International Journal of Research and Publication
ISSUE 5 VOL. 3 (MAY 2024)

Enhancing Maritime Border Security Management with AI-Powered Object Detection and Tracking Systems

Author(s): [1]Echezona Chukwujekwu Davidson
[2]Obi Somto
Corresponding Email: echezona.chukwujekwu@unn.edu.ng

Abstract

Maritime border security faces growing challenges due to increasing threats such as smuggling, illegal fishing, piracy, and unauthorized vessel entry. Traditional surveillance methods often struggle with limitations in real-time detection, tracking accuracy, and scalability. This study explores the integration of AI-powered object detection and tracking systems as a transformative solution for enhancing maritime border security using a novel ship detection approach called YOLOv9 with Adan optimizer (YOLOv9-Adan). Leveraging advanced deep learning algorithms and computer vision, these systems enable accurate identification and continuous monitoring of vessels across vast oceanic areas, including under adverse weather and low-visibility conditions. Our model is trained on a drone-image dataset comprising 3200 images of maritime scenes and ship types in drone views collected from various sources. The experimental results show that our approach using the YOLOv9-Adan model achieves 65.5% mAP, which exceeds the mAP of YOLOv9 by 4.3%. Additionally, this article also provides a comparative analysis of our model YOLOv9-Adan with other existing models in literature with consistently surpassing existing approaches.

Keywords: Ship Detection, Deep Learning, YOLO (You Only Look Once), Adan (ADAptive Nesterov momentum algorithm), PGI (Programmable Gradient Information), GELAN (Generalized Efficient Layer Aggregate Network).

Introduction

The waterway management and Maritime management have been disrupted by numerous issues including ship accidents, smuggling and illegal fishing. All these issues arise from the lack of adequate real-time monitoring of waterways. Thus, making the real-time ship detection very crucial for improving surveillance waterways.

Traditionally, the data sources of ship detection exist from three primary sources: optical remote sensing (ORS) images, Synthetic Aperture Radar (SAR) images and visual images. SAR images are vividly used for ship detection tasks. However, they do have their own limitations as SAR have low resolution which can lead to missing detection of small ships, ORS images cannot be obtained at night time and it is heavily affected by meteorological factors such as clouds and rains. In contrast, visual images which are captured from optical cameras provide more colour and texture information allowing us to detect even small ships more easily and also it is inexpensive and easily accessible. As technology advances in the computer vision and drone technology field, the use of cameras on drones with high resolution have become more effective means for ship detection while lowering the monitoring costs per unit water area. The objective of this work is to enhance the accuracy of the ship detection model to detect ships from drone-imagery. Object detection from images is one area of computer vision which tries to identify the presence of a particular type of object from images or videos. Generally, object detection algorithms are developed using machine learning or deep learning classifiers in order to enhance the detection accuracy and to reduce the detection time. Prior detection there are two steps: feature extraction and feature selection. When machine learning algorithms are used for object detection, separate algorithms are used in the identification of features which are more relevant and can contribute to the classifier performance.

In addition, another detection algorithm using selected features is used to classify the objects to correctly identify the desired objects. Ship detection from drone images is also an object detection problem in which the videos taken by the drone are converted into frames and the key frames are extracted using feature analysis. Moreover, the features are optimised by the application of a detection algorithm which is specialised in feature selection. Finally, an additional detection algorithm is employed to distinguish the ship images from other images. When deep learning algorithms are used, one single algorithm is able to perform all the tasks namely feature extraction, feature selection and detection.

Maritime management and waterway surveillance play pivotal roles in ensuring the safety and security of maritime activities. The surveillance is carried out using drones to get a better visual coverage of the waterways. However, these domains grapple with persistent challenges, including ship accidents, illegal fishing and smuggling, primarily attributed to the shortcomings in real-time monitoring of waterways. In such cases ship detection is crucial for the maritime management from drone captured images.

Literature Survey

Moving Object Segmentation in the video is one of the most crucial tasks in the current stage. Therefore, new techniques were developed by many researchers for moving object detection from a video sequence. The works on unsupervised moving object segmentation are focusing on automatic segmentation of moving objects in the video without prior understanding. Their main objective is to design new techniques for extracting the object of interest from a series of successive video keyframes rather than all frames. In many existing systems, the object segmentation is performed using machine learning algorithms and hence they require a tool to select an object area and such systems take a lot of time. However, the use of deep learning for object detection will address these issues.

Many works are present in the literature that use deep learning algorithms for object detection. Among them, an efficient deep learning-based network, namely the YOLOv5-ODConvNeXt proposed by Cheng S et al [1] for accurate ship detection from drone-captured images. Deep learning-based object detectors can be roughly divided into two types namely, transformer [2] based detectors and Convolutional Neural Networks (CNN) based detectors. The outstanding performance of Transformer models in the natural language Processing in broad range lead to the interest of applying it to computer vision tasks. A Transformer structure is made of two blocks: an encoder and decoder, which uses a self-attention mechanism to detect relationships among the elements of a sequence. An image can be considered as a sequence by dividing into N × N image patches, which serves the basic concept of Transformer. Detection Transformer (DETR) [3] is the first end-to-end transformer-based object detector which considers object detection as a set prediction problem.

International Journal of Research and Publication
ISSUE 5 VOL. 3 (MAY 2024)

To eliminate the series of non-maximum suppression (NMS) strategy and anchor boxes which are handcrafted, the CNN-generated features processed by Transformer are used by DETR. Although DETR achieves high performance on the MS-COCO [4] dataset, it faces various obstacles including missed detection on small objects and a long training duration. Inspired by this, Dai et al. [5], Zhu et al. [6] proposed Deformable DETR, which aims to accelerate the convergence speed and promote the accuracy on detecting small objects of DETR. The Deformable attention module of DETR focuses on partial important points around a reference point of feature maps.

You Only Look Once (YOLO) [7] is the first onestage CNN-based detector. Two-stage detectors make predictions on various regions of interest, which ignore the global information of the whole input image. To address this, YOLO considers object detection as a single regression problem. Compared to two-stage detectors the neural network can directly obtain bounding boxes and its probabilities from image pixels of each area. The input image is divided into a series of grids, with each grid responsible for detecting objects in that region of the image. Many works are present in the literature that used deep learning algorithms for object detection. Among them, Xu et al [8] provided a survey on the works that considered the application of Swarm Intelligence Optimization Algorithms in Image Processing. Zhou et al. [9] proposed an improved beetle swarm optimization algorithm for the smart navigation control.

Gupta et al [10] proposed an improved PSO algorithm-based Convolutional Neural Network approach for ship detection using deep classifications. Iwin Thanakumar Joseph et al. [11] proposed a new ship detection and classification model by using a hybrid optimization enabled Deep Learning approach. Wang et al [12] proposed the concept of programmable gradient information for implementing the various changes that are required by deep neural networks to achieve multiple objectives in Yolov9. This architecture has confirmed that the programmable gradient information introduced in the deep learning classifier has gained superior results on lightweight models used in object detection. Wang et al [13] proposed a deep learning classifier called Cross Stage Partial Network (CSPNet) to handle the problem of heavy inference computations required by many existing deep neural networks based on the perspective of network architecture. This work considered the duplicate gradient information within network optimization for effectively handling the variability of the gradients by integrating feature maps from the beginning and the end of a network stage. Their experiments proved that their model has reduced the computations by 20% and provided superior accuracy than the state-of-the-art approaches.

Wang et al [14] proposed gradient path design strategies for designing a high-efficiency and high-quality expressive network architecture. The authors provided new strategies at the layer-level, the stage-level, and the network-level for enhancing the quality of architecture design. They also proved that their design strategies are superior and feasible from theoretical analysis and experiments. Bo et al. [15]

gave a detailed survey on the existing works of ship detection and classification on images acquired from ORS source. They explained the methods of ship detection and classification for practically testing in optical remote sensing images, and provided the corresponding feature extraction strategies and statistical data.

De Arijit et al. [16] addressed the maritime inventory routing problem by satisfying the demand at different ports. A mixed integer non-linear programming model was presented by the authors considering various scheduling and routing constraints,

International Journal of Research and Publication
ISSUE 5 VOL. 3 (MAY 2024)

loading/unloading constraints and vessel capacity constraints. Moreover, non-linear equation between fuel consumption and vessel speed was incorporated in their model for capturing the sustainability aspects. As per their model, penalty costs are incurred if the ship arrives early before the starting of the time window or if it finishes its operation after the ending of the time window. They used a swarm intelligence approach to solve the problem and to provide optimal results when compared to other related models.

Escorcia-Gutierrez et al [17] proposed for the small ships that the autonomous ship detection is efficient with optimal mask regional convolutional neural network technique. They aimed at the minimization of the challenging issues in ship detection such as the unpredictable errors of manual navigation for reducing the human labour and also for increasing the navigation security and profit margin. Moreover, the data augmentation process is performed in their system to resolve the issue of the limited number of real-world samples of small ships and hence it is able to detect small ships in most cases accurately. The Mask RCNN with SqueezeNet model is used by the authors to detect ships accurately. Furthermore, the Colliding Body's Optimization (CBO) algorithm with the weighted regularised extreme learning machine (WRELM) technique is also employed in this system to classify detected ships more effectively. Their comparative results analysis has demonstrated that their technique achieved the maximum accuracy of 98.63% over the current methods available in the literature.

Goodfellow Ian et al. [18] proposed an improved R3Det algorithm for handling the task of ship target detection on ORS images. This model has been developed by the authors to solve the problem of low accuracy in ship target detection in optical remote sensing ship images that occur due to complex scenes and large-target-scale differences. An efficient channel attention (ECA) module is also added in their model to make the network gather in the target area. The improved algorithm is applied to the ship data in the remote sensing image data set and the effectiveness of this improved model detection has been verified by the authors through experiments with R3Det and other models in a complex environment and for a small target ship.

Proposed Methodology

Block Diagram

Figure 1 shows the overall structure of the proposed ship detection system that uses YOLO v9 with Adan optimiser.

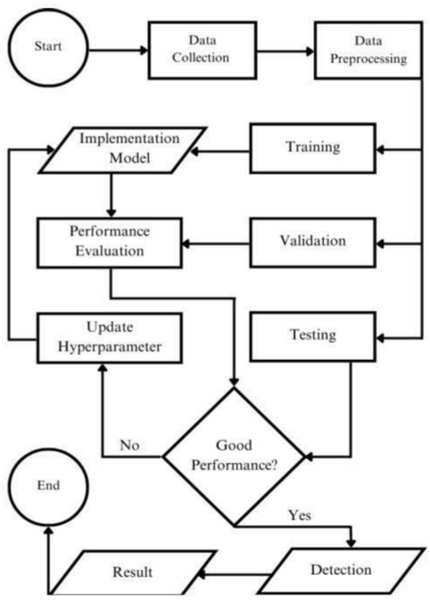


Figure 1: Flowchart of the ship detection system using proposed YOLOv9-Adan model

The proposed model is built upon YOLOv9 while integrating it with an adaptive optimisation mechanism adjusts the learning rate effectively depending on the performance of the network. YOLOv9 uses Adam optimiser as its default optimiser.

International Journal of Research and Publication
ISSUE 5 VOL. 3 (MAY 2024)

Through extensive experimentation and comparative analysis against the existing literature, our findings demonstrate higher precision, recall and overall performance metrics.

Dataset Description

The dataset was initially shared by Shuxiao Cheng, for the paper titled "Deep learning based efficient ship detection from drone-captured images for maritime surveillance.". Dataset was obtained for this study from their provided GitHub link and was made accessible publicly with the access key: cuc6. There were no modifications made to the original dataset and was used for the purpose of training the model as proposed in this research.

The dataset consists of 3200 images of ships taken from drones or drone view. The images in the datasets have three sources including MS-COCO dataset (Lin et al., 2014), Pascal VOC (Everingham et al., 2010, 2015) dataset and images captured by drones. All the images are in JPEG format with the resolution of 1080 x 1920. The dataset is divided in the ratio of 7:1:2 for training (2240 images), validation (320 images) and testing (640 images) respectively. Figure 2 shows the sample images of the dataset.

Figure 2: Sample images from the dataset

Deep Learning for Classification

Using a set of labelled data instances known as training, classification is used to create a model termed a classifier. Testing is the process of applying the learnt model to classify a test instance into one of the classes. Techniques for object detection based on classification function in a similar two-phase manner. Using the available labelled training data, a classifier is learned during the training phase. Using the classifier, the testing step assigns a test instance to either the provided object or not. Deep learning in artificial intelligence (AI) teaches machines to process information similarly to the human brain. Deep learning algorithms can recognise complex patterns in photos, text, audio, and other kinds of data to produce accurate insights and projections.

YOLOv9

An end-to-end neural network that predicts bounding boxes and class probabilities simultaneously is the method that You Only Look Once (YOLO) suggests using. YOLO outperforms other existing real-time object detection algorithms and produces state-of-art results. Unlike other methods like Region Based Neural Network (RCNN), which operates using Region Proposal Network to detect the potential region of interest, YOLO uses a single fully connected layer to make all of its predictions. Moreover YOLO, as the name suggests, only one iteration is needed whereas other network models require many iterations for the same image. The YOLOv9 architecture introduces two new methods: Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregate Network (GELAN), resulting in improved gradient flow and information retention. In our implementation, we used Adan optimizer in YOLOv9 for better results.

Programmable Gradient Information

PGI is an auxiliary supervision framework which oversees the flow of gradient information across various semantic levels, improving performance of the model. There are three main components in PGI, named, main branch, auxiliary reversible branch, and multi-level auxiliary information. The main branch handles both forward and back propagation during the interference process. As the network goes deeper, leading to information bottleneck, the auxiliary reversible branch preserves the information integrity and lowers the information loss in the main branch by employing reversible functions. Moreover, the issue of error accumulation from deep supervision mechanisms is addressed by multi-level auxiliary information by introducing supplementary information at various levels. This forms the theoretical basis to have an excellent performance of YOLOv9 in ship detection. **3.4.2 Generalized Efficient Layer Aggregate Network**

GELAN enhances the information integration and propagation efficiency in model training. GELAN combines the strengths of Cross Stage Partial Network (CSPNet) [19] and Efficient Layer Aggregate Network (ELAN) [20] to migrate information loss in propagation and improve inter-layer information interaction. This architecture makes it perfectly suitable for ship detection using drone imagery.

Adaptive Nesterov Momentum Algorithm

Optimizers are very crucial and useful in training deep neural networks, as it mitigates the error or loss function by refining the model parameters in an iterative approach. Stochastic Gradient Descent (SGD) is one of the several optimization algorithms which is one of the simplest and widely used optimization algorithms, where subsets of training data are randomly selected for computing gradient and the parameters are updated based on the computed gradient.

Adaptive Nesterov Momentum Algorithm (Adan) is an efficient Deep Neural Network optimiser, which utilises modified Nesterov Momentum Estimation (NME) approach. Adan employs a novel approach by estimating not only first-order moments of gradient but also the second-order moments of gradient, as it predicts the future direction of gradient and helps the optimiser to function more effectively.

Algorithm

Step 1: Read one image.

Step 2: Set the layers for feature selection and feature extraction for each classifier.

Step 3: Apply YOLOv9 classifier and extract the features.

International Journal of Research and Publication
ISSUE 5 VOL. 3 (MAY 2024)

Step 4: Perform feature selection using the activation functions, convolution operators and pooling operators by applying YOLOv9.

Step 5: Perform detection using the fully connected network of YOLOv9.

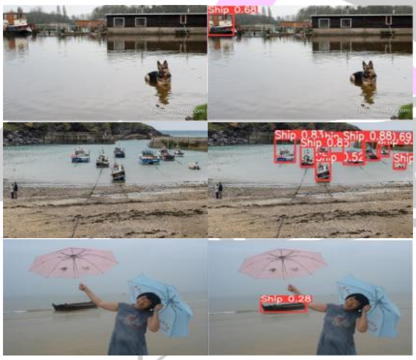
Step 6: Apply the Adan optimiser for checking the error rate and change the parameters.

Step 7: Repeat Steps 1-5 for all the images present in the database.

Step 8: Find Precision, Recall, Accuracy.

Step 9: Send the results to the user interface.

Figure 3 shows the sample input and out images of our proposed YOLOv9-Adan model



(a) (b)

Figure 3: (a) Sample input images containing ships, (b) corresponding output images using proposed YOLOv9-Adan model.

Experimental Results

Performance Metrics

The performance of a machine learning model is measured and evaluated using Evaluation Metrics. The evaluation metrics used for evaluating our model are Precision, Recall and mean Average Precision(mAP). The evaluation metric, precision is the ratio of true positives to the resultant sum of true positives and false positives. True positives are the objects correctly detected as a ship and false positives are the non-ship object detected as a ship object. Precision is calculated using the formula as seen in equation (1).

$$\boldsymbol{Precision} = \boldsymbol{True\ positives} \,/\, (\boldsymbol{True\ positives} + \boldsymbol{False\ positives}) \qquad (1)$$

The evaluation metric, recall is the proportion of true positives to the sum of true positives and false negatives. True negatives are the objects which are correctly detected as non-ship objects and false negatives are the ship objects detected as non-ship objects. Recall is calculated using the formula as seen in equation (2).

$$\boldsymbol{Recall} = \boldsymbol{True\ positives} \,/\, (\boldsymbol{True\ positives} + \boldsymbol{False\ negatives}) \qquad (2)$$

The evaluation metric, mAP is equipped to evaluate the accuracy of the models. It calculates the average precision value for recall value across 0 to 1. It is based on the precision-recall curve.

Result Analysis

Table 1 shows the comparison of some state-of-art object detectors. Overall, the best performing methods is YOLOv9 with Adan optimizer with mAP score of 65.5%, Precision score of 74.2% and Recall score of 57.7% with the epoch of 100. It

International Journal of Research and Publication
ISSUE 5 VOL. 3 (MAY 2024)

performs better than other state-of-model such as CNN, CNN-Adan, YOLOv5, YOLOv5-ODConvNeXt, YOLOv8, YOLOv9 with adan and SGD optimizer as it seen from the above table.

Model	**Precision (%)**	**Recall (%)**	**mAP(%)**
CNN	51.2	40.7	44.9
CNN-Adan	53.9	44.5	46.2
YOLOv5	54.6	46.8	47.0
YOLOv5-ODConvNeXt	57.3	48.5	50.9
YOLOv8	61.8	50.7	59.4
YOLOv9-Adam	66.4	52.1	61.2
YOLOv9-SGD	69.7	54.6	63.7
YOLOv9-Adan(proposed)	**74.2**	**57.7**	**65.5**

Table 1: Comparison of state-of-art object detectors

Figure 4 shows the bar chart of various state-of-art object detector's precision, from which the proposed YOLOv9-Adan model outperforms by having 74.2% precision.

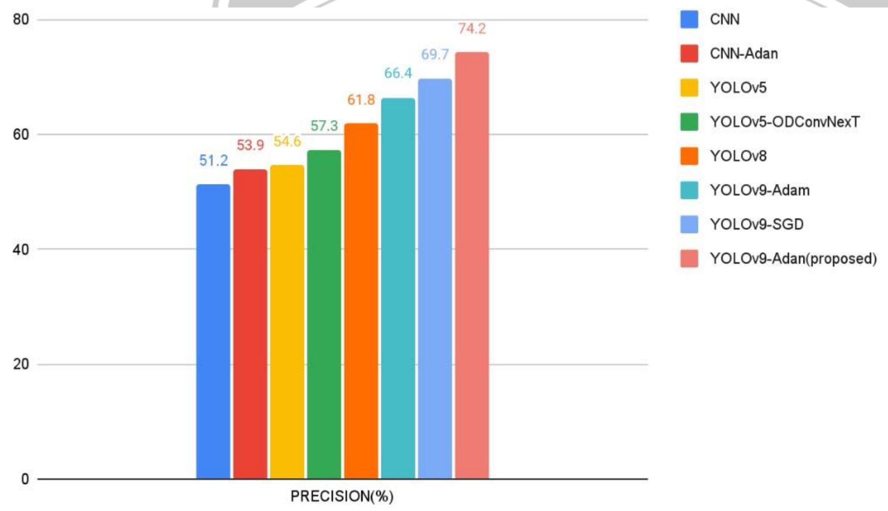


Figure 4: Comparative Analysis of state-of-art object detectors based on precision

International Journal of Research and Publication
ISSUE 5 VOL. 3 (MAY 2024)

Figure 5 shows the bar chart of various state-of-art object detector's recall, from which the proposed YOLOv9-Adan model outperforms by having 57.7% recall.

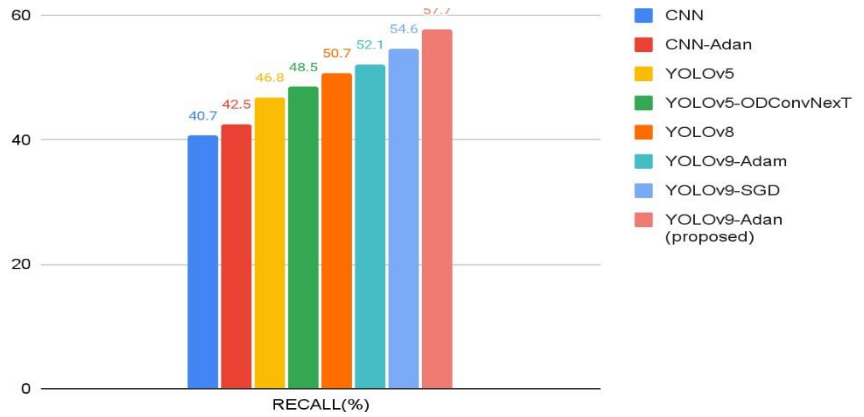


Figure 5: Comparative Analysis of state-of-art object detectors based on precision

Figure 6 shows the bar chart of various state-of-art object detector's mAP, from which the proposed YOLOv9-Adan model outperforms by having 65.5% mAP.

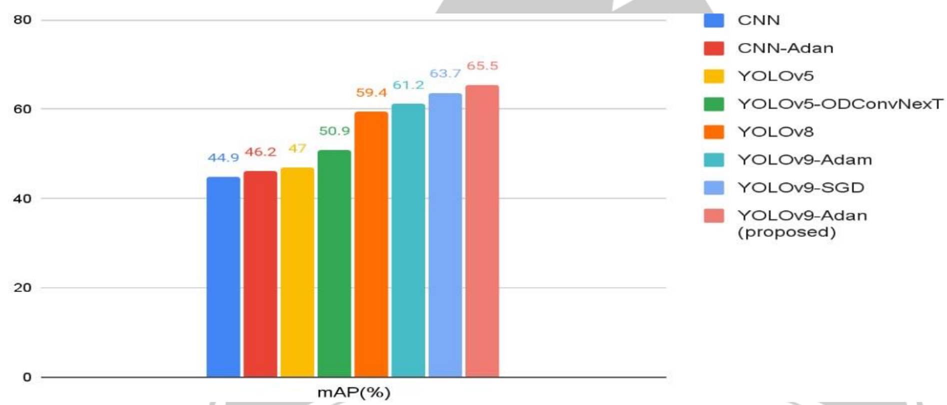


Figure 6: Comparative Analysis of state-of-art object detectors based on mAP

Conclusion

This paper provided a new deep learning approach by the development of a new model yolov9 with adan optimiser. The major advantages of the proposed algorithm are its ability to identify the ship objects more accurately with reduced false positive rate and increased precision and recall values. The experimental results proved that the proposed YOLOv9-adan model achieves better performance with a Precision score of 74.2%, Recall score of 57.7% and 65.5% when compared to existing literature. Future works aim to use fuzzy rules in the place of crisp rules which are used in this work in the form of IF THEN rules.

References

1. Cheng, S., Zhu, Y. and Wu, S. (2023). Deep learning based efficient ship detection from drone-captured images for maritime surveillance. Ocean Engineering, 285, 115440.
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. and Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, vol 30.
3. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In European conference on computer vision (pp. 213229). Cham: Springer International Publishing.
4. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D. and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13 (pp. 740-755). Springer International Publishing.
5. Dai, Jifeng, Yi Li, Kaiming He, and Jian Sun. "R-FCN: Object detection via region- based fully convolutional networks." Advances in neural information processing systems 29 (2016).
6. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J., 2021. Deformable DETR: Deformable transformers for end-to-end object detection. In: International Conference on Learning Representations.
6. Maity, M., Banerjee, S., and Chaudhuri, S. S. (2021, April). Faster r-cnn and yolo based vehicle detection: A survey. In 2021 5th international conference on computing methodologies and communication (ICCMC) (pp. 1442-1447). IEEE.
7. Xu M, Cao L, Lu D, Hu Z, Yue Y. Application of Swarm Intelligence Optimization Algorithms in Image Processing: A Comprehensive Review of Analysis, Synthesis, and Optimization. Biomimetics. 2023; 8(2):235.
8. Zhou, L., Chen, K., Dong, H., Chi, S. and Chen, Z., 2020. An improved beetle swarm optimization algorithm for the intelligent navigation control of autonomous sailing robots. IEEE Access, 9, pp.52965311.
9. Gupta, V. and Gupta, M., 2022. Improved PSO Algorithm-Based Convolutional Neural Network Approach for Ship Detection and Classifications. SN Computer Science, 3(4), p.318.
10. Iwin Thanakumar Joseph, S., Shanthini Pandiaraj, N., Sarveshwaran, V. and Mythily, M., 2022. Ship Detection and Classification Using Hybrid Optimization Enabled Deep Learning Approach by Remote Sensing Images. Cybernetics and Systems, pp.1-28.
11. Wang, C.Y., Yeh, I.H., Liao, H.Y.M.: Yolov9: Learning what you want to learn using programmable gradient information. arXiv preprint arXiv:2402.13616 (2024).
12. Wang, C.Y., et al.: Cspnet: A new backbone that can enhance learning capability of cnn. In: Proceedings of the IEEE/CVF conference on computer vision and pattern

recognition workshops, pp. 390–391. (2020)

13. Wang, C.Y., Liao, H.Y.M., Yeh, I.H.: Designing network design strategies through gradient path analysis. arXiv preprint arXiv:2211.04800 (2022)
14. Bo, L. I., X. I. E. Xiaoyang, W. E. I. Xingxing, and T. A. N. G. Wenting. "Ship detection and classification from optical remote sensing images: A survey." Chinese Journal of Aeronautics 34, no. 3 (2021): 145-163.
15. De, Arijit, Sri Krishna Kumar, Angappa Gunasekaran, and Manoj Kumar Tiwari. "Sustainable maritime inventory routing problem time window constraints." Engineering Applications of Artificial Intelligence 61 (2017): 77-95.
16. Escorcia-Gutierrez, José, Margarita Gamarra, Kelvin Beleño, Carlos Soto, and Romany F. Mansour. "Intelligent deep learning-enabled autonomous small ship detection and classification model." Computers and Electrical Engineering 100 (2022): 107871.
17. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016. 6. Li, Jianfeng, Zongfeng Li, Mingxu Chen, Yongling Wang, and Qinghua Luo. "A new ship detection algorithm in optical remote sensing images based on improved R3Det." Remote Sensing 14, no. 19 (2022): 5048.
18. Wang, Chien-Yao, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and IHau Yeh. "CSPNet: A new backbone that can enhance learning capability of CNN." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp. 390-391. 2020.
19. Wang, Chien-Yao, Hong-Yuan Mark Liao, and I-Hau Yeh. "Designing network design strategies through gradient path analysis." arXiv preprint arXiv:2211.04800 (2022).