without the need for extensive retraining. Through this approach, the research aims to deliver valuable insights into the adaptability of deep learning models in diverse maritime surveillance contexts, thereby enhancing their practical utility in real-world scenarios.

2.5. Model robustness evaluation

The robustness of each YOLOv8 model variant was evaluated by comparing performance metrics from intra-domain and cross-domain evaluations, highlighting the ability of each model to maintain consistency under different environmental conditions—a key factor for real-world applications. Each model underwent a standardised training regimen, keeping all parameters consistent except for the model variant type, to ensure fair comparison. Crucial training parameters such as learning rate, batch size, and number of epochs are explicitly listed in Table 2. Considering that Detectron2 is a separate library to YOLOv8 the configuration may differ. However, it was noted when training the Faster R-CNN model it would match any applicable argument from the YOLOv8 parameters to ensure an evaluation when comparing between both model frameworks. Furthermore, it's noted that any training parameter not specifically mentioned defaults to Ultralytics or Detectron2 documentation settings respectively, ensuring that the study adheres to established machine learning practices and that differences in model performance are attributable to the model variants themselves, rather than variations in training approach.

Table 2. Training arguments for YOLOv8.

Argument	Argument Value
Epochs	50
Patience	3
Batch	5
Imgsz	640

The evaluation phase was systematically structured to ensure consistency across all model variants and types, with specific parameters detailed in Table 3. Like the training phase, any evaluation parameter not explicitly specified defaults to the recommended settings by Ultralytics and Detectron2 configuration files, standardising the evaluation process.

Table 3. Evaluation arguments for YOLOv8 and detectron2.

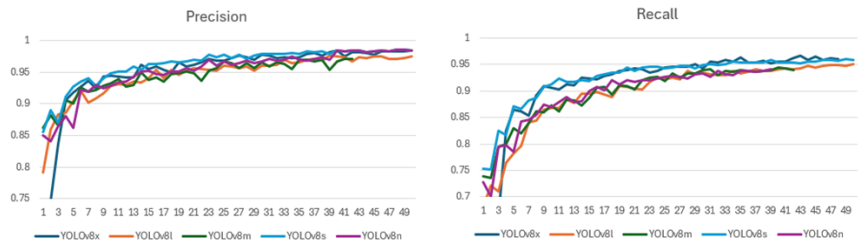
Argument	Argument Value
IOU	0.50
Confidence	0.50
Max Detection	30

This approach allows for a direct comparison between control model evaluation metrics and cross-domain metrics, effectively determining whether there has been a significant drop in model performance.

3. Results and Discussion

3.1. Performance evaluation across datasets

Figure 6 displays the precision and recall curves for the SMD model across 50 training epochs, revealing varied performances among different YOLO variants. Initially, all models exhibit low precision; however, they quickly improve, with the YOLOv8x variant consistently achieving superior precision, effectively adapting to the SMD dataset. Significant improvements in recall are observed in the early epochs across all variants, with rapid enhancements in their ability to identify true positives. By the 50th epoch, both precision and recall stabilise near the optimal score of 1.0, demonstrating the models' high accuracy and robustness in detecting true instances without overfitting.



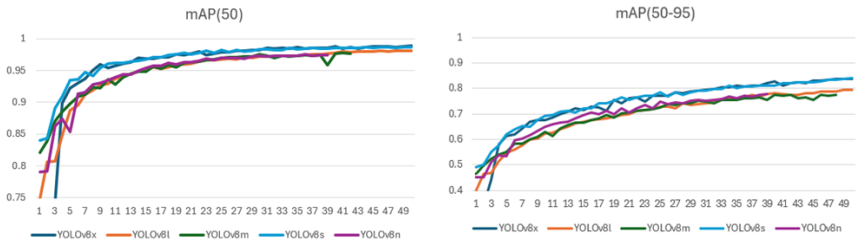
(a) **(b)**

Fig. 6. (a) SMD-model precision values across epoch step (b) SMD-model recall.

Figure 7 presents the mAP(50) and mAP(50-95) metrics for the same training period. All variants start with similar mAP(50) values, suggesting uniform initial detection capabilities. Early in training, a steep increase highlights a swift enhancement in the models' ability to localise and classify objects accurately. The mAP(50) values approach a perfect score, indicating a potential saturation in model performance given the current architecture and dataset. In contrast, the mAP(50-95) metrics, which evaluate performance across stricter IoU thresholds, show continuous improvement, suggesting ongoing refinement in object detection precision and potential for further optimization across a broader IoU range.

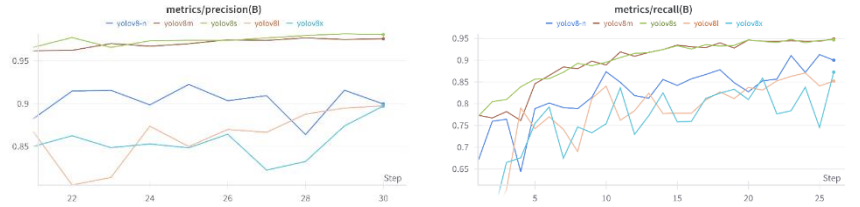
3.2. Intra- and cross-domain analysis

Figure 8 tracks precision and recall metrics for various YOLO model variants over 50 training epochs on the SeaShip dataset. Precision varies initially among the models but rises quickly, with the YOLOv8n variant displaying the highest and most consistent precision, indicating an optimal fit for the SeaShip dataset. Recall also improves swiftly, particularly for YOLOv8n, which leads in consistently identifying all relevant instances. As training progresses, precision and recall trends for all models gradually plateau, indicating that additional training yields diminishing returns and suggesting a learning threshold.



(a) **(b)**

Fig. 7. (a) SMD-model mAP(50) across epoch step (b) SMD-model mAP(50-95) across epoch step.



(a) **(b)**

Fig. 8. (a) SeaShip-model precision across epoch step (b) SeaShip-model recall across epoch step.

Table 4 provides a comparative analysis of the YOLOv8x and YOLOv8m variants on the SMD and SeaShip datasets, respectively. This table highlights the performance metrics of each variant, showing how well each adapts to different maritime environments. The YOLOv8x variant demonstrates remarkable effectiveness on the SMD dataset with higher precision and mAP scores compared to YOLOv8m, suggesting its suitability for environments like SMD. Conversely, YOLOv8m shows superior performance metrics on the SeaShip dataset, emphasising its adaptability to different maritime contexts.

Table 4. Evaluated control models for YOLOv8 models.

Metrics	Dataset	
	SMD-YOLOv8x	SeaShip-YOLOv8m
Precision	0.983	0.913
Recall	0.961	0.889
mAP(50)	0.988	0.954
mAP(50-95)	0.839	0.686

Regarding the evaluation for the Faster R-CNN model, precision could not directly be accessed through the Detectron2 framework; hence, recall and mAP will be analysed. Table 5 shows that for both SMD and SeaShip Faster R-CNN models

have fair effectiveness, with SMD showing slightly significant performance in terms of the mAP(50) scores compared to the SeaShip model. SeaShip was identified to have an underdeveloped recall rate, scoring 79.0%, compared to the SMD's 90.2%, suggesting that the RPN was not able to fully propose suitable bounding boxes for further classification. When directly compared to the YOLOv8 architecture, both models have a slight underperformance in all metrics, however none fall below a 10.0% difference.

Table 5. Evaluated control models for faster R-CNN.

Metrics	Dataset	
	SMD-X101-FPN	**SeaShip- X101-FPN**
Recall	0.902	0.801
mAP(50)	0.979	0.927
mAP(50-95)	0.772	0.650

The evaluation of precision and recall metrics for the SMD model offers significant insights into the classification model's performance. Notably, the YOLOv8x variant shows consistently superior performance with a precision rate of 98.3% and a recall rate of 96.1% on the SMD dataset, confirming the completeness and accuracy of its predictions. This variant also demonstrates significant improvements in the mAP(50-95) metric compared to the mAP(50), due to its larger feature extraction network enhancing detection capabilities across various IoU thresholds.

In contrast, the SeaShip dataset results indicate that the YOLOv8m variant excels in both precision and recall, as well as in the mAP(50) and mAP(50-95) metrics, leading to its selection alongside YOLOv8x for cross-domain evaluation. This evaluation assesses performance degradation when models trained on one dataset are applied to another, highlighting the adaptability and limitations of each model variant as detailed in Tables 6 and 7 for YOLOv8 and Faster R-CNN models, respectively.

Table 6. Cross-domain evaluation for YOLOv8.

Metrics	SMD-Model with SeaShip Dataset-YOLOv8x Model	SeaShip-Model with SMD Dataset-YOLOv8m Model
Precision	0.167	0.508
Recall	0.424	0.301
mAP(50)	0.144	0.446
mAP(50-95)	0.0514	0.240

Table 7. Cross-domain evaluation for faster R-CNN.

Metrics	SMD Faster R-CNN X101-FPN Model with SeaShip Dataset	SeaShip Faster R-CNN X101-FPN Model with SMD Dataset
Recall	0.557	0.524
mAP(50)	0.612	0.483
mAP(50-95)	0.318	0.208

Furthermore, Table 8 explores this degradation, revealing a significant drop across all metrics in the YOLOv8 and Faster R-CNN model. However, it should be noted the drop of these scores is more apparent in the YOLOv8 model with mAP(50) having a loss of 84.4% in the SMD cross evaluation. However, the Faster R-CNN SMD model has a loss of 36.7%, suggesting an improvement difference of 47.4%. Regardless of this it shows that the substantial performance disparity emphasises the challenges of model generalisation and the impact of domain-specific characteristics on model efficacy.

Table 8. Cross-domain degradation.

Metrics	SMD-Model with SeaShip Dataset-YOLOv8x Model (Δ%)	SeaShip-Model with SMD Dataset-YOLOv8x Model(Δ%)	SMD Faster R-CNN X101-FPN Model with SeaShip Dataset (Δ%)	SeaShip Faster R-CNN X101-FPN Model with SMD Dataset
Precision	-0.816	-0.405	-	-
Recall	-0.537	-0.588	-0.345	-0.266
mAP(50)	-0.844	-0.508	-0.367	-0.444
mAP(50-95)	-0.7876	-0.446	-0.454	-0.442

3.3. Discussion

Whilst both models face some loss in cross-domain performance, the extent of the degradation varies between the model and dataset, with YOLOv8 appearing to be more sensitive to the domain shift from the SMD to the SeaShip datasets. On the other hand, Faster R-CNN seems to be more stable with no degradation change exceeding 50% in any of the metrics. This could be attributed to the Faster R-CNN's two stage object detector: its RPN and classification stage. The RPN will propose appropriate bounding box locations and filter irrelevant regions, ensuring that the impact of domain shift is minimal [24].

These findings aligned itself with researcher such as Huang et al. who compared the performances of multiple architectures namely Faster R-CNN, R-FCN and SSD suggesting Faster R-CNN provided the most accurate results and maintained optimal model performance between speed and accuracy if the proposal candidates remained minimal [25]. In training simple ground truths are used, this makes the model susceptible under complex scenarios hence Cho et al. proposed a method to improve mAP scores for Faster R-CNN by further training the data with false positives in the scenario where the RPN introduces too many hard negatives as candidates. Applying this method could lessen the impact of the domain shift [26].

Another key difference that could account for the drastic change in degradation between models could be attributed to each architectures' backbone. YOLOv8 applies CSPDarknet which is recognised to be compact and efficient for real application, whilst Faster R-CNN uses ResNeXt-101 that has a more complex deeper model that utilises parallel branches for broader feature extraction which in turn can enable the model to understand more complex scenarios. Hence, the choice of method and backbone can also greatly affect the mAP scores [27, 28].

Nakamura et al. conducted similar research comparing the mAP scores between YOLOv5 and Faster R-CNN X101-FPN for various datasets: Mask-Detection, Blood Count Cell Detection, Vehicle Detection with training set sizes following the ratio 105:765:435 respectively. In general, Faster R-CNN had significant

improvement, except in the situation for the Blood Count Cell Detection, in which YOLOv5 outperformed. However, it can be noted that the training set for this was much larger in comparison to the other datasets, indicating that Faster R-CNN is suitable for situations where the training set is limited [29]. Additionally further studies have shown Faster R-CNN performing better than YOLOv5 under some situational cases, however it was noted that the inference speed of Faster R-CNN was significantly slower to YOLOv5 [30]. Considering these factors, it is proposed that the significant degradation value during YOLOv8 cross evaluation arise due to its inability to identify more complex features that are to be expected during a sudden shift of domain prior to classification. Furthermore, the lack of varying data types could have undermined the mAP scores.

4. Conclusions

This study systematically evaluated the performance of various regression YOLO model variants, namely YOLOv8x and YOLOv8m in addition to a region based Faster R-CNN detection method, across the SMD and SeaShip datasets. Our findings reveal that YOLOv8x consistently outperforms other variants on the SMD dataset, achieving high precision and recall rates, which underscores its suitability for environments like SMD. This variant also demonstrated superior performance in handling diverse IoU thresholds, evidenced by its significantly improved mAP(50-95) scores. Conversely, the YOLOv8m variant proved more effective on the SeaShip dataset, indicating that different model architectures may be optimised for specific types of datasets. In comparison, the Faster R-CNN model variant has a slight underperformance with metrics differing by less than 10.0%. Suggesting that in a controlled and consistent domain YOLOv8 would be an ideal model for development. However, upon cross-domain evaluation it highlighted the challenges associated with applying models to new environments, as each performance metric suffered drastically especially the YOLOv8 variants, calling for the need of some optimization. From the results it was highlighted that Faster R-CNN under similar training scenarios was able to reduce the extent of degradation. In the case of the SMD model, Faster R-CNN showed high effectiveness in preserving its mAP(50) seeing a drop of 36.7% as opposed to the YOLOv8 variant's 84.4%.

These results emphasise the importance of selecting an appropriate model and objection detection algorithm to maximise detection performance and utilising domain adaptation techniques to reduce the effects of domain shift when the need of a critical robust model is needed across multiple domains in practical applications. Future studies should focus on developing more adaptive model algorithms that can be trained on constrained datasets and dynamically adjusted to allow varying dataset characteristics without substantial degradation in performance, to allow better practicality in complex and varied environments.

Acknowledgement

The author gratefully acknowledges the support from the IUM Engineering Merit Scholarship (KOEIEMS) for Outstanding Students. This scholarship has significantly contributed to the completion of the research presented in this paper. This work was supported in part by the Ministry of Higher Education (MoHE) Malaysia and International Islamic University Malaysia (IIUM); in part by the

Research Management Centre; and in part by the project MyLAB22-002-0002 CXSense System for the SURAYA Unmanned Surface Vessel Fleet.

Nomenclatures

b	Bounding Boxes
b_{gt}	Ground Truth Bounding Boxes
FN	False Negative
FP	False Positive
IoU	Intersection over Union
mAP	Mean Average Precision
P	Precision
R	Recall
TP	True Positive

Abbreviations

CNNs	Convolutional Neural Networks
ID	Identification
R-CNNs	Regional Convolutional Neural Networks
R-FCNs	Regional Fully Convolutional Networks
RPN	Region Proposal Network
SMD	Sea Maritime Dataset
SSD	Single Shot Detector
YOLO	You Only Look Once

References

1 Qiao, D.; Liu, G.; Lv, T.; Li, W.; and Zhang, J. (2021). Marine vision-based situational awareness using discriminative deep learning: A survey. *Journal of Marine Science and Engineering*, 9(4), 397.

2 Haldorai, A.; Lincy, B.R.; Murugan, S.; and Balakrishnan, M. (2024). *Marine vision-based situational automatic ship detection using remote sensing images.* In Haldorai, A.; Lincy, B.R.; Murugan, S.; and Balakrishnan, M. (Eds.), *Artificial Intelligence for Sustainable Development*. Springer, 341-357.

3 Sun, Z.; Hu, X.; Qi, Y.; Huang, Y.; and Li, S. (2023). MCMOD: The multi-category large-scale dataset for maritime object detection. *Computers, Materials & Continua*, 75(1), 1657-1669.

4 Jin, J. et al. (2020). Vision-based target tracking for unmanned surface vehicle considering its motion features. *IEEE Access*, 8, 132655-132664.

5 Prasad, D.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; and Quek, C. (2017). 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.

6 Liu, H.; Shao, M.; Ding, Z.; and Fu, Y. (2019). Structure-preserved unsupervised domain adaptation. *IEEE* Transactions *on Knowledge and Data Engineering*, 31(4), 799-812.

7 Chen, Y.; Li, W.; Sakaridis, C.; Dai, D.; and Gool, L.V. (2018). Domain adaptive faster R-CNN for object detection in the wild. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*, 2018), Salt Lake City, UT, USA, 3339-3348.

8 Yang, B.; Hossain, M.Z.; and Rahman, S. (2023). SS-faster-RCNN: A domain adaptation-based method to detect whether people wear masks correctly. *Proceedings of the International Joint Conference on Neural Networks* (*IJCNN*, 2023), Gold Coast, Australia, 1-8.

9 Redmon, J.; Divvala, S.; Girshick, R.; and Farhadi, A. (2016). You only look once: unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*, 2016), Las Vegas, NV, USA, 779-788.

10 Zhang, S.; Tuo, H.; Hu, J.; and Jing, Z. (2021). Domain adaptive YOLO for one-stage cross-domain detection. *Proceedings of the* 13*th Asian Conference on Machine Learning*, Virtual, 785-797.

11 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.

12 Sharma, N.; Baral, S.; Paing, M.P.; Chawuthai, R. (2023). Parking time violation tracking using YOLOv8 and tracking algorithms. *Sensors*, 23(13), 5843.

13 LeCun, Y.; Bengio, Y.; and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-44.

14 Terven, J.; Córdova-Esparza, D.-M.; and Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4), 1680-1716.

15 Shao, Z.; Wang, L.; Wang, Z.; Du, W.; and Wu, W. (2020). Saliency-aware convolution neural network for ship detection in surveillance video. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(3), 781-794.

16 Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.-Y.; and Girshick, R. (2019). Detectron2. Retrieved July 1, 2024, from https://github.com/facebookresearch/detectron2

17 Ren, S.; He, K.; Girshick, R.B.; and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137-1149.

18 Oksuz, K.; Cam, B.C.; Akbas, E.; and Kalkan, S. (2018). Localization recall precision (LRP): A new performance metric for object detection. *Proceedings of the* 15*th European Conference on Computer Vision* (*ECCV*, 2018), Munich, Germany, 504-519.

19 Pham, V.; Pham, C.; and Dang, T. (2020). Road Damage detection and classification with detectron2 and faster R-CNN. *Proceedings of the IEEE International Conference on Big Data* (*Big Data*, 2020), Atlanta, GA, USA, 5592-5601.

20 Shao, Z.; Wu, W.; Wang, Z.; Du, W.; and Li, C. (2018). SeaShips: A large-scale precisely annotated dataset for ship detection. *IEEE Transactions on Multimedia*, 20(10), 2593-2604.

21 Visa, G.P.; and Salembier, P. (2014). Precision-recall-classification evaluation framework: Application to depth estimation on single images. *Proceedings of the* 13*th European Conference - Computer Vision* (*ECCV*, 2014), Zurich, Switzerland, 648-662.

22 Everingham, M.; Gool, L.V.; Williams, C.K.I.; Winn, J.; and Zisserman, A. (2010). The Pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88, 303-338.

23 Zaidi, S.S.A. et al. (2022). A survey of modern deep learning-based object detection models. *Digital Signal Processing*, 126, 103514.

24 Lu, Y.; Javidi, T.; and Lazebnik, S. (2015). Adaptive object detection using adjacency and zoom prediction. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*, 2016), 2351-2359.

25 Huang, J.; et al. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*, 2017), Honolulu, HI, USA, 7310-7311.

26 Cho, M.; Chung, T.-Y.; Lee, H.; and Lee, S. (2019). N-RPN: Hard example learning for region proposal networks. *Proceedings of the IEEE International Conference on Image Processing* (*ICIP*, 2019), Taipei, Taiwan, 3955-3959.

27 Elharrouss, O.; Akbari, Y.; Almaadeed, N.; and Al-Ma'adeed, S. (2022). Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. *arXiv preprint arXiv*:2206.08016, 1-23.

28 Bochkovskiy, A.; Wang, C.-Y.; and Liao, H.-Y.M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv*:2004.10934, 1-17.

29 Nakamura, A. et al. (2022). Transfer learning for the ICU: Comparing the performance of YOLOv5 and faster R-CNN X101-FPN - Kachachan Em Chotitamnavee. Stanford University, CS231n. Retrieved July 1, 2024, from http://cs231n.stanford.edu/reports/2022/pdfs/110.pdf

30 Mahendrakar, T. et al. (2022). Performance study of YOL Ov5 and faster R-CNN for autonomous navigation around non-cooperative targets. *Proceedings of the IEEE Aerospace Conference* (*AERO*, 2022), Big Sky, MT, USA, 1-12.