

FEDERATED LEARNING FOR SHIP OBJECT DETECTION: A SURVEY OF METHODS, CHALLENGES, AND FUTURE DIRECTIONS

Vo Thanh Nhat Khanh^{1,*}, Duong Tan Manh², Huynh Minh Tri³, Nguyen Tan Y²

¹Interdisciplinary Research Transfer and Innovation Center, University of Phan Thiet, Vietnam

²Faculty of Engineering and Technology, University of Phan Thiet, Vietnam

³Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology and Engineering, Viet Nam

Abstract: Maritime domains are critical to economic development and international trade, while also playing a key role in the protection of national sovereignty. The exploitation of marine resources relies heavily on maritime vessels, which are central to transportation and offshore operations. For coastal regions frequently affected by natural disasters such as storms and tsunamis, there is a pressing need for reliable ship detection systems using aerial drones, satellites, radar, or surveillance cameras. Such systems enable timely identification and coordination of vessels at sea, facilitating evacuation to safe zones during extreme events. In addition, ship detection is essential for monitoring unauthorized maritime activities and illegal incursions into territorial waters. While recent studies have explored the application of Machine Learning, Deep Learning, and Federated Learning in maritime environments, a gap remains in integrating these approaches into unified frameworks. This work aims to review and analyze these methodologies for ship detection, with the goal of identifying suitable algorithms and models for practical deployment under diverse operational conditions.

Keyword: deep learning, federated learning, machine learning, maritime areas, object detection

1. INTRODUCTION

Federated Learning is a machine learning technique that differs from traditional machine learning. It allows devices or edge nodes to collaboratively train a common model without transferring their local data. In other words, Federated Learning distributes the global model to devices or edge nodes, enabling training to occur locally so that data remains on each respective device or node, thereby

ensuring security and privacy instead of centralizing data as in traditional machine learning methods. The server aggregates the local model updates using aggregation algorithms, such as FedAvg (McMahan et al., 2017), FedProx (Li et al., 2020), and FedOpt (Reddi et al., 2021), to update the global model, with FedAvg serving as one of the fundamental approaches in Federated Learning (McMahan et al., 2017). Total number of local samples at round t :

$$m_t = \sum_{k \in S_t} n_k \quad (1)$$

Global model update (Federated Averaging):

$$w_{t+1} = \sum_{k \in S_t} \frac{n_k}{m_t} w_{t+1}^k \quad (2)$$

where:

t : communication round

S_t : set of selected clients at round t

n_k : number of local training samples at client k

m_t : total number of samples from participating clients

w_{t+1}^k : local model of client k after local training

w_{t+1} : updated global model

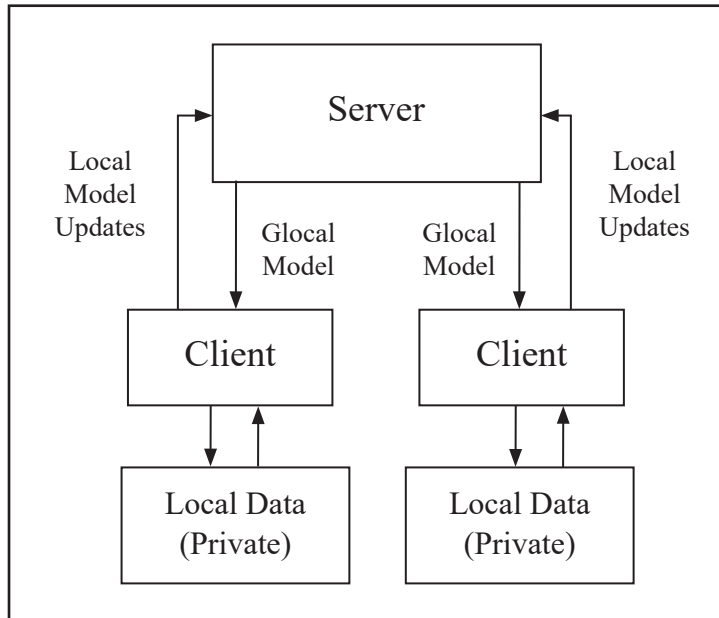


Figure 1. Flowchart of Federated Learning

Federated Learning (FL), introduced by McMahan et al. (2017), has been widely adopted across data-sensitive domains, including healthcare (Nguyen et al., 2022), (Joshi et al., 2022), transportation (Pandya et al., 2023), smartphones (Lim et al., 2020), finance (D'souza et al., 2023), and smart agriculture (Zhang et al., 2025). In the IoT sector, the global FL market is projected to grow from USD 127 million in 2023 to USD 210 million by 2028, corresponding to a CAGR of 10.6% (MarketsandMarkets, 2023).

In maritime applications, FL has been explored for ship object detection; however, challenges remain in terms of accuracy, system cost, and deployment under harsh environmental conditions with limited infrastructure. This paper reviews existing approaches and analyzes representative models to identify key challenges and practical limitations in applying FL to maritime object detection systems.

2. MACHINE LEARNING FOR SHIP OBJECT DETECTION

There are many studies on ML applications in object detection, and we have compiled the latest published works.

Aneesha Guna et al. developed a machine learning model that integrates multiple methods to achieve an increasing level of automation. The team created a new dataset by collecting videos of moving objects and labeling them using a combination of automated and manual techniques. The model was trained using You Only Look Once Version 4 (YOLOv4) and Region-Based Convolutional Neural Network (R-CNN) and deployed on Parrot Mambo drones for real-time object detection and tracking. The results were significant, achieving an average loss of 0.1942 and 96% accuracy (Guna et al., 2024).

Wang Yong et al. proposed a rotational multimodal ship object detection algorithm. Feature extraction is performed separately for optical data and Synthetic Aperture Radar (SAR) data, and an intermediate fusion strategy is adopted to combine

optical and SAR features for detection. The introduction of the Convolutional Block Attention Module (CBAM) into the backbone enables automatic weighting of feature channels during the learning process, thereby improving feature extraction efficiency and reducing noise and irrelevant information. The proposed multimodal detection algorithm was validated on the constructed QXS-SAROPT-SHIP dataset. Compared with conventional horizontal bounding box detection algorithms and single-model detection approaches, the proposed multimodal algorithm achieved significantly improved accuracy. Specifically, the accuracy increased by 1.73% with the inclusion of the CBAM module (Yong et al., 2024). Originally proposed by Woo et al., CBAM sequentially applies channel attention and spatial attention to refine feature representations by emphasizing informative features while suppressing irrelevant background information. The corresponding channel attention and spatial attention maps are mathematically formulated as follows (Woo et al., 2018):

Channel Attention

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (3)$$

where:

$F \in \mathbb{R}^{C \times H \times W}$: input feature map

$M_c(F) \in \mathbb{R}^{C \times 1 \times 1}$: channel attention map

$\text{AvgPool}(F)$: average-pooled feature descriptor

$\text{MaxPool}(F)$: max-pooled feature descriptor

MLP: shared multi-layer perceptron with one hidden layer

σ : sigmoid activation function

Spatial Attention

$$M_s(F) = \sigma(f^{7 \times 7}[\text{AvgPool}(F)(\text{MaxPool}(F))]) \quad (4)$$

where:

$M_s(F) \mathbb{R}^{1 \times H \times W}$: spatial attention map

$\text{AvgPool}(F)$: average-pooled feature map across the channel dimension

$\text{MaxPool}(F)$: max-pooled feature map across the channel dimension

$[\cdot]$: channel-wise concatenation.

$f^{7 \times 7}$: convolution operation with a 7 x 7 kernel

σ : sigmoid activation function

In maritime ship object detection, channel attention emphasizes ship-related feature channels, while spatial attention highlights informative regions and suppresses sea clutter, wave reflections, and other background interference.

Domenico Barretta et al. investigated the performance of binary machine learning classification models in terms of false alarm probability for detecting targets such as ships from satellite data. The study employed decision functions based on learned training data. The results showed that, under appropriate conditions, the models achieved improved accuracy, making them suitable for remote sensing applications, including maritime navigation (Barretta et al., 2024).

WenxinZhangetal.proposedPillarDANet, an improved 3D point cloud-based ship detection method for large vessels based on PillarNet. This method incorporates a Position-Aware Module (PAM) and a Channel-Aware Module (CAM) to enhance sensitivity to the geometric structure of ships. Experimental results demonstrated improvements of 3.1% and 3.9% in the mAP metric over PillarNet, with PillarDANet achieving 66.2% and 87.8%, respectively (Zhang and Ren, 2024).

Shuqi Sun et al. proposed a method to address noise in ship detection from Synthetic Aperture Radar (SAR) images by combining the localization capability

of the Constant False Alarm Rate (CFAR) algorithm with the classification strength of machine learning. Targets were detected and localized using a two-stage CFAR approach, followed by noise removal and extraction of candidate targets. A combination of Gaussian distribution modeling and global CFAR was applied. After obtaining candidate targets, a Gradient Boosting Decision Tree (GBDT) classifier was used for training and evaluation. Simulation results demonstrated 96.1% accuracy and strong detection performance (Sun et al., 2023).

The above studies demonstrate that machine learning techniques can reduce errors and improve accuracy in ship object detection, and these methods are summarized in Table 1.

3. DEEP LEARNING FOR SHIP OBJECT DETECTION

Deep learning is a subfield of machine learning driven by multilayer neural networks, designed with inspiration from the structure of the human brain, and widely applied in many fields, including maritime domains. Artificial neural networks are structured as multiple layers of interconnected computational nodes, where each node applies an activation function to transform its input into an output signal (IBM, n.d.).

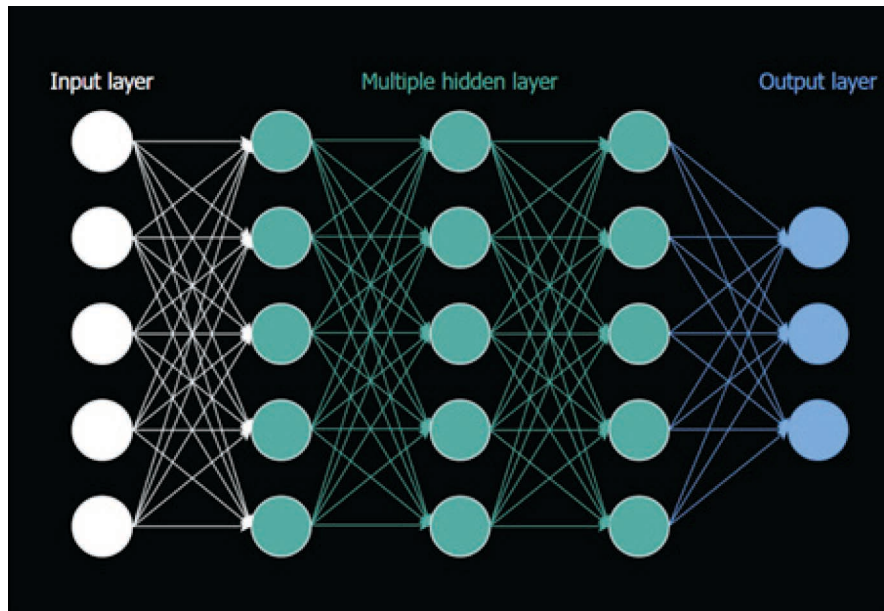


Figure 2. General architecture of a deep neural network

Lixin Zhang et al. conducted a deep learning-based ship detection study using Faster Region-Based Convolutional Neural Network (Faster R-CNN), YOLOv4, and Single Shot MultiBox Detector (SSD). The results showed that each algorithm has its own advantages and disadvantages depending on object size. For medium and large ships, Faster R-CNN achieved the highest accuracy, followed by YOLOv4, while SSD showed the lowest accuracy. In terms of detection speed, SSD was the fastest, followed by YOLOv4, and Faster R-CNN was the slowest. For small ships, YOLOv4 achieved the best detection accuracy, while SSD had the fastest detection speed. Faster R-CNN was neither as accurate nor as fast as the other two algorithms. Therefore, it is necessary to clearly define the objectives in order to select the appropriate algorithm (Zhang and Yin, 2022).

Shuxiao Cheng et al. proposed a deep learning model for ship detection from drone images using YOLOv5-ODConvNext. They incorporated multi-dimensional

convolution (ODConv) into YOLOv5 to improve accuracy. Experimental results showed that their model achieved 48% accuracy, exceeding YOLOv5s by 1.2%. The detection speed was 8.3 ms per image on an NVIDIA RTX3090 GPU, which is 13.3% faster than YOLOv5s (Cheng et al., 2023).

Esra Yildirim et al. conducted a comparative study of deep learning algorithms using State-of-the-Art (SOTA) datasets for ship detection. The performance of five models (Faster R-CNN, YOLOv8, CenterNet, EfficientDet, and SSD) was evaluated. The results indicated that YOLOv8 achieved the highest detection performance ($AP@0.50 = 96.33\%$), followed by CenterNet ($AP@0.50 = 88.80\%$), Faster R-CNN ($AP@0.50 = 85.30\%$), EfficientDet ($AP@0.50 = 78.91\%$), and SSD ($AP@0.50 = 75.98\%$) (Yildirim and Kavzoglu, 2025).

Chun-Ming Wu et al. developed the Ship_YOLO algorithm based on You Only Look Once version 11 (YOLOv11). As a

one-stage detection algorithm, YOLOv11 prioritizes speed, achieves lower missed detection and false positive rates, and is well-suited for real-time maritime monitoring. In contrast, two-stage detection algorithms and multimodal approaches offer higher accuracy but are unsuitable for real-time detection due to slow inference speeds (Badgujar et al., 2024). To further improve ship detection performance, Wu et al. introduced an Adaptive Multi-scale Fusion (AMF) structure to effectively integrate multi-scale features and proposed a Dual-Teacher Adaptive Dynamic Head (DTADH) to align adaptive detection boxes for various tasks under different lighting conditions. These components improve the model's robustness in challenging maritime environments by enhancing feature representation across different ship scales and reducing the influence of complex backgrounds and illumination variations, thereby enabling more reliable ship detection under low-light conditions and other harsh environmental scenarios. Experimental results demonstrated that Ship_YOLO outperformed baseline models on ship datasets, providing a cost-effective solution for improving future ship detection systems (Wu et al., 2025).

Research has focused on developing deep learning models based on YOLO architectures and noise-processing techniques to mitigate the effects of weather conditions and improve ship detection accuracy. These techniques and algorithms are summarized in Table 2.

3.1. Overview of CNNs (Convolutional Neural Networks)

Convolutional Neural Networks (CNNs) are a class of deep learning models designed to efficiently process structured grid data, particularly digital images. Due to their ability to automatically learn features directly from raw data, CNNs have become a fundamental approach for many computer vision tasks such as image classification, object detection, and image segmentation (Goodfellow et al., 2016).

A typical CNN architecture consists of convolutional layers, nonlinear activation functions, and pooling layers. Convolutional layers utilize filters learned during training to extract local features such as edges, textures, and shapes. The convolution operation is defined as follows (LeCun et al., 1998):

$$Y(i, j) = \sum_{c=1}^C \sum_{u=0}^{k_h-1} \sum_{v=0}^{k_w-1} K(u, v, c) X(i+u, j+v, c+b) \quad (5)$$

where:

X : denotes the input feature map

K : represents the learnable convolutional kernel

C : the number of input channels

b : the bias term

$Y(i, j)$: output feature value at position i, j

k_h, k_w : height and width of the convolution kernel

Nonlinear activation functions, most commonly ReLU, enable the network to learn complex representations and improve convergence speed during training (Nair and Hinton, 2010). The ReLU activation function is defined as:

$$f(x) = \max(0, x) \quad (6)$$

where:

x : input value

$f(x)$: output value after applying the activation function

Pooling layers are used to reduce the spatial dimensions of feature maps, thereby decreasing computational cost and increasing robustness to small spatial variations in the input (LeCun et al., 1998).

$$Y(i, j) = \max_{(u, v) \in \Omega} X(i + u, j + v) \quad (7)$$

where:

Ω : denotes a local pooling region

A key characteristic of CNNs is the combination of weight sharing and local connectivity. These properties significantly reduce the number of parameters compared to traditional fully connected networks while improving generalization capability (LeCun et al., 1998). As a result, CNNs demonstrate superior performance over classical machine learning methods in image processing tasks.

In modern object detection models such as YOLO, CNNs are not only used for feature extraction but also form the backbone of the entire network architecture. Components such as the backbone, neck, and head are built upon convolutional layers, enabling end-to-end object detection and supporting real-time processing requirements (Redmon et al., 2016).

3.2. Overview of YOLO

YOLO (You Only Look Once) is an object detection model based on Convolutional Neural Networks (CNNs), widely used for detecting objects in images and videos

in computer vision. Unlike traditional multi-stage object detection methods, YOLO performs detection in a single forward pass of the neural network, simultaneously predicting object locations and class labels (Redmon et al., 2016). YOLO divides the input image into an $S \times S$ grid, where each grid cell is responsible for detecting an object if its center falls within that cell. A CNN is used to extract features and directly predict bounding boxes and class probabilities, enabling YOLO to achieve high processing speed and real-time performance (Redmon et al., 2016).

4. FEDERATED LEARNING FOR SHIP OBJECT DETECTION

4.1. General Optimization Objective

The studies reviewed in this paper employ various object detection models as local models within the Federated Learning framework. Since different detection architectures adopt different

optimization strategies, the corresponding loss functions are not identical across all studies. Nevertheless, object detection is generally formulated as a multi-task optimization problem, where the overall objective consists of multiple task-specific loss components, such as localization, objectness, and classification losses (Chan et al., 2024). Within the Federated Learning framework, this detection loss serves as the sample-level objective for local model optimization at each client. The local objective is then defined as the

average detection loss over the client's local dataset, while the global objective is formulated as the weighted aggregation of all local objectives according to the data distribution among clients (McMahan et al., 2017). After local training, the global model is updated according to the adopted federated optimization strategy, such as FedAvg, FedProx, or FedOpt. Following the general object detection loss formulation presented by Chan et al. (Chan et al., 2024), L_{total} can be expressed as follows:

$$L_{\text{total}} = \lambda_{\text{box}} L_{\text{box}} + \lambda_{\text{obj}} L_{\text{obj}} + \lambda_{\text{class}} L_{\text{class}} \quad (8)$$

where:

L_{total} : overall optimization objective

L_{box} : localization loss for bounding box regression

L_{obj} : objectness loss used to determine whether an object is present in a predicted region

L_{class} : classification loss used to predict the object category

λ_{box} : weighting coefficient for the localization loss

λ_{obj} : weighting coefficient for the objectness loss

λ_{class} : weighting coefficient for the classification loss

Local Optimization Objective for FedAvg

Based on the detection loss defined above, the local optimization objective for client is formulated as follows (McMahan et al., 2017):

$$F_k(w) = \frac{1}{n_k} \sum_{i \in P_k} L_{\text{total}}(w; x_i, y_i) \quad (9)$$

where:

$F_k(w)$: local optimization objective of client k

w : model parameters

P_k : set of training samples belonging to client k

$n_k = |P_k|$: number of local training samples at client k

$L_{\text{total}}(w; x_i, y_i)$: detection loss calculated for sample i using model parameters w

x_i : input image

y_i : corresponding ground-truth label

Global Optimization Objective for FedAvg

The global optimization objective is then obtained by aggregating the local optimization objectives of all participating clients according to their local dataset sizes, following McMahan et al. (McMahan et al., 2017):

$$\min_w f(w) = \sum_{k=1}^K \frac{n_k}{n} F_k(w) \quad (10)$$

with:

$$n = \sum_{k=1}^K n_k \quad (11)$$

where:

$f(w)$: global optimization objective in the Federated Learning framework

K : total number of clients

n : total number of training samples from all clients

n_k/n : aggregation weight of client k

$F_k(w)$: local objective function of client k

Global Model Update for FedAvg

The global objective defined above represents the general optimization formulation in the Federated Learning framework. In practice, this objective can be optimized using different federated optimization strategies. The standard FedAvg approach aggregates locally optimized models based on the relative number of training samples at each client (McMahan et al., 2017), while other approaches, such as FedProx (Li et al., 2020) and FedOpt (Reddi et al., 2021), introduce additional mechanisms to improve robustness under heterogeneous and non-IID data distributions. These optimization strategies are discussed in detail in the following sections.

4.2. Federated Learning Optimization Algorithms

During the development of FL, federated optimization algorithms have continuously evolved to overcome the limitations of earlier

optimization approaches. Building upon FedAvg introduced in the previous section, several advanced optimization algorithms have been proposed to improve optimization efficiency, convergence stability, and model robustness under heterogeneous (non-IID) data distributions. Among them, FedProx and FedOpt are two representative optimization algorithms that enhance the federated training process from different perspectives, including local objective regularization and adaptive global model optimization. These advancements have significantly expanded the applicability of FL in real-world scenarios, including maritime ship detection, where client data are inherently distributed, heterogeneous, and non-IID.

Tian Li et al. proposed FedProx to address the challenges of statistical heterogeneity and system heterogeneity by introducing a proximal regularization term into the local objective function, while retaining the same server-side

aggregation strategy as FedAvg. The proximal term constrains local model updates to remain closer to the current global model, thereby reducing client drift and improving convergence under non-IID data distributions (Li et al., 2020). Therefore, the FedProx algorithm

is particularly well suited for maritime ship detection applications, where data collected from satellites, UAVs, ships, and coastal monitoring stations are inherently heterogeneous and exhibit non-IID characteristics. The corresponding local objective function is defined as follows:

$$h_k(w; w_t) = F_k(w) + \frac{\mu}{2} \|w - w_t\|^2 \quad (12)$$

where:

$h_k(w; w_t)$: local objective function of client k

$F_k(w)$: local loss function of client k

w : local model parameters

w_t : global model at communication round t

μ : proximal coefficient controlling the regularization strength

Sashank J. Reddi et al. subsequently proposed FedOpt, a generalized adaptive federated optimization framework that applies adaptive optimization algorithms on the server side, including FedAdagrad, FedAdam, and FedYogi, to improve convergence speed and optimization performance (Reddi et al., 2021). The global model is updated as follows:

$$x^{t+1} = \text{SERVEROPT}(x^t, -\Delta^t, \eta, t) \quad (13)$$

where:

x^t : global model at communication round t

x^{t+1} : updated global model at communication round $t+1$

Δ^t : average client update aggregated from the selected clients

η : server learning rate

SERVEROPT : server-side adaptive optimizer (FedAdagrad, FedAdam, or FedYogi)

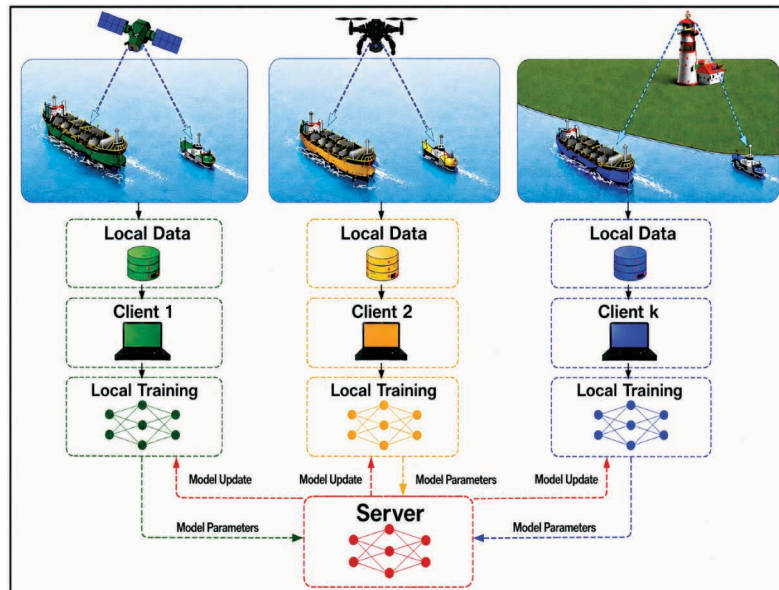


Figure 3. Federated Learning framework for ship object detection

4.3. Federated Learning for Ship Object Detection

Federated Learning has emerged as a promising approach for ship object detection by enabling maritime monitoring platforms to jointly train a model while preserving local data privacy.

Figure 3 illustrates the architecture of Federated Learning applied to ship object detection across different maritime surveillance devices, including satellites, UAVs, and coastal surveillance stations. Each device acts as a distributed client: the observed ship data are stored locally and used to perform local training of a neural network model directly on the client without transmitting the raw data. After each training round, clients send only the model updates to the central server; the server aggregates these updates into a global model and sends the resulting model parameters back to each client to continue the next training round. As a result, different maritime surveillance devices can collaboratively train a shared ship object detection model while the data

remain stored locally, thereby preserving data privacy.

Priyanka Nandal et al. proposed combining Federated Learning with YOLOv11 and a transformer-based attention module to improve real-time accuracy while ensuring data security through Federated Learning. YOLOv11 was adopted as the backbone because of its enhanced architecture for efficient multi-scale feature extraction. Specifically, the C3K2 block improves feature representation while reducing computational complexity, the Spatial Pyramid Pooling Fast (SPPF) module aggregates contextual information at multiple scales to enhance the detection of small ships, and the C2PSA block incorporates spatial attention to emphasize salient ship features while suppressing complex maritime backgrounds. In addition, the redesigned detection head integrates a multi-scale Feature Pyramid Network (FPN) with an anchor-free prediction mechanism, enabling more accurate detection of ships with different sizes

and orientations. The transformer-based attention mechanism further strengthens long-range spatial feature learning, allowing the model to better identify small or partially occluded vessels in challenging maritime environments (Nandal et al., 2025). The proposed method achieved high accuracy (95%), recall (90%), mAP50 of

92%, and mAP50-95 of 54.7%. Moreover, the communication overhead decreased from approximately 500 MB at round 1 to 380 MB at round 10, corresponding to a reduction of approximately 24%. The research team asserted that the model provides real-time inference speed and high applicability (Nandal et al., 2025).

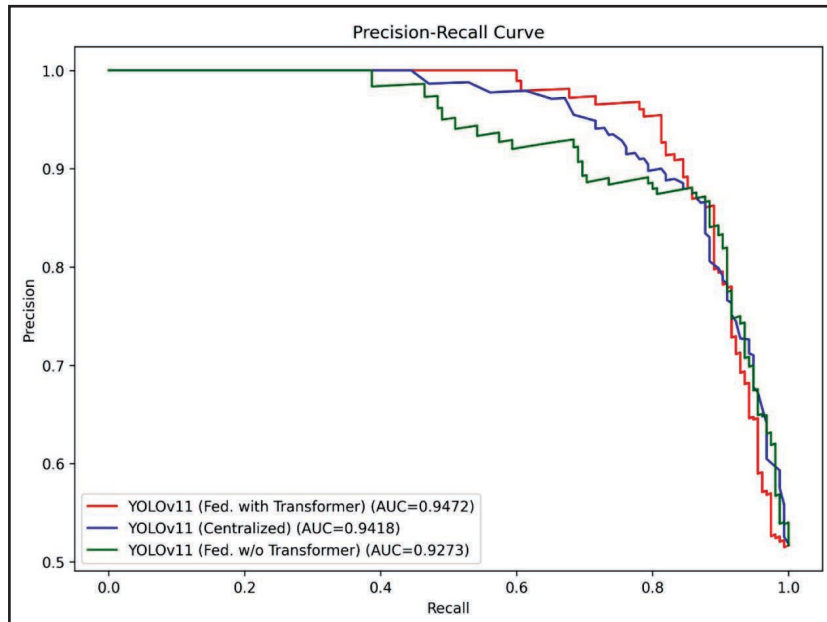


Figure 4. Precision-Recall curve (Nandal et al., 2025)

Figure 4 presents the Precision-Recall (P-R) curve reported by Nandal et al. (2025). The results demonstrate that the federated YOLOv11 model integrated with the transformer-based attention mechanism achieved the highest detection performance ($AUC = 0.9472$), slightly outperforming the centralized YOLOv11 model ($AUC = 0.9418$) and clearly surpassing the federated model without the transformer ($AUC = 0.9273$). The results

indicate that the attention mechanism effectively enhances feature representation, reduces false positives caused by complex maritime backgrounds, and improves the detection of small or partially occluded ships. Furthermore, the close performance between the federated and centralized models highlights the feasibility of privacy-preserving Federated Learning for maritime ship detection without sacrificing detection accuracy (Nandal et al., 2025).

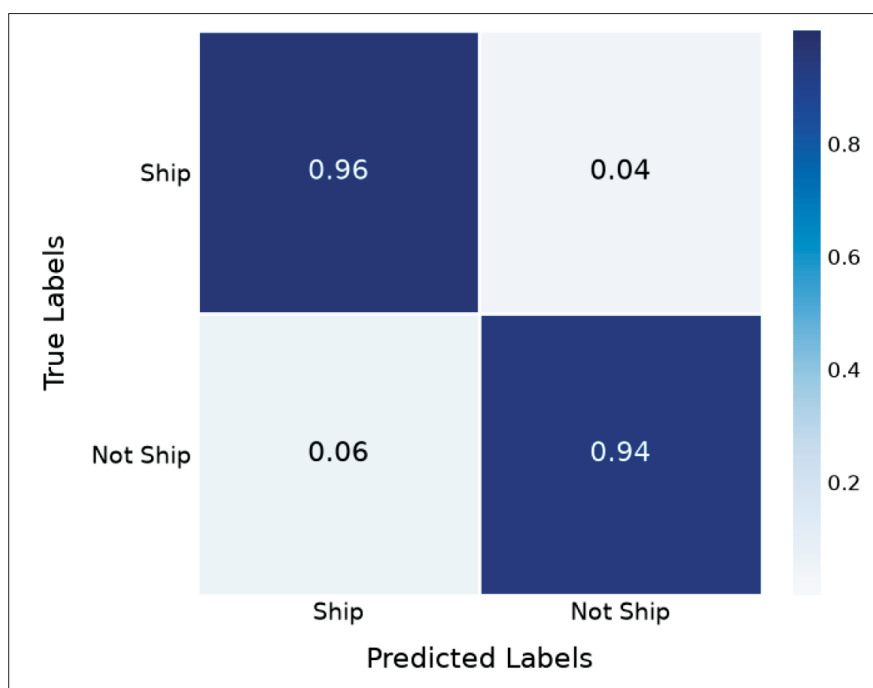


Figure 5. Confusion Matrix (Nandal et al., 2025)

Figure 5 presents the confusion matrix of the proposed model. The results show that the model achieves a true positive rate of 0.96 and a true negative rate of 0.94, while maintaining low false positive and false negative rates. These results demonstrate the model's strong ability to distinguish between classes and its high performance in ship detection. The Transformer-based attention mechanism enhances spatial feature representation and discrimination, thereby contributing to more accurate ship detection in complex maritime environments, particularly for small or densely distributed ships (Nandal et al., 2025).

Anh-Kiet Duong et al. proposed FedShip, a Federated Learning

framework for ship detection using multi-source optical satellite images. The study evaluated three representative federated optimization algorithms, namely FedAvg, FedProx, and FedOpt, on two benchmark datasets: Ship4FL for heterogeneous multi-source satellite imagery and ShipGE for homogeneous satellite imagery. The Ship4FL dataset comprises four subsets corresponding to four optical satellite sensors, namely PlanetScope, Sentinel-2A/B, Gaofen-1/6, and Landsat-8/9. In addition, a combined validation set, referred to as Group4, integrates the validation subsets from all four satellite sources to evaluate the overall generalization performance of the models across heterogeneous multi-source data.

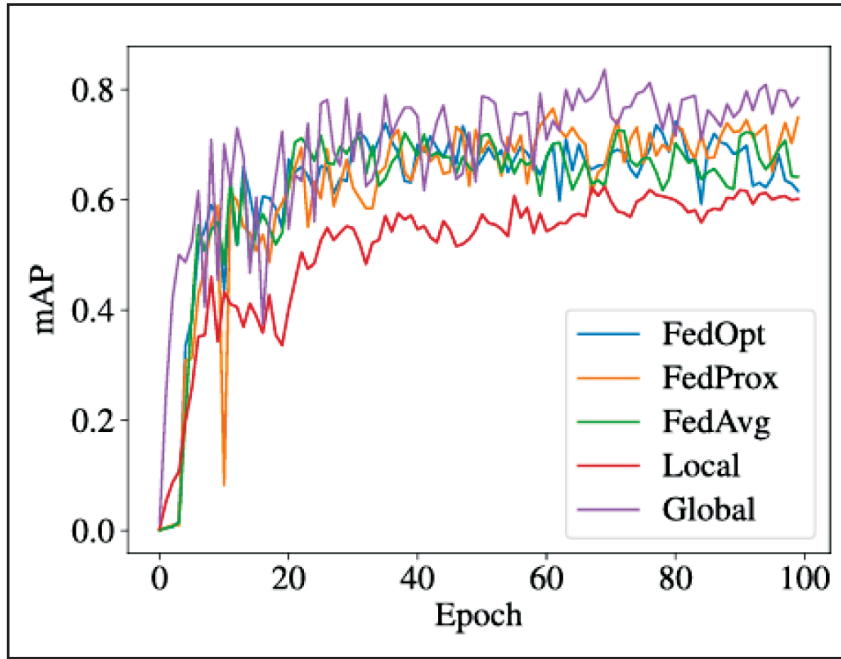


Figure 6. Performance of different training strategies on the Ship4FL dataset
Duong et al., 2025)

Figure 6 shows that the centralized training strategy achieved the highest detection performance throughout the training process, whereas local training consistently produced the lowest mAP50. Although the three Federated Learning algorithms did not reach the performance of centralized training, they substantially outperformed local training while preserving data privacy. Among the evaluated algorithms, FedProx exhibited slightly better performance than FedAvg and FedOpt during the later stages of training (after approximately 60 epochs), indicating that the proximal regularization term effectively mitigates client drift under heterogeneous (non-IID) data distributions. Quantitative results further supported these observations. The centralized model achieved the highest mAP50 on the Planet (0.903) and the combined Group4 (0.832) validation sets. Among the Federated Learning algorithms, FedOpt achieved the

highest mAP50 on the Gaofen-1/6 dataset (0.674), while FedAvg achieved the best performance on the Landsat-8 dataset (0.838), outperforming the centralized model on their respective validation subsets. Overall, the authors reported that the mAP50 of the FedAvg and FedProx algorithms was approximately 7%-10% lower than that of centralized training but 10%-15% higher than that of local training on average. These findings demonstrate that advanced federated aggregation algorithms can effectively improve ship detection performance on heterogeneous multi-source satellite datasets while maintaining the privacy-preserving advantages of Federated Learning (Duong et al., 2025).

To further assess how data distribution affects Federated Learning performance, the authors subsequently evaluated the same training strategies on the homogeneous ShipGE dataset.

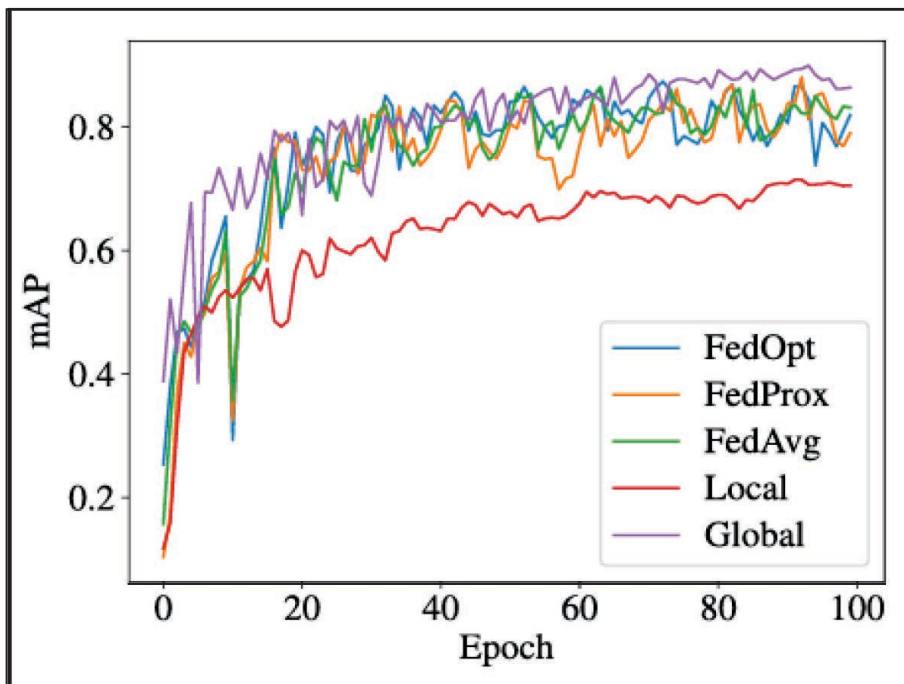


Figure 7. Performance of different training strategies on the ShipGE dataset
(Duong et al., 2025)

Figure 7 shows that the centralized training strategy again achieved the highest detection performance, while local training consistently produced the lowest mAP50. The three Federated Learning algorithms significantly outperformed local training and achieved performance very close to centralized training. Unlike the heterogeneous Ship4FL dataset, the performance differences among FedAvg, FedProx, and FedOpt became marginal, with their convergence curves closely overlapping throughout the training process. This behavior indicates that when client data are relatively homogeneous, the influence of the aggregation algorithm on convergence and detection performance is considerably reduced. Quantitative results further showed that the Federated

Learning models achieved mAP50 values of approximately 86.2%-87.4%, compared with 89.8% for centralized training, while improving detection performance over local training by approximately 10%-25%, depending on the client. These findings demonstrate that Federated Learning can achieve detection accuracy comparable to centralized training under homogeneous data distributions while preserving data privacy (Duong et al., 2025).

Tran-Vu La et al. investigated the application of Federated Learning for ship detection while ensuring privacy and security. The study evaluated and compared several FL models-FedAvg, FedProx, FedOpt, and FedMedian-where the YOLOv8 model was trained independently on each dataset for ship detection.

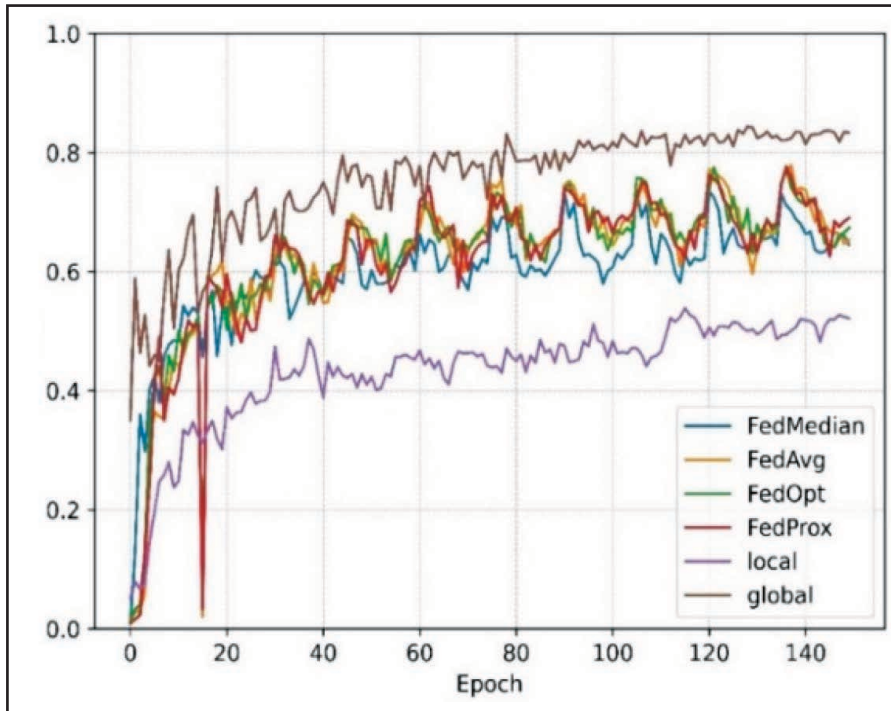


Figure 8. Performance of local and global training strategies (La et al., 2025)

Figure 8 shows that the global training strategy achieved the highest detection performance with an mAP of 0.843, whereas local training produced the lowest average mAP of 0.545 across the eight clients. Among the Federated Learning algorithms, FedAvg, FedProx, and FedOpt achieved comparable mAP values of approximately 0.743, while FedMedian exhibited slightly lower performance. Although the Federated Learning algorithms did not reach the accuracy of global training, they substantially outperformed local training, demonstrating the effectiveness of collaborative model learning without requiring raw data sharing.

Figure 9 illustrates the impact of different communication rounds and local training epoch configurations on Federated Learning performance, based on the quantitative results reported by authors (La et al., 2025). To further investigate

the effect of communication rounds, the authors evaluated four combinations of communication rounds (T) and local training epochs (E): Opt-1 (3×50), Opt-2 (5×30), Opt-3 (10×15), and Opt-4 (15×10). The results showed that Opt-4 achieved the highest average mAP of 0.748, whereas Opt-1 obtained the lowest average mAP of 0.692. At the dataset level, the highest detection performance was achieved on the Geosat-2 dataset with an mAP of 0.781 under the Opt-4 configuration, while the lowest mAP of 0.649 was observed on the Gaofen dataset under the Opt-1 configuration. However, although increasing the number of communication rounds from 3 to 15 improved detection performance, it also substantially increased the computation time, from 6,261 s to 22,700 s. Overall, the results suggest that Opt-3 provides the most balanced communication configuration by offering an effective trade-off between detection performance and computational

efficiency. These findings indicate that selecting an appropriate communication configuration is essential for balancing detection performance and computational efficiency in federated ship detection systems (La et al., 2025).

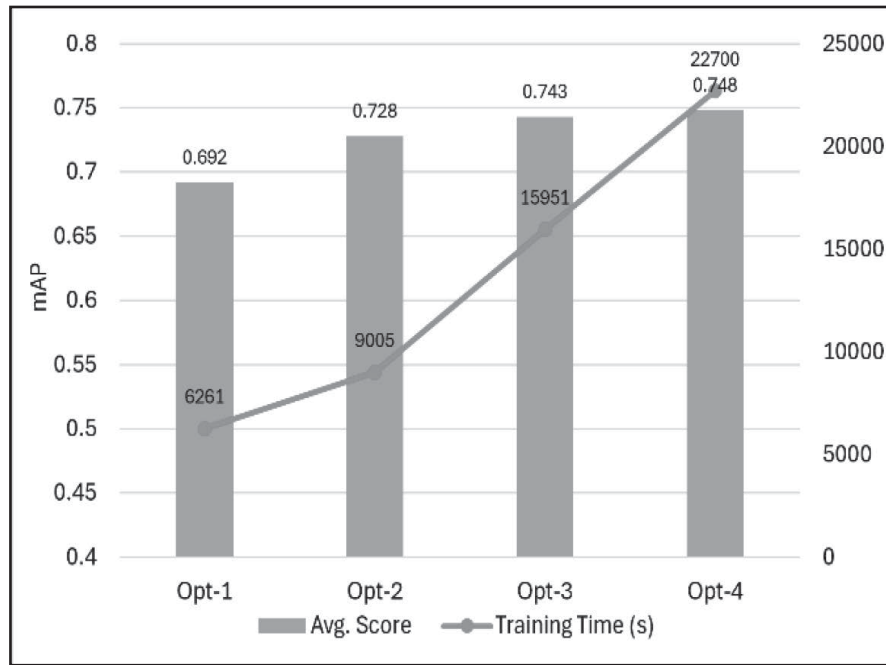


Figure 9. Average mAP and training time for different training strategies

YanhongHuang et al. proposed FLCSDet, a Federated Learning-based ship detection framework for maritime surveillance while preserving data privacy among maritime authorities. Instead of sharing raw data, the proposed framework enables collaborative model training by exchanging only model parameters. The FLCSDet architecture consists of two main components: the local ship detection model (CSDet) and the Global Adaptive Aggregation Model (GAM). Built upon YOLOv8, the CSDet model replaces the conventional C2F module with the CS-C2F module integrated

with the Efficient Multi-scale Attention (EMA) mechanism, enabling more effective multi-scale feature extraction and improving the detection of ships with various sizes under complex maritime conditions such as waves, reflections, and occlusions. In addition, the conventional CIoU loss is replaced with the dynamic non-monotonic Wise-IoU (WIoU) loss, which suppresses harmful gradients from low-quality anchor boxes to improve localization accuracy and accelerate convergence. The total loss function is defined as follows (Huang et al., 2025):

$$L = \omega_1 L_{\text{WIoU}} + \omega_2 L_{\text{cls}} + \omega_3 L_{\text{dfl}} \quad (14)$$

where:

L :Total loss function

L_{WIoU} : Wise-IoU loss for bounding-box regression

L_{cls} : Classification loss for predicting ship categories

L_{dfl} : Distribution Focal Loss

ω_1, ω_2 and ω_3 : Weight values for the WIoU loss, classification loss, and DFL

At the global aggregation stage, instead of employing simple averaging as in FedAvg, the authors proposed an adaptive aggregation strategy that assigns aggregation weights according to the effective data ratio of each client, thereby

balancing the generalization capability of the global model while preserving the characteristics of heterogeneous non-IID local datasets. The adaptive global aggregation strategy is given by (Huang et al., 2025):

$$\omega^t = \frac{1}{K} \sum_{k=1}^K \frac{n^{(k)}}{n} \omega_k \quad (15)$$

where:

$\omega^{(t)}$: global model parameters after round t

K : total number of participating clients

$n^{(k)}$: number of effective training images at client k

$n = \sum_{k=1}^K n^{(k)}$: total number of effective training images across all clients

ω_k : local model parameters uploaded by client k

Furthermore, transfer learning, pre-trained on the SeaShips and UAV datasets, was employed to accelerate local convergence and reduce communication rounds between clients and the server. Owing to the lack of a standardized benchmark for Federated Learning in maritime applications, the authors constructed three heterogeneous maritime datasets representing shore-based, UAV-mounted, and multi-device monitoring scenarios to simulate realistic non-IID data distributions across different maritime surveillance platforms (Huang et al., 2025). Based on the quantitative results reported by the authors, we visualized the results.

As illustrated in Figure 10, the mAP@0.5 performance of FLCSDet consistently improved as T increased across all three datasets. When T increased from 5 to 10, the detection performance improved substantially, with the mAP@0.5 increasing from 0.849 to 0.938 on the shore-based dataset, from 0.847 to 0.926 on the UAV-mounted dataset, and from 0.788 to 0.897 on the multi-device dataset. These results indicate that additional rounds of global model aggregation effectively enhanced collaborative learning and accelerated model convergence during the early stage of federated training.

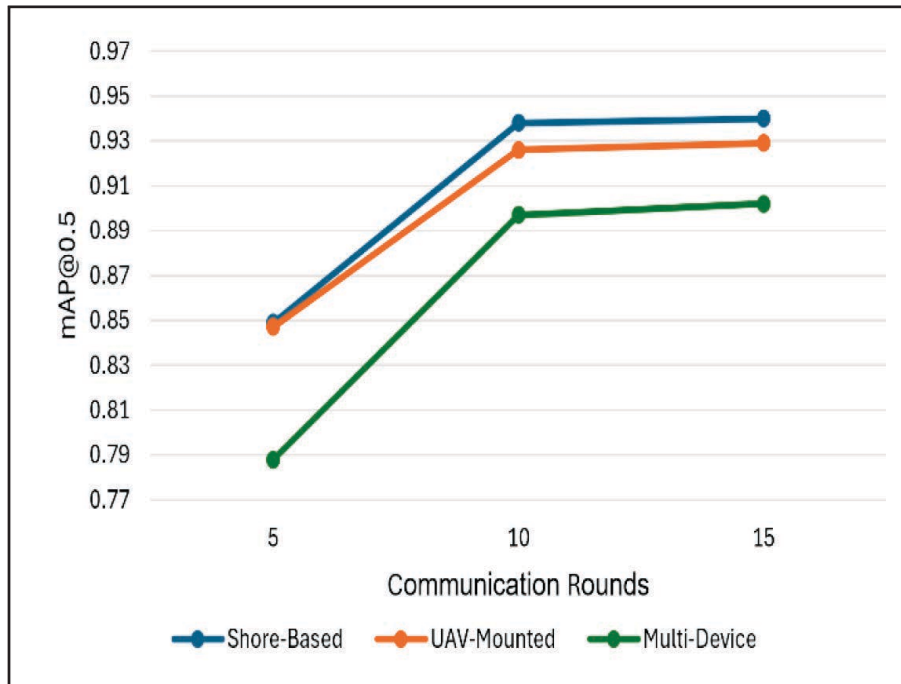


Figure 10. $mAP@0.5$ of FLCSDet under different communication rounds (T)

However, when T was further increased from 10 to 15, only marginal performance gains were observed, with improvements of 0.002, 0.003, and 0.005 on the shore-based, UAV-mounted, and multi-device datasets, respectively. This trend suggests that the FLCSDet model had nearly converged after approximately $T = 10$, and further increasing T provided only limited improvements while increasing communication and computational costs. Among the three datasets, the shore-based dataset consistently achieved the highest $mAP@0.5$ under all T configurations, followed by the UAV-mounted dataset, whereas the multi-device dataset consistently exhibited the lowest performance. The comparatively

lower performance of the multi-device dataset can be attributed to its higher degree of data heterogeneity arising from multiple sensing platforms, which makes collaborative optimization more challenging. Nevertheless, the consistent improvement observed across all three datasets demonstrates the robustness of the proposed adaptive aggregation strategy under heterogeneous data distributions (Huang et al., 2025).

The above studies demonstrate the potential of Federated Learning for ship object detection. Their methods, datasets, results, convergence rates, communication costs, and privacy levels are summarized in Tables 3 and 4.

Table 1. Machine Learning for Ship Object Detection

Ref	Methods / Models	Objective	Dataset	Results
Aneesha Guna et al.	YOLOv4, R-CNN; semi-auto labeling; UAV deployment	Real-time object detection on drones	Video data	Accuracy = 96%, Loss = 0.19
Wang Yong et al.	Multimodal Optical+SAR; CBAM	Improve multi-source ship detection	QXS-SAROPT-SHIP	AP = 79.84% (+1.73%)
Domenico Barretta et al.	Binary ML classifiers	Reduce false alarms	Satellite imagery	Improved accuracy estimation
Wenxin Zhang et al.	3D detection; PillarNet + PAM + CAM	Large-ship 3D detection	3D point clouds	mAP improved by 3.1% and 3.9% over PillarNet
Shuqi Sun et al.	Two-stage CFAR + GBDT	Reduce SAR noise	SAR images	Accuracy = 96.1%

Table 2. Deep Learning for Ship Object Detection

Ref	Methods / Models	Objective	Dataset	Results
Lixin Zhang et al.	Faster R-CNN, YOLOv4, SSD comparison	Analyze performance across ship sizes	Ship images of various sizes	Faster R-CNN best for medium/large; YOLOv4 best for small; SSD fastest; Faster R-CNN slowest
Shuxiao Cheng et al.	YOLOv5-ODConvNext (ODConv)	Improve drone-based ship detection	Drone imagery	AP = 48% (+1.2% vs YOLOv5s); Speed = 8.3 ms/img (13.3% faster)
Esra Yildirim et al.	Faster R-CNN, YOLOv8, CenterNet, EfficientDet, SSD	Benchmark SOTA models	SOTA ship detection datasets	YOLOv8 AP@0.50 = 96.33% (best); CenterNet = 88.80%; Faster R-CNN = 85.30%; EfficientDet = 78.91%; SSD = 75.98%
Chun-Ming Wu et al.	Ship_YOLO (YOLOv11), AMF, DTADH	Enhance multi-scale fusion & robustness	Ship detection datasets	Outperforms YOLOv11; better multi-scale and adaptive detection

Table 3. Federated Learning for Ship Object Detection

Ref	Methods / Models	Objective	Dataset	Results
Priyanka Nandal et al.	FL-YOLOv11 with Transformer-based attention	Real-time ship detection with data privacy	Distributed ship datasets	Accuracy = 95%, Recall = 90%, mAP50 = 92%, mAP50-95 = 54.7%
Anh-Kiet Duong et al.	FedShip with (FedAvg, FedProx, FedOpt)	Privacy-preserving ship detection on multi-source satellite imagery	Ship4FL, ShipGE	Ship4FL: FedAvg = 73.1%, FedProx = 76.8%, FedOpt = 75.1%; ShipGE: FedAvg = 86.2%, FedProx = 87.4%, FedOpt = 86.1% (mAP50)
Tran-Vu La et al.	FedAvg, FedProx, FedOpt, FedMedian with YOLOv8	Privacy-preserving ship detection on multi-source satellite imagery	8 satellite datasets	Global mAP = 0.843; Local = 0.545; FL $\approx$ 0.743
Yanhong Huang et al.	FLCSDet	Privacy-preserving ship detection under heterogeneous non-IID data	Shore-Based, UAV-Mounted, Multi-Device	mAP@0.5 at T = 10: Shore-Based 0.938, UAV-Mounted 0.926, Multi-Device 0.897

Table 4. Comparison of Communication Efficiency and Privacy in Federated Learning for Ship Detection Methods

Ref.	Methods / Models	Convergence Rate	Communication Cost	Privacy Level
Priyanka Nandal et al.	FL-YOLOv11 with Transformer-based Attention	10 rounds (95% accuracy)	$\approx 500 \rightarrow \approx 380$ MB (-24%)	Standard FL; raw data remain local
Anh-Kiet Duong et al.	FedShip (FedAvg, FedProx, FedOpt)	≈ 10 rounds	NR	Standard FL
Tran-Vu La et al.	FedAvg, FedProx, FedOpt, FedMedian with YOLOv8	≈ 10 rounds; T = 15 provides marginal improvement	NR	Standard FL
Yanhong Huang et al.	FLCSDet	≈ 10 rounds; marginal improvement beyond T = 10	NR	Standard FL

*NR: Not Reported

5. CHALLENGES IN FEDERATED LEARNING FOR SHIP OBJECT DETECTION

Some challenges in Federated Learning identified from the aforementioned studies are illustrated in Figure 11:

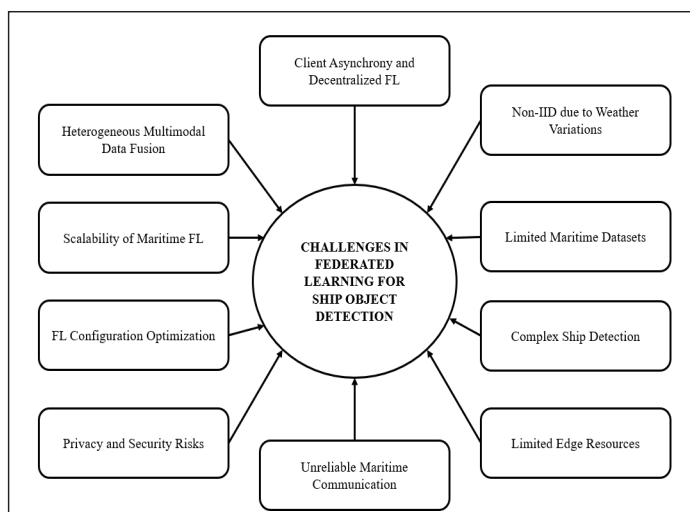


Figure 11. Challenges in Federated Learning for ship object detection

Different weather conditions and external environments lead to variations in the collected data, resulting in non-uniform data distribution (non-IID) across clients, which affects model accuracy and convergence (Kairouz et al., 2021), (Qi et al., 2023).

The lack of standardized datasets, as well as datasets covering diverse weather conditions, makes model training and evaluation challenging (Zhao et al., 2019).

In scenarios with multiple ships, varying orientations and rotations make it difficult to adjust bounding boxes accurately and avoid duplication (Zhao et al., 2024).

In maritime remote sensing imagery, ships often appear at extremely small scales, especially in satellite or high-altitude UAV data. The limited pixel representation, combined with dynamic sea backgrounds such as wave patterns and specular reflections, significantly degrades localization accuracy. Although small object detection is inherently challenging in centralized training frameworks, this issue is further exacerbated

in federated settings due to heterogeneous data distributions across clients (Yang et al., 2021), (Bochkovski et al., 2020).

Edge devices deployed on ships or offshore stations typically have limited CPU/GPU capacity, memory, and power supply. Training deep detection networks locally may result in excessive computation time and energy consumption, thereby limiting full client participation in FL rounds (Zhao et al., 2018).

In maritime environments, unstable bandwidth leads to high communication costs and significant delays (Li et al., 2020). Due to unstable maritime connectivity, some clients may fail to complete local training or upload updates. Client dropout leads to uneven contributions during aggregation and may degrade overall model performance (Yang et al., 2019).

Privacy and security remain significant challenges in Federated Learning because sharing model parameters or gradients does not completely eliminate the risk of sensitive information leakage. Adversaries may exploit

exchanged model updates through attacks such as model inversion or membership inference to infer information about local training data. Therefore, additional protection mechanisms, including Differential Privacy (DP), Secure Aggregation, Secure Multi-Party Computation (SMPC), and homomorphic encryption, may be required to strengthen privacy and security in maritime FL systems (Nandal et al., 2025).

FL configuration optimization is another important challenge because the selection of communication rounds and local training epochs directly affects both model performance and computational cost. Increasing the number of communication rounds may improve detection accuracy but can substantially increase training and communication overhead, while the performance improvement may become marginal after a certain point. Therefore, determining an appropriate FL configuration that balances detection performance and computational efficiency remains an important issue for practical maritime ship detection systems (La et al., 2025).

Scalability is an important consideration when deploying FL for large-scale maritime monitoring, where heterogeneous nodes such as satellites, ground stations, ships, and other edge devices may participate in collaborative training. As the number and diversity of participating clients increase, computational requirements and system coordination become more complex, particularly for resource-constrained nodes. Therefore, lightweight and computationally efficient model architectures are necessary to support scalable FL deployment across heterogeneous maritime networks (Nandal et al., 2025).

Multimodal data integration remains challenging because maritime surveillance systems can collect heterogeneous information from multiple sensors and platforms, including RGB imagery,

Automatic Identification System (AIS) data, radar, LiDAR, infrared cameras, and satellite imagery. Differences in data representations, sensing characteristics, and computational requirements make it difficult to effectively integrate these heterogeneous sources within a unified FL framework. Moreover, multimodal fusion can increase computational complexity and require more sophisticated fusion mechanisms, particularly when multiple satellite and sensor data sources are involved (Huang et al., 2024), (Duong et al., 2025).

Client asynchrony and delayed updates are significant challenges in maritime FL because unstable communication links may cause clients to disconnect or submit model updates at different times. Conventional synchronous FL may therefore suffer from delays because the aggregation process can depend on the availability of participating clients. Asynchronous FL can reduce this dependency by allowing model aggregation without requiring all clients to participate simultaneously, while decentralized approaches such as Peer-to-Peer FL and Swarm Learning may further reduce reliance on a central server. Nevertheless, efficiently handling disconnected clients, delayed updates, and system failures remains an important challenge for reliable maritime FL deployment (Nandal et al., 2025).

5.1. Future Research Directions for 2026–2027

Future research on Federated Learning for ship object detection is expected to be driven by several emerging research trends that address the current limitations identified in this survey, including non-IID data distribution, limited annotated maritime datasets, communication constraints, resource-limited edge devices, and the increasing demand for scalable and privacy-preserving maritime surveillance systems.

One of the most promising research trends is the integration of Vision Transformers into Federated Learning-based ship detection frameworks. Compared with conventional convolutional neural networks, Vision Transformers have demonstrated superior capability in capturing long-range dependencies and global contextual information, making them particularly suitable for detecting multi-scale and small ship targets in complex maritime environments. Lightweight Transformer architectures and hybrid CNN-Transformer models are expected to improve detection accuracy while maintaining computational efficiency on resource-constrained edge devices. In addition, the combination of Vision Transformers with self-supervised pretraining and few-shot learning is likely to reduce the dependence on large-scale annotated maritime datasets while enhancing model generalization across heterogeneous clients (Dosovitskiy et al., 2021).

Another emerging research trend is the adoption of Foundation Models in Federated Learning. Large-scale vision foundation models pretrained on diverse image datasets have shown remarkable transferability across downstream computer vision tasks and can substantially reduce the need for extensive labeled maritime data. Parameter-efficient adaptation methods, including adapters and Low-Rank Adaptation, are expected to facilitate the efficient fine-tuning of foundation models on heterogeneous maritime datasets while preserving data privacy. Furthermore, multimodal foundation models capable of jointly processing optical imagery, SAR, AIS, radar, infrared imagery, and other sensing modalities are likely to improve maritime situational awareness and ship detection performance through richer feature representation and cross-modal information fusion (Hatfaludi and Serban, 2025).

Self-supervised Federated Learning is also emerging as an important research

direction for maritime applications, where acquiring large-scale annotated ship images remains costly and time-consuming. Learning discriminative feature representations from abundant unlabeled data collected by satellites, UAVs, ships, and coastal surveillance systems is expected to substantially reduce annotation requirements before collaborative model optimization. Meanwhile, adaptive aggregation algorithms, personalized Federated Learning, and communication-efficient optimization strategies are likely to receive increasing research attention because of their potential to improve convergence, robustness, and adaptability in highly heterogeneous non-IID maritime environments while reducing communication overhead (Hatfaludi and Serban, 2025),(Wei et al., 2026).

Edge AI is anticipated to become a key enabling technology for the practical deployment of Federated Learning-based maritime surveillance systems. Lightweight object detection networks, together with model compression, pruning, quantization, and knowledge distillation techniques, are expected to enable real-time inference on computationally constrained platforms, including UAVs, autonomous vessels, satellites, and coastal edge devices. In addition, hardware-aware optimization and energy-efficient model architectures are likely to play an increasingly important role in balancing detection accuracy, computational cost, and energy consumption, thereby supporting long-term autonomous maritime operations (Wei et al., 2026).

The convergence of Vision Transformers, Foundation Models, Self-supervised Federated Learning, and Edge AI is expected to significantly advance the development of Federated Learning for ship object detection. Together with more efficient aggregation algorithms, privacy-preserving mechanisms, and communication-efficient protocols,

these emerging research trends have the potential to improve model robustness, scalability, and deployment capability, thereby facilitating the next generation of intelligent and privacy-preserving maritime surveillance systems.

6. CONCLUSION

6.1. Conclusion

This paper focuses on the broad potential of Federated Learning (FL) to improve ship detection and maritime surveillance. By enabling models to be trained collaboratively on distributed datasets from satellites, Unmanned Aerial Vehicles (UAVs), and radar sensors, FL overcomes limitations related to privacy, bandwidth, and data sharing, while maintaining high detection accuracy.

In this study, we reviewed and compared a range of deep learning models, including Faster R-CNN, YOLO (versions v4, v5, v8, v11), SSD, CenterNet, EfficientDet, and PillarDANet, as well as models incorporating attention mechanisms. We evaluated the advantages and limitations of each model based on factors such as ship size, observation conditions, and processing speed.

By integrating FL with these models, particularly through attention-based feature extraction, the capability to combine data from multiple sources is significantly enhanced, resulting in more effective ship detection under challenging conditions such as nighttime, dense fog, or rough seas.

For future research, practical implementation, collaboration among relevant agencies, and the development of flexible FL strategies suitable for diverse maritime environments should be prioritized. The combination of FL and deep learning models provides a scalable, privacy-preserving, and efficient solution for search and rescue operations, enabling real-time ship detection and contributing to improved safety, security, and efficiency in maritime operations.

6.2. Recommendations

Improving the quality and diversity of data collected from participating devices or systems is important for enhancing the robustness and generalization capability of the global model. Maritime datasets obtained from different sources, such as satellites, coastal monitoring systems, and radar sensors, should be combined to better represent diverse environmental conditions, including weather, lighting, and sea states.

Another important issue in Federated Learning is data heterogeneity among participating devices. Data collected from different maritime regions may follow different distributions, which can affect the performance of the global model. Therefore, exploring more advanced model aggregation methods or personalized Federated Learning approaches can help improve model performance when dealing with non-independent and identically distributed (non-IID) data.

Communication efficiency between participating devices and the central server should also be considered. Methods such as model compression, parameter pruning, or adaptive communication strategies can help reduce communication costs and accelerate the training process. This is particularly important in distributed maritime surveillance systems where network resources may be limited.

In addition, strengthening security mechanisms and protecting data privacy are essential when deploying Federated Learning frameworks. Techniques such as secure aggregation and differential privacy can help safeguard sensitive maritime surveillance data while still enabling collaborative model training among different participants.

Acknowledgments: *We would like to express our sincere gratitude to the reviewers for their valuable comments and constructive suggestions, which contributed significantly to improving the overall quality.*

Author Information:

Vo Thanh Nhat Khanh (*Corresponding author), Interdisciplinary Research Transfer and Innovation Center, University of Phan Thiet, Lam Dong Province, Vietnam

Email: vothanhnhatkhanh@gmail.com

Duong Tan Manh, University of Phan Thiet, Lam Dong Province, Vietnam

Email: dtmanh@upt.edu.vn

Huynh Minh Tri, Ho Chi Minh City University of Technology and Engineering, Ho Chi Minh City, Viet Nam

Email: huynhminhtri643@gmail.com

Dr. Nguyen Tan Y, University of Phan Thiet, Lam Dong Province, Vietnam

Email: nty@upt.edu.vn

Author Contributions:

Vo Thanh Nhat Khanh: Conceptualization, methodology, literature search and selection, literature synthesis and analysis, visualization, preparation of the original manuscript, and manuscript review and revision.

Duong Tan Manh: Assistance with literature search.

Huynh Minh Tri: Assistance with literature search.

Dr. Nguyen Tan Y: Conceptualization and provision of comments and feedback on the manuscript.

Research and Publication Ethics & Artificial Intelligence (AI) Transparency Statement: If accepted, the author is kindly requested to tick the box

☒ The author(s) confirm(s) that they have read, understood, and agreed to comply with the Regulations on Ethics in Research and Academic Publishing of the University of Phan Thiet Journal of Science (UPTJS).

AI use disclosure: ☐ No AI tools were used.

☒ AI tools were used — Tool/platform:ChatGPT.....; model/version:..... GPT-5.6; purpose/use:... Assistance with image generation and refinement of academic language ...

☒ The author(s) have reviewed and take full responsibility for all AI-assisted content; AI tools cannot be credited as authors or co-authors.

Article Information:

Received: April 12, 2026

Revised: August 17, 2026

Accepted: September 06, 2026

Note

The authors declare no competing interests regarding this article.

REFERENCES

- Badgujar, C. M., Poulose, A., & Gan, H. (2024). *Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review*. Computers and Electronics in Agriculture, 223, 109090. doi: 10.1016/j.compag.2024.109090.
- Barretta, D., Millefiori, L. M., & Braca, P. (2024). *Analytical Classification Performance Analysis Of Machine-Learning-Based Ship Detection From Optical Satellite Imagery*. IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, 8861-8865. doi: 10.1109/IGARSS53475.2024.10642217.
- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *YOLOv4: Optimal Speed and Accuracy of Object Detection*. arXiv. doi: 10.48550/arXiv.2004.10934.
- Chan, S., Wu, B., Zhang, G., Yao, Y., & Wang, H. (2024). *Learning Discriminatory Information for Object Detection on Urine Sediment Image*. Computer Modeling in Engineering & Sciences, 138(1), 411-428. doi: 10.32604/cmcs.2023.029485.
- Cheng, S., Zhu, Y., & Wu, S. (2023). *Deep learning based efficient ship detection from drone-captured images for maritime surveillance*. Ocean Engineering, 285(2), 115440. doi: 10.1016/j.oceaneng.2023.115440.
- D'Souza, K., Puthusseri, S., & Samuel, A. G. (2023). *Scalable Federated Learning for Privacy-Preserving Credit Card Fraud Detection*. 2023 IEEE International Carnahan Conference on Security Technology (ICCST), 1-6. doi: 10.1109/ICCST59048.2023.10726848.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale*. International Conference on Learning Representations (ICLR).
- Duong, A. K., La, T. V., Lê, H.-Â., & Pham, M.-T. (2025). *FedShip: Federated Learning for Ship Detection From Multi-Source Satellite Images*. IEEE Geoscience and Remote Sensing Letters, 22, 1-5. doi: 10.1109/LGRS.2024.3511122.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA, USA: MIT Press.
- Guna, A., Ganeriwala, P., & Bhattacharyya, S. (2024). *Exploring Machine Learning Engineering for Object Detection and Tracking by Unmanned Aerial Vehicle (UAV)*. 2024 International Conference on Machine Learning and Applications (ICMLA), 1001–1004. doi: 10.1109/ICMLA61862.2024.00149.
- Hatfaludi, C.-A., & Serban, A. (2025). *Foundational Models and Federated Learning: Survey, Taxonomy, Challenges and Practical Insights*. PeerJ Computer Science, 11, e2993. doi: 10.7717/peerj-cs.2993.
- Huang, Y., Liu, W., Lin, Y., Kang, J., Zhu, F., & Wang, F.-Y. (2025). *FLCSDet: Federated Learning-Driven Cross-Spatial Vessel Detection for Maritime Surveillance With Privacy Preservation*. IEEE Transactions on Intelligent Transportation Systems, 26(1), 1177-1192. doi: 10.1109/TITS.2024.3488497.
- IBM. (n.d.). *What is deep learning?* IBM Think. Retrieved December 01, 2025, from <https://www.ibm.com/think/topics/deep-learning>
- Joshi, M., Pal, A., & Sankarasubbu, M. (2022). *Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges*. ACM Transactions on Computing for Healthcare, 3(4), 40. doi: 10.1145/3533708.

- Kairouz, P., & McMahan, H. B. (2021). *Advances and Open Problems in Federated Learning*. Foundations and Trends in Machine Learning, 14(1-2), 1-210. doi: 10.1561/22000000083.
- La, T.-V., Pham, M.-T., Li, Y., Matgen, P., & Chini, M. (2025). *Adaptive Federated Learning for Ship Detection Across Diverse Satellite Imagery Sources*. IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium, 6596-6600. doi: 10.1109/IGARSS55030.2025.11243906.
- Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). *Federated Learning: Challenges, Methods, and Future Directions*. IEEE Signal Processing Magazine, 37(3), 50-60. doi: 10.1109/MSP.2020.2975749.
- Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). *Federated Optimization in Heterogeneous Networks*. Proceedings of Machine Learning and Systems (MLSys). doi: 10.48550/arXiv.1812.06127.
- Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y.-C., Yang, Q., Niyato, D., & Miao, C. (2020). *Federated Learning in Mobile Edge Networks: A Comprehensive Survey*. IEEE Communications Surveys & Tutorials, 22(3), 2031-2063. doi: 10.1109/COMST.2020.2986024.
- MarketsandMarkets. (2023). *Federated Learning Solutions Market*. MarketsandMarkets. Retrieved November 29, 2025, from <https://www.marketsandmarkets.com/Market-Reports/federatedlearning-solutions-market-151896843.html>.
- McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). *Communication-Efficient Learning of Deep Networks from Decentralized Data*. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), JMLR: W&CP, 54. doi: 10.48550/arXiv.1602.05629.
- Nair, V., & Hinton, G. E. (2010). *Rectified linear units improve restricted boltzmann machines*. Proceedings of the 27th International Conference on Machine Learning (ICML), 807-814.
- Nandal, P., Bohra, N., Mann, P., & Das, N. N. (2025). *YOLOv11 with transformer attention for real-time monitoring of ships: A federated learning approach for maritime surveillance*. Results in Engineering, 27, 106297. doi: 10.1016/j.rineng.2025.106297.
- Nguyen, D. C., Pham, Q.-V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., Dobre, O., & Hwang, W.-J. (2022). *Federated Learning for Smart Healthcare: A Survey*. ACM Computing Surveys, 55(3), 60. doi: 10.1145/3501296.
- Pandya, S., Srivastava, G., Jhaveri, R., Babu, M. R., Bhattacharya, S., Maddikunta, P. K. R., Mastorakis, S., Piran, M. J., & Gadekallu, T. R. (2023). *Federated learning for smart cities: A comprehensive survey*. Sustainable Energy Technologies and Assessments, 55, 102987. doi: 10.1016/j.seta.2022.102987.
- Qi, Z., Meng, L., Chen, Z., Hu, H., Lin, H., & Meng, X. (2023). *Cross-Silo Prototypical Calibration for Federated Learning with Non-IID Data*. Proceedings of the 31st ACM International Conference on Multimedia, 3099-3107. doi: 10.1145/3581783.3612481.
- Reddi, S., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Konečný, J., Kumar, S., & McMahan, H. B. (2021). *Adaptive Federated Optimization*. arXiv preprint arXiv:2003.00295. doi: 10.48550/arXiv.2003.00295.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-Time Object Detection*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788. doi: 10.1109/CVPR.2016.91.

- Sun, S., Wang, J., & Tan, D. (2023). *Ship Detection in SAR Images Based on CFAR and Machine Learning*. 2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 1-6. doi: 10.1109/CISP-BMEI60920.2023.10373385.
- Wei, H., Lee, D.Y., Rohal, S. et al. (2026). *A survey of foundation models for IoT: taxonomy and criteria-based analysis*. CCF Trans. Pervasive Comp. Interact. 8, 1-29. doi: 10.1007/s42486-025-00208-w.
- Woo, S., Park, J., Lee, J.Y., & Kweon, I.S. (2018). *CBAM: Convolutional Block Attention Module*. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11211. Springer, Cham. doi: 10.1007/978-3-030-01234-2_1.
- Wu, C.-M., Lei, J., Li, Z.-Q., & Ren, M.-L. (2025). *Ship_YOLO: General ship detection based on mixed distillation and dynamic task-aligned detection head*. Ocean Engineering, 323, 120616. doi: 10.1016/j.oceaneng.2025.120616.
- Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). *Federated Machine Learning: Concept and Applications*. ACM Trans. Intell. Syst. Technol., 10(2), 12. doi: 10.1145/3298981.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 86(11), 2278-2324. doi: 10.1109/5.726791.
- Yang, X., Yan, J., Ming, Q., Wang, W., Zhang, X., & Tian, Q. (2021). *Rethinking Rotated Object Detection with Gaussian Wasserstein Distance Loss*. Proceedings of the 38th International Conference on Machine Learning, 139, 11830-11841.
- Yildirim, E., & Kavzoglu, T. (2025). *Deep convolutional neural networks for ship detection using refined DOTA and TGRS-HRRSD high-resolution image datasets*. Advances in Space Research, 75(2), 1871-1887. doi: 10.1016/j.asr.2024.10.028.
- Yong, W., Ling, L., Xiao, T., & Bing, J. (2024). *Multi-Modal Ship Object Detection in Optical and SAR Images*. 2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA), Shenyang, China, 1321-1325. doi: 10.1109/ICIPCA61593.2024.10709219.
- Zhang, L., & Yin, H. (2022). *Research on Ship Detection Method of Optical Remote Sensing Image Based on Deep Learning*. 2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), 1-6. doi: 10.1109/ICSMD57530.2022.10058312.
- Zhang, S., Han, Q., Wang, H., Liu, J., & Li, B. (2025). *Federated learning with dual dynamic quantization optimization in smart agriculture*. Internet of Things, 34, 101798. doi: 10.1016/j.iot.2025.101798.
- Zhao, T., Wang, Y., Li, Z., Gao, Y., Chen, C., Feng, H., & Zhao, Z. (2024). *Ship Detection with Deep Learning in Optical Remote-Sensing Images: A Survey of Challenges and Advances*. Remote Sensing, 16(7), 1145. doi: 10.3390/rs16071145.
- Zhang, W., & Ren, M. (2024). *PillarDANet: An Improved PillarNet with Dual-Attention Mechanism for Ship Detection*. 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), 1605-1610. doi: 10.1109/ICICML63543.2024.10958078.
- Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). *Federated Learning with Non-IID Data*. arXiv. doi: 10.48550/ARXIV.1806.00582.
- Zhao, Z.-Q., Zheng, P., Xu, S.-T., & Wu, X. (2019). *Object Detection With Deep Learning: A Review*. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212-3232. doi: 10.1109/TNNLS.2018.2876865.

HỌC LIÊN KẾT ĐỂ PHÁT HIỆN ĐỐI TƯỢNG TÀU: KHẢO SÁT CÁC PHƯƠNG PHÁP, THÁCH THỨC VÀ HƯỚNG ĐI TRONG TƯƠNG LAI

Võ Thanh Nhật Khánh^{1,*}, Dương Tấn Mạnh², Huỳnh Minh Trí³, Nguyễn Tấn Ý²

¹Trung tâm Nghiên cứu Chuyển giao và Đổi mới sáng tạo liên ngành, Đại học Phan Thiết, Việt Nam

²Khoa Kỹ thuật Công nghệ, Đại học Phan Thiết, Việt Nam

³Khoa cơ khí chế tạo máy, Đại học Công nghệ Kỹ thuật Thành phố Hồ Chí Minh, Việt Nam

Tóm tắt: Lĩnh vực hàng hải rất quan trọng đối với phát triển kinh tế và thương mại quốc tế, đồng thời đóng vai trò quan trọng trong việc bảo vệ chủ quyền quốc gia. Việc khai thác tài nguyên biển phụ thuộc nhiều vào tàu biển, là trung tâm của vận tải và hoạt động ngoài khơi. Đối với các vùng ven biển thường xuyên bị ảnh hưởng bởi thiên tai như bão và sóng thần, nhu cầu cấp thiết về hệ thống phát hiện tàu đáng tin cậy sử dụng máy bay không người lái, vệ tinh, radar hoặc camera giám sát. Các hệ thống như vậy cho phép xác định và phối hợp kịp thời các tàu trên biển, tạo điều kiện thuận lợi cho việc sơ tán đến các khu vực an toàn trong các sự kiện cực đoan. Ngoài ra, phát hiện tàu là điều cần thiết để giám sát các hoạt động hàng hải trái phép và xâm nhập bất hợp pháp vào lãnh hải. Mặc dù các nghiên cứu gần đây đã khám phá ứng dụng của học máy, học sâu và học liên kết trong môi trường hàng hải, nhưng vẫn còn một lỗ hổng trong việc tích hợp các phương pháp này vào các khuôn khổ thống nhất. Bài báo này nhằm mục đích xem xét và phân tích các phương pháp phát hiện tàu, với mục tiêu xác định các thuật toán và mô hình phù hợp để triển khai thực tế trong các điều kiện hoạt động đa dạng.

Từ khóa: học liên kết, học máy, học sâu, khu vực hàng hải, phát hiện đối tượng

Thông tin tác giả:

Võ Thanh Nhật Khánh (*Tác giả liên hệ), Trung tâm Nghiên cứu Chuyển giao và Đổi mới sáng tạo liên ngành, Trường Đại học Phan Thiết, tỉnh Lâm Đồng, Việt Nam

Email: vothanhnhatkhanh@gmail.com

Dương Tấn Mạnh, Trường Đại học Phan Thiết, tỉnh Lâm Đồng, Việt Nam

Email: dtmanh@upt.edu.vn

Huỳnh Minh Trí, Trường Đại học Công nghệ Kỹ thuật TP. Hồ Chí Minh, TP. Hồ Chí Minh, Việt Nam

Email: huynhminhtri643@gmail.com

TS. Nguyễn Tấn Ý, Trường Đại học Phan Thiết, tỉnh Lâm Đồng, Việt Nam

Email: nty@upt.edu.vn

Ghi chú

Các tác giả xác nhận không có tranh chấp về lợi ích đối với bài báo này.