Journal of Field Robotics

WILEY

RESEARCH ARTICLE OPEN ACCESS

Prioritized Real-Time UAV-Based Vessel Detection for Efficient Maritime Search

Lyes Saad Saoud[1] | Zikai Jia[2] | Siyuan Yang[2] | Muhayy Ud Din[1] | Lakmal Seneviratne[1] | Shaoming He[1] | Irfan Hussain[1]

[1]Khalifa University Center for Autonomous Robotic Systems (KUCARS), Khalifa University, Abu Dhabi, United Arab Emirates | [2]School of Aerospace Engineering, Beijing Institute of Technology, Beijing, China

Correspondence: Shaoming He (shaoming.he@bit.edu.cn) | Irfan Hussain (irfan.hussain@ku.ac.ae)

Received: 14 June 2024 | **Revised:** 9 July 2025 | **Accepted:** 9 August 2025

Funding: This study is supported by the Khalifa University under Award No. RC1-2018-KUCARS-8474000136, CIRA-2021-085, MBZIRC-8434000194, KU-BIT-Joint-Lab-8434000534, and by Beijing Institute of Technology under National Key Research and Development Program Grant No. 2022YFE0204400.

ABSTRACT

Real-time vessel detection in maritime environments is crucial for diverse applications requiring speed and accuracy. Static camera views often introduce blind spots, compromising detection efficiency. This paper proposes a novel, real-time UAV-based system that uses a dynamic camera control strategy to address this limitation. This strategy leverages pre-defined search patterns, historical data (if available), and real-time sensor information (e.g., radar or LiDAR) to dynamically adjust the UAV's camera gimbal angles. This ensures comprehensive search area coverage while minimizing the risk of undetected vessels. Beyond dynamic camera control, our system incorporates a unique feature-based prioritization scheme for real-time target vessel identification. This scheme analyzes features extracted from captured images, including object size and shape. Additionally, movement analysis helps distinguish stationary objects from potential vessels. The combined approach of dynamic camera control and feature-based prioritization offers significant advantages. Firstly, it enhances search efficiency by systematically scanning the area and prioritizing promising candidates based on dynamic camera adjustments and feature analysis. Secondly, it improves detection accuracy by employing feature similarity (cosine similarity with a reference vessel stored in the system using a ResNet50 module) to reduce false positives and expedite target identification, especially in scenarios with multiple vessels. A comprehensive evaluation process has been conducted to validate the effectiveness of our proposed system in diverse simulated and real-world environments encompassing various conditions (weather, traffic density, background clutter). The results from this evaluation are highly promising and suggest the system's strong potential for real-time vessel detection in maritime environments.

1 | Introduction

Real-time vessel detection using unmanned aerial vehicles (UAVs) has become a crucial technology for various applications, including maritime security, environmental monitoring, and search and rescue operations (Marques et al. 2021). This capability enables continuous surveillance over vast maritime areas, ensuring timely responses to critical situations. However, achieving effective real-time vessel detection with UAVs remains challenging due to environmental conditions, dynamic target movement, and system constraints. These challenges are particularly evident in high-stakes competitions such as the Mohamed Bin Zayed International Robotics Challenge (MBZIRC) Maritime Grand Challenge (Xiu et al. 2019; Silva et al. 2016).

To address these challenges, the proposed system employs a UAV equipped with dynamic camera control to systematically scan the task area while continuously detecting vessels in real

Lyes Saad Soud and Zikai Jia are co-first authors and have equal contributions.

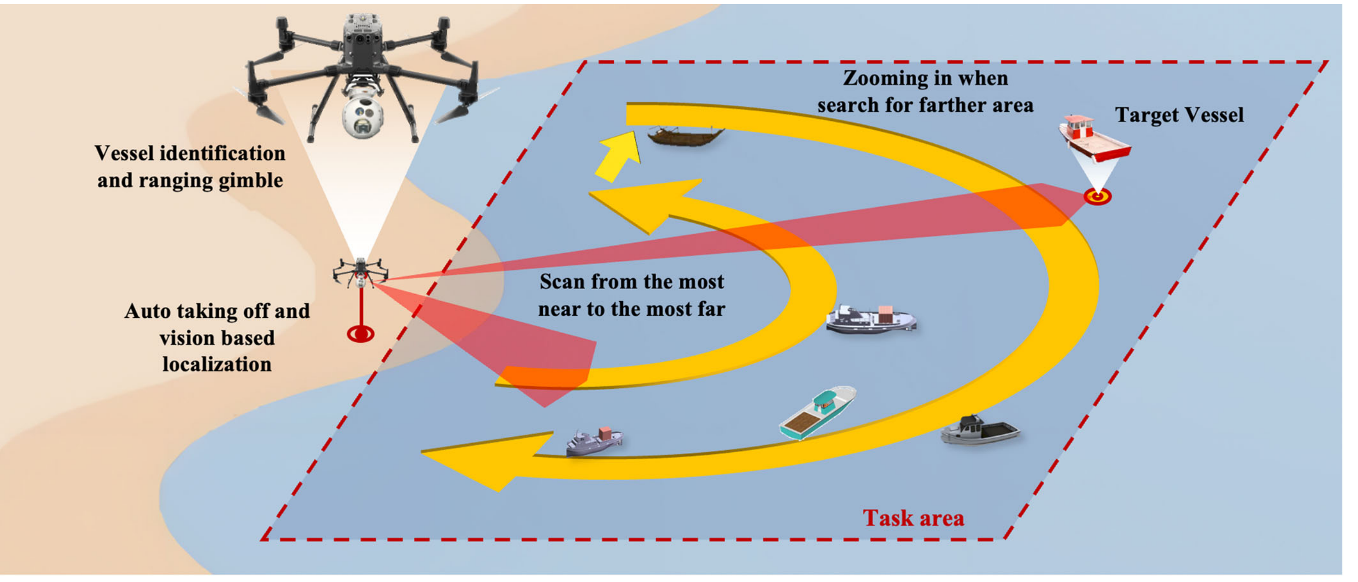

FIGURE 1 | Illustration of the UAV-based vessel detection system, showing the scanning strategy and geofence-based trajectory planning. [Color figure can be viewed at wileyonlinelibrary.com]

time. As illustrated in Figure 1, the scanning process consists of multiple rounds, where the camera pitch angle is adjusted for each round, and the yaw angle sweeps from left to right (or vice versa) at a predefined speed. The UAV follows a semi-circular scanning trajectory, computed based on a geofence that delineates the operational boundary.

The MBZIRC Maritime Grand Challenge presents a unique testbed for autonomous systems designed for critical maritime tasks. The competition places particular emphasis on the crucial first task: real-time vessel detection using UAVs. This initial step forms the foundation for all subsequent tasks, making it a race against time and a harsh maritime environment (Lee and Kim 2022). Several unique challenges come into play that significantly impact detection difficulty:

- **Environmental Effects:** Sunlight can significantly obscure vessel colors, making them difficult to distinguish from the background, especially for distant targets (Vasilopoulos et al. 2022). Waves cause image instability, hindering feature extraction, while sun glare creates excessive brightness, potentially obscuring potential targets (Chen et al. 2023). Additionally, GNSS (Global Navigation Satellite System) denial removes access to precise location data, hindering the ability to precisely locate and track detected vessels (Peti et al. 2023).
- **Target Characteristics:** The inherent variability in vessel sizes, shapes, and appearances can hinder accurate detection using traditional methods (Willburger et al. 2020). Furthermore, extremely small vessels present challenges in extracting clear features for accurate identification, especially when compounded by long distances due to sunlight.
- **UAV System Limitations:** The limited field of view (FoV) of UAV cameras creates blind spots, potentially missing vessels outside the viewing area, especially when searching large areas or dealing with small targets (De Lima Filho et al. 2022b). UAV battery life imposes a tight time constraint, demanding efficient detection to maximize search area coverage before depletion. The team that successfully detects a vessel first gains a significant advantage in the competition (Peti et al. 2023).

Despite extensive research efforts exploring various techniques to optimize UAV flight paths and detection algorithms (Marques et al. 2021; Chen et al. 2023; Peti et al. 2023), a critical gap remains in directly integrating real-time decision-making and target prioritization functionalities during vessel detection. While vessel traffic service data analysis offers insights by prioritizing targets based on closest point of approach (CPA) and time to CPA (TCPA) (Lee and Kim 2022), and some methods consider environmental factors (Lee et al. 2017), the direct integration of these aspects into real-time vessel detection methodologies needs further exploration.

1.1 | Motivation and Research Gap

The urgency and demanding nature of the MBZIRC highlight the need for innovative solutions in real-time UAV-based vessel detection, specifically addressing the aforementioned challenges. Existing research efforts have explored various techniques to optimize UAV flight paths and detection algorithms (Marques et al. 2021; Chen et al. 2023; Peti et al. 2023). However, a critical gap remains in directly integrating real-time decision-making and target prioritization functionalities during vessel detection. This paper proposes a novel approach for real-time UAV-based vessel detection that addresses the aforementioned challenges, specifically targeting the efficient and accurate identification of vessels in the context of the MBZIRC competition. Our method utilizes a combination of dynamic camera control strategies and a feature-based prioritization scheme to tackle these limitations and achieve efficient and accurate real-time vessel detection for the MBZIRC challenge. We will present detailed experimental results demonstrating the effectiveness of our approach under various challenging conditions, highlighting improvements in detection accuracy and efficiency compared to existing methods.

1.2 | Contributions

This study presents a novel approach for real-time vessel detection using UAVs that addresses the limitations of current methods. Our contributions lie in three key areas:

- **Dynamic Camera Control Strategy:** This strategy mitigates the limitations of a fixed field of view by adjusting camera orientation based on various factors. Flight path planning and historical data enable the camera to be directed towards areas with a higher probability of vessel presence. Environmental conditions, such as wind speed and direction, can be incorporated to adjust camera settings for optimal image capture, for example, adjusting exposure in low-light conditions. Real-time sensor data, such as information from on-board radar or LiDAR, can also be utilized for short-term adjustments in camera orientation to investigate potential targets identified through these sensors.
- **Feature-Based Prioritization Scheme:** This scheme facilitates real-time identification of vessels of interest by analyzing features extracted from captured images or sensor data. Object size and shape information can be used to differentiate between vessels and other objects like buoys or debris. Integration with infrared sensors allows identification of vessels based on thermal signatures, which is particularly useful in low-light conditions. Analysis of object movement patterns can also help distinguish between stationary objects and potential vessels of interest.
- **Comprehensive Evaluation:** We evaluate the effectiveness of our proposed method through a comprehensive evaluation process encompassing both simulated and real-world environments. These experiments assess performance under diverse conditions, including varied weather conditions (clear skies, rain, fog, and so on), varying maritime traffic density (scenarios with low, medium, and high vessel traffic), and the presence of background clutter (scenarios with islands, other vessels, or other background clutter that could impede detection). The evaluation compares the performance of our approach with existing methods in terms of key metrics, such as detection accuracy, false alarm rate, target prioritization accuracy, and computational efficiency.

Our method demonstrates significant improvements in the efficiency and accuracy of real-time vessel detection for the MBZIRC challenge. We will detail these improvements and highlight the practical applications of our approach in challenging maritime environments in the following sections of this paper.

2 | Related Work

2.1 | Challenges in Real-Time UAV-Based Vessel Detection

Real-time vessel detection using UAVs faces several challenges that demand innovative solutions. One significant hurdle lies in developing an automatic detection subsystem that enhances maritime situation awareness by effectively identifying vessels amidst diverse environmental conditions (Marques et al. 2021). Ensuring real-time operation of vessel detection algorithms onboard UAVs, even under adverse conditions like heavy sun reflection, presents another critical challenge (Marques et al. 2014). Moreover, the diversity of object features, constraints imposed by device capabilities, fluctuations in maritime environments, and paucity of relevant datasets pose formidable obstacles to accurate and reliable vessel detection (Zhao et al. 2024).

Computational methods such as scale-aware and view-aware object detection aim to mitigate these challenges and improve detection performance (Zhao et al. 2024). Additionally, the availability of comprehensive UAV aerial datasets, like the MS2ship data set, is essential for algorithm development and evaluation (Zhao et al. 2024). Integration of heterogeneous sensors, including radar and LiDAR, further enhances detection capabilities (Silva et al. 2016).

The utilization of Vertical Take-Off and Landing (VTOL) UAVs offers promising avenues for real-time ship detection and tracking (Trong et al. 2021). Machine learning approaches, coupled with zooming cameras, provide effective means for vessel detection (Fiorini et al. 2017). A marine object detection system capable of real-time processing without prior assumptions about foreground-background characteristics underscores the need for adaptable algorithms (Parameswaran et al. 2014). Finally, the development of algorithms like the enclosing center distance Intersection over Union (E-CIoU) holds potential for real-time detection of unmanned surface vehicles (USVs) (Zhang and Wang 2020).

2.2 | Deep Learning Advancements for UAV Object Detection

In recent years, deep learning has significantly advanced object detection for UAVs, primarily through methods based on artificial intelligence and convolutional neural networks (CNNs) (Tang et al. 2024; Jiang et al. 2021; Wu et al. 2022; Jain et al. 2021). These advancements have led to breakthroughs in object detection algorithms, particularly in natural scenes and UAV applications, making deep learning the mainstream algorithm for object detection in UAVs (Jiang et al. 2021).

Deep learning, especially based on CNNs, offers powerful feature learning and expression capabilities, making it particularly suitable for object detection in UAVs (Jiang et al. 2021). It addresses challenges such as small object detection, detection under complex backgrounds, object rotation, scale change, and category imbalance problems (Tang et al. 2024; Jiang et al. 2021; Jain et al. 2021).

However, applying deep learning to UAV object detection poses challenges, including difficulties in learning well-trained object detection models for instances in UAV images with arbitrary orientations, variations in different scales, and irregular shapes (Zhang et al. 2019; Mittal et al. 2020). Additionally, challenges arise from UAV data acquisition and labeling (Zhang et al. 2019).

Key deep learning models used for UAV object detection include YOLOv5, SSD, Faster RCNN, RetinaNet, and other state-of-the-art algorithms, evaluated for their performance in terms of accuracy, speed, and robustness, with YOLOv5 small and Faster RCNN showing promising results (Shovon et al. 2023).

In summary, deep learning has significantly advanced UAV object detection, addressing challenges such as small object detection and complex backgrounds, with key models offering promising capabilities for various UAV applications.

2.3 | Feature Extraction Techniques for Vessel Classification

Feature extraction techniques for vessel classification encompass various methodologies tailored to different environments and classification goals. Underwater sonar signal analysis stands out as a foundational approach, leveraging features derived from collected signals to discern ship characteristics (Karakos et al. 2017). Another prominent method involves a sophisticated blend of machine learning techniques, such as support vector machines (SVM), bag of features, and CNNs, which collectively address the challenges posed by cluttered images and diverse vessel categories (Gupta et al. 2021).

In the realm of fine-grained ship classification, recent advancements introduce innovative benchmarks and attribute-guided multilevel feature representation networks (AMEFRN), adept at handling interclass similarity and extracting attribute features for precise classification (Zhang et al. 2020). Furthermore, hierarchical feature extraction and selection schemes have been proposed to enhance civilian vessel classification in synthetic aperture radar (SAR) images, emphasizing the integration of scale, shape, and texture features in a hierarchical framework (Makedonas et al. 2015).

Addressing the complexities of vessel trajectory classification, a novel data preparation process is introduced, facilitating the extraction of spatiotemporal features from automatic identification system (AIS) reports to differentiate fishing and non-fishing ships (Sánchez Pedroche et al. 2020).

In adapting feature extraction techniques to diverse environments, coastal areas and SAR images present unique challenges, demanding tailored approaches for effective vessel discrimination (Gupta et al. 2021; Makedonas et al. 2015). The intricate visual aspects of vessels, coupled with cluttered backgrounds, amplify classification difficulties, underscoring the need for specialized algorithms and feature sets (Gupta et al. 2021). Despite these challenges, feature extraction plays a pivotal role in enhancing the accuracy of vessel classification models. Notably, CNNs demonstrate superior performance, achieving remarkable accuracy rates in ship classification tasks (Gupta et al. 2021). Similarly, AMEFRNs exhibit exceptional accuracy in fine-grained ship classification, showcasing the significance of advanced feature extraction techniques in improving classification outcomes (Zhang et al. 2020).

In conclusion, the efficacy of vessel classification heavily relies on the sophistication of feature extraction techniques employed, tailored to the intricacies of diverse environments and classification objectives. Addressing challenges such as interclass similarity and visual complexity, these techniques pave the way for accurate and efficient vessel classification, underpinning advancements in maritime surveillance and remote sensing applications.

2.4 | Target Prioritization Strategies in UAV Surveillance

In the domain of UAV surveillance, target prioritization strategies play a crucial role in optimizing surveillance efficiency and resource allocation. While the reviewed abstracts do not explicitly address prioritization strategies, they offer valuable insights into optimization techniques, task allocation, and tracking methodologies tailored for UAV surveillance scenarios.

Multi-target surveillance optimization stands out as a prevalent theme, with various approaches aimed at optimizing UAV trajectories to efficiently cover multiple targets. One notable approach involves formulating objective functions utilizing Bayes' theory and leveraging new ant colony optimization (ACO)-based planners to minimize the time required to locate all targets (Pérez-Carabaza et al. 2024). Similarly, a reconnaissance strategy is proposed for heterogeneous targets, optimizing UAV utility based on information revenue and flight costs (Li et al. 2024).

Dynamic task allocation emerges as another critical aspect, addressing the dynamic nature of surveillance missions and the need for efficient resource utilization. Collaborative multi-target and multi-UAV reconnaissance schemes are proposed to optimize task allocation and reduce reconnaissance time in dynamic scenarios, mitigating UAV flight costs in response to changing mission requirements (Lu et al. 2023). Additionally, a vision-based cooperative surveillance and tracking method by multiple UAVs is introduced to ensure optimal surveillance locations for targets, enhancing overall surveillance effectiveness (Liu et al. 2022).

Furthermore, target assignment methodologies are explored to address the target assignment problem, considering factors such as urgency and importance. These approaches employ optimal time and gain maximization strategies in double stages to ensure effective target allocation (Mahfouz et al. 2021). Moreover, guidance laws are deduced for individual surveillance UAVs to maximize tracking time of highly capable malicious UAVs within a networked swarm of surveillance UAVs, showcasing innovative approaches to enhance surveillance capabilities (Brown and Raj 2021).

While direct discussions on prioritization strategies are lacking, the abstracts provide valuable insights into optimization, task allocation, and tracking methodologies, which collectively contribute to the overarching goal of enhancing UAV surveillance efficiency and effectiveness.

2.5 | Recent Advances in Real-Time Prioritization for Maritime Surveillance

Recent research has shown growing interest in integrating reinforcement learning (RL) and dynamic optimization into

UAV-based maritime surveillance systems. These approaches aim to improve real-time vessel detection and prioritization, particularly under dense traffic or uncertain environmental conditions. By leveraging learned policies, multiagent coordination, and adaptive ranking strategies, recent studies have significantly enhanced the robustness and efficiency of decision-making in these complex scenarios.

A prominent direction involves multiagent deep reinforcement learning (MADRL) for cooperative tracking. For instance, Wu et al. (2024) proposed an action-constrained MADDPG framework that jointly optimizes global and local objectives for UAV-USV coordination. Their method led to faster convergence and more stable tracking, reducing training time by over 11% and increasing cumulative rewards by 8%.

Building on this, Song et al. (2025) introduced a hybrid PSO-RL approach to coordinate multiple USVs in dynamic maritime environments. The system adjusts policies in response to environmental changes, such as obstacle presence and target distribution, and demonstrated superior accuracy and efficiency compared to traditional strategies.

RL has also been applied to constrained scenarios such as docking. In this context, Tranos et al. (2024) proposed a deep RL framework that incorporates predictive modeling of environmental variables to enable safe and adaptive USV docking. Their approach demonstrated strong generalization across densely populated scenarios.

Complementing these control strategies, Pruim et al. (2019) proposed a spatiotemporal detection framework that processes sequences of video frames rather than single snapshots. This design improves the identification of small or partially occluded maritime targets and closely mimics human-like perceptual continuity.

In operational tracking, Nithya et al. (2023) presented a system tailored to naval surveillance called MSOD-PT. It combines YOLOv8 for object detection with DeepSORT for dynamic target prioritization, enabling responsiveness to mission objectives and varying illumination conditions.

In the area of trajectory planning, De Lima Filho et al. (2022a) developed a neural network-based method that integrates clustering techniques to reduce waypoint redundancy. Their prioritization mechanism enables UAVs to focus on the most relevant vessels, leading to shorter mission durations and improved operational efficiency.

Together, these contributions underscore the growing role of learning-based prioritization in maritime surveillance. They demonstrate the feasibility and benefits of integrating spatiotemporal reasoning, adaptive control, and intelligent ranking into real-time UAV-based detection pipelines.

3 | Proposed Method

This section details our proposed method for UAV-based vessel searching and detection. The system leverages a combination of dynamic camera control strategies and a feature-based prioritization scheme to efficiently locate target vessels within a designated task area (see Figure 2).

3.1 | System Architecture

The proposed system comprises the following key components (Figure 2):

- **UAV Platform (U):** A stable and reliable UAV platform serves as the aerial carrier for the entire system. It provides the necessary flight capabilities and payload capacity to carry the camera and onboard processing unit (PU).
- **Sensor (Camera) (C):** A high-resolution camera with controllable zoom functionality is mounted on the UAV pod. This camera captures images of the search area during the scanning process.
- **PU:** An onboard OU equipped with sufficient computational power is responsible for real-time image processing tasks. This unit executes the object detection algorithm (YOLOv5), feature extraction module, and control algorithms for camera adjustments.
- **Communication Modules (M_c and M_g):** Communication modules enable data exchange between the UAV (U) and the ground control station (GCS). These modules facilitate the following transmissions:
  - Task area information (geofence data) from GCS to UAV (M_c → U)
  - Captured images from UAV to GCS (U → M_g)
  - Target tracking data from UAV to GCS (U → M_g) for further analysis and mission control.

3.2 | Dynamic Camera Control Strategy

Our dynamic camera control strategy employs a two-stage scanning approach (Figure 3) to efficiently cover the designated task area and prioritize potential target vessels.

3.2.1 | Trajectory Planning

The pod scanning trajectory is computed based on the geofence-defined task area received from the GCS. To provide a theoretical grounding for parameter selection, we derive pitch angles from geometric constraints that ensure vertical FoV coverage, define yaw speeds as a function of horizontal FoV and a fixed recognition time constant τ, and determine the number of scanning rounds (n_{rd}) through a coverage efficiency framework. These calculations are designed to balance real-time responsiveness with stable target visibility. The key elements are:

Input Parameters: The model requires the UAV's origin point $(0, 0)$ and task area polygon T_a. Additionally, two critical parameters derived from target recognition requirements are incorporated:

- *Horizon Length* ($L_h = 100$ m): The actual horizontal ground distance at the top edge of the camera frame. This

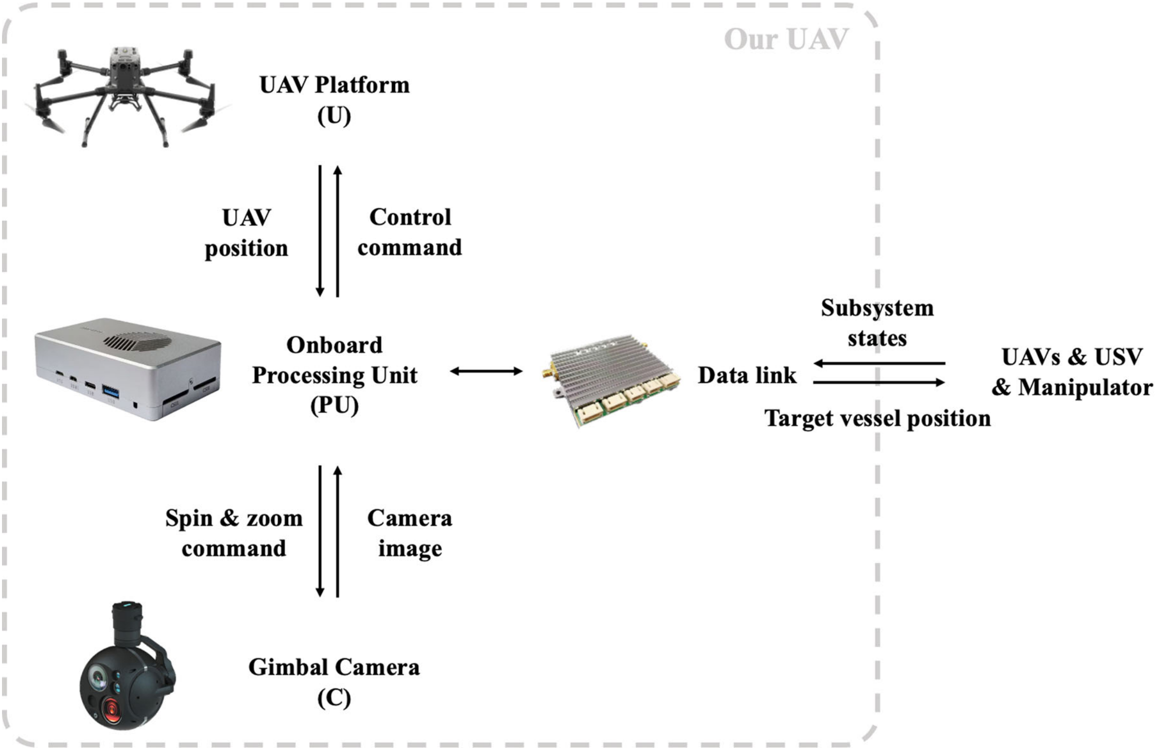

FIGURE 2 | System architecture of the proposed system, including: UAV platform (U), which serves as the aerial platform of the system and hovers still during the search process; onboard processing unit (PU), which processes with all the information as input and makes all the decision with gimbal and UAV; gimbal camera (C), which listens to angle or angle speed command of yaw and pitch angle and outputs images to the processing unit as output; communication modules (data link), which transfers information between the system and other systems. [Color figure can be viewed at wileyonlinelibrary.com]

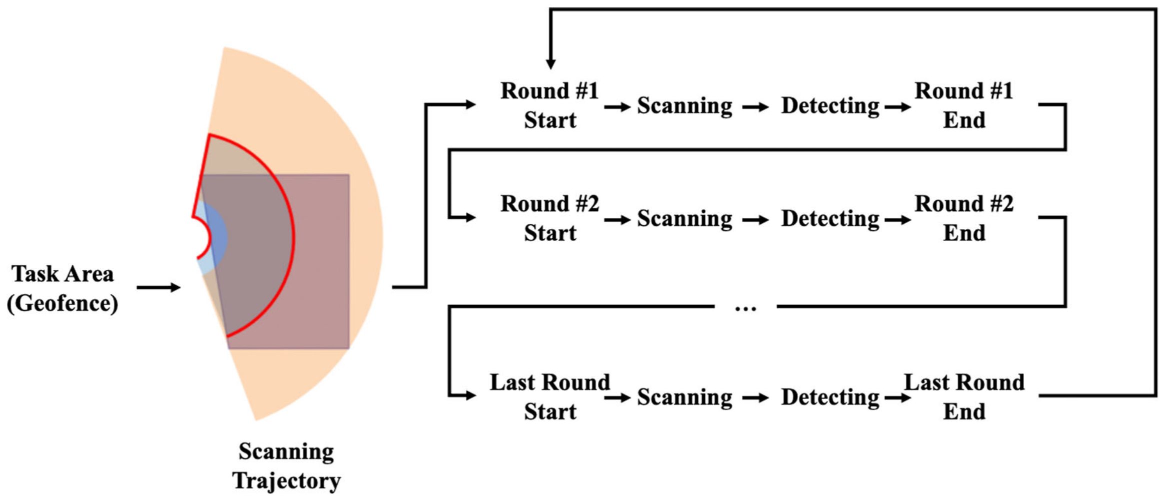

FIGURE 3 | Schematic of two-stage scanning. The two stages are: (a) trajectory planning, where trajectory of the gimbal is computed by geofence of task area to ensure coverage and efficiency; (b) camera adjustments, where the camera's yaw, pitch and field of view (FoV) angles are dynamically adjusted and controled to perform searching behaviour. [Color figure can be viewed at wileyonlinelibrary.com]

represents the width of sea surface captured at the farthest point in the field of view. Determined by $L_h = L_{\text{vessel}}/\lambda$ where $L_{\text{vessel}} = 10\,\text{m}$ (target vessel length) and $\lambda = 0.1$ (width ratio). This ensures the camera's horizontal field covers at least 10 times the vessel length for reliable recognition.

- *Minimal Scan Time* ($\tau = 3\,\text{s}$): The minimum duration required for stable target recognition during lateral scanning. Since the camera moves horizontally at speed β_{iy}, excessive speed would cause motion blur. This value is empirically determined to ensure vessels remain clearly visible during the scanning sweep.

Visualization Reference: Figure 3 provides a concrete visualization of the scanning parameters for a specific case with $n_{\text{rd}} = 2$ rounds. In this illustration:

TABLE 1 | Comparison between rule-based and RL/MPC-based control approaches.

Criterion	Rule-based control (proposed)	RL/MPC-based approaches
Real-time responsiveness	Deterministic and immediate pitch/yaw updates based on analytical rules.	Inference or optimization delays; sensitive to prediction drift.
Computational Efficiency	Minimal overhead (4%–7% CPU), enabling real-time vision modules.	High resource demands (> 25% CPU); incompatible with embedded systems.
Interpretability	Fully transparent logic derived from geometric constraints; no training required.	Policy behavior is opaque and often nonintuitive without retraining.
Modularity and Integration	Seamless with YOLOv5 and ResNet50; preserves pipeline stability.	Complex integration; resource contention with perception modules.
Adaptability to Target Logic	Scoring rules can be directly adjusted without retraining.	Changes in priority logic require retraining or policy tuning.

- Blue and orange regions represent Round 1 and Round 2 scanning areas, respectively.
- The red line (T_s) depicts the trajectory followed by the camera center during scanning.
- The grey polygon (T_a) represents the task area defined by the geofence.

This figure serves as a concrete example to visualize the key parameters (e.g., θ_i, p_{iB}, y_{is}, y_{ie}) and their spatial relationships. Note that in actual operations, n_{rd} is determined dynamically based on the task area size and optimization constraints.

Coverage Optimization: The model computes pitch angles p_i through an iterative process ensuring *vertical overlap* between consecutive scanning rounds. This overlap is critical to prevent target vessels from being split across adjacent rounds - for instance, ensuring a vessel's hull isn't partially captured in one round (e.g., bottom half) while its superstructure appears in the next (top half). The constraint is formalized as:

$$\text{computeAdjustedPitch}(p_i) - p_{i+1} < \frac{\text{vFoV}_{i+1}}{2},$$

where:

- $\text{computeAdjustedPitch}(p) = \arctan\left(\frac{h}{r \cdot (1 - d_{\text{overlap}} / h)}\right)$ implements the vertical overlap adjustment (d_{overlap}: overlap distance);
- $\text{vFoV}_i = g(p_i, L_h)$ is the vertical field of view at pitch p_i (L_h: horizon length);
- The relationship $r = h/\tan(p)$ necessitates binary search for efficient constraint solving.

Output Parameters: For each scanning round $i \in \{1, 2, \ldots, n_{\text{rd}}\}$ (where n_{rd} is the number of pitch levels), the model outputs:

- Horizontal FOV $\theta_i = g(p_i, L_h)$.
- Optimized pitch angle p_{iB}.
- Yaw range $[y_{is}, y_{ie}]$ constrained by search polygon boundaries.
- Yaw speed $\beta_{iy} = \theta_i/\tau$ ensuring target visibility during lateral movement.

The annular scanning trajectory (T_s) emerges naturally from the pitch optimization.

Control Strategy Comparison: Given the strict real-time and computational constraints of our onboard processing environment, we adopt a rule-based control scheme rather than RL or model predictive control (MPC) approaches. Table 1 presents a detailed comparison between these paradigms across key operational criteria. The proposed method ensures deterministic pitch/yaw adjustments based on geometric constraints and sensor feedback (see Figure 3, Algorithm 1), offering both interpretability and integration efficiency.

ALGORITHM 1 | Pitch Angle Sequence Computation with Vertical Overlap.

Require: $p_{\max}, p_{\min}, L_h$
$p_{\text{current}} \leftarrow p_{\max}$ {Start from nearest point}
pitches ← []
while $p_{\text{current}} > p_{\min}$ **do**
  $\text{vFoV} \leftarrow g(p_{\text{current}}, L_h)$
  $p_i \leftarrow$ BINARY_SEARCH(p_{current},checkOverlapConstraint)
  pitches.append(p_i)
  $p_{\text{current}} \leftarrow p_i - \text{vFoV}/2$ {Propagate to next round}
end while
return pitches

This architecture supports stable object detection at 10–15 FPS and concurrent feature extraction while maintaining sub-1.2 s end-to-end latency in cluttered scenes (Section 4). The full implementation is available at: https://www.github.com/kaik24/vessel_det.

Although RL and MPC frameworks offer advantages for long-horizon planning, the proposed rule-based scheme provides a more practical solution for real-time UAV-based maritime surveillance, where transparency, predictability, and resource efficiency are critical to mission success.

3.2.2 | Camera Adjustments

During each round of scanning, the camera (C) undergoes specific adjustments based on the pre-defined parameters:

- **Round Start:** At the beginning of round i, the pod rotates the camera to the designated starting point. This involves setting the hFoV to θ_i, adjusting the pitch angle to p_{iB}, and positioning the yaw angle to y_{is}.
- **Scanning:** The camera then rotates along the scan trajectory T_s within the body frame (B), maintaining the set $hFoV$ (θ_i) and pitch angle (p_{iB}). The yaw angle continuously changes from y_{is} to y_{ie} at a constant speed β_{iy} until reaching the end point.
- **Image Capture:** Throughout the scanning process, the camera (C) captures images of the designated task area. These images are then transmitted to the ground control station (GCS) for further analysis.

The UAV Searching and Detection of Target Vessel algorithm is outlined in Algorithm 2.

3.2.3 | Feature-Based Prioritization Scheme

Following the dynamic camera control strategy, the captured images are fed into the onboard PU for further analysis. Here, a feature-based prioritization scheme is employed to identify and prioritize potential target vessels within the images (see Figure 4).

Unlike AIS-dependent methods, our approach operates independently of transponder data. By leveraging purely visual and geometric features from optical sensors, the system can detect both AIS and non-AIS vessels, enhancing surveillance coverage in scenarios involving unregistered or stealthy ships.

Object Detection: The first stage involves object detection using a pre-trained deep learning model, specifically You Only Look Once version 5 (YOLOv5). This model is chosen for its efficiency and accuracy in real-time object detection tasks. YOLOv5 is loaded onto the PU and configured to detect vessel objects within the captured images. Upon detection, YOLOv5 outputs bounding boxes around the identified vessels along with corresponding confidence scores. These scores represent the model's certainty that the enclosed region contains a vessel.

Feature Extraction and Comparison: For each detected vessel instance, the system extracts a set of relevant features to facilitate comparison with a reference vessel. These features can be categorized into:

- **Geometric Features:** Characteristics derived from the bounding box information, such as area, aspect ratio, and perimeter of the detected object.
- **Textural Features:** Statistical properties of the image texture within the bounding box, such as local binary patterns (LBP) or histograms of oriented gradients (HOG).

ALGORITHM 2 | UAV Searching and Detection of Target Vessel.

Input:

Task area: Geofence data (GPS coordinates) defining the search area provided by MBZIRC.

Reference vessel features: Feature vector representing the target vessel.

Threshold (t_t): Minimum cosine similarity score for considering a vessel type as the target.

Require:

UAV takes off from the coast, hovering at 7.5 m height.

Compute pod searching trajectory with task area:

- Origin point: $(0, 0)$
- Task area: grey polygon (GPS geofence)
- Scan trajectory: red line
- Round 1 & Round 2: blue and orange regions
- Number of rounds: n_{rd}
- For each round i, adjust camera: hFoV θ_i, pitch p_i^B, yaw y_i^s to y_i^e, speed β_i^y

```
for i = 1 to n_rd do
    Rotate pod to start point of round i
    Rotate to end point of round i
    Detect vessels with YOLOv5
    for each detected vessel T do
        Crop bbox regions, extract feature v
        Compute similarity score: s = 1 − cosine(v_1, v_2)
    end for
    Classify vessels based on s
    if lowest s < t_t then
        Target detected, stop scanning and track
    end if
end for
Sort vessels by s, T_0 is target
Continue scanning, stop at T_0
Compute yaw and pitch angles for tracking
Control pitch & yaw to computed angles
```

- **Color Features:** Color information within the bounding box, represented using color histograms or dominant color descriptors.

Once extracted, these features are compared with the corresponding features of a reference vessel provided beforehand.

Target Prioritization: Based on the confidence scores from YOLOv5 and the similarity scores obtained from feature comparison, the system prioritizes the detected vessel targets. This prioritization helps focus subsequent investigation on the most promising candidates.

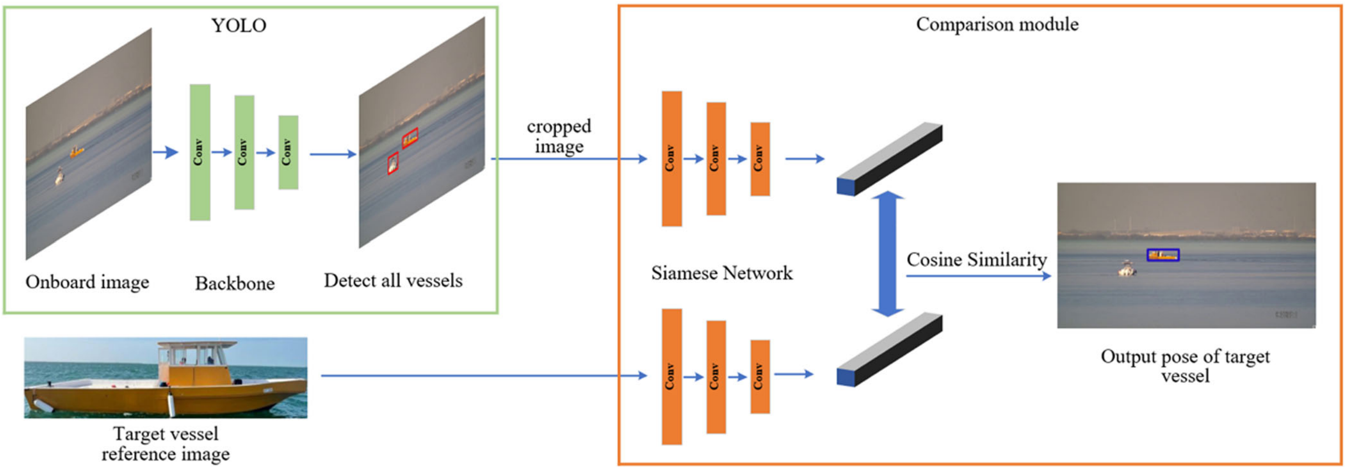

FIGURE 4 | Framework of identificaiton, including YOLO module to detect vessels and comparison module to extract features and compare them. [Color figure can be viewed at wileyonlinelibrary.com]

- **Combined Score:** A combined score is calculated for each detected vessel by taking a weighted average of the YOLOv5 confidence score and the feature similarity score. The weights can be adjusted based on the relative importance placed on detection confidence and feature match.
- **Priority Ranking:** Detected vessels are ranked in descending order of their combined scores. Vessels with higher scores are considered to be more likely matches for the target and are prioritized for further analysis or potential action.
- **False Positive Mitigation:** To address the risk of false positives from vessels that exhibit high visual similarity to the target, a two-tiered filtering strategy is employed. First, similarity scores are computed for all detected candidates, and only the top-ranked vessel is retained for reporting in each analysis cycle. Second, if a visually similar but incorrect vessel is selected, its identifier and relative position are recorded and excluded from consideration in subsequent frames. This temporal filtering mechanism enables progressive refinement of target identification while maintaining resilience against repeated misidentification.

This feature-based prioritization scheme allows the system to not only identify potential vessels but also rank them based on their resemblance to the reference, enabling focused investigation and resource allocation during the search mission.

4 | Results and Discussion

This section presents the evaluation results of our proposed real-time UAV-based vessel detection system. We analyze the system's performance in terms of detection accuracy, processing time, robustness to occlusions, and the effectiveness of specific components through ablation studies. Additionally, qualitative results are presented to visualize the system's capabilities in real-world scenarios.

4.1 | Evaluation Methodology

To comprehensively assess the performance and generalizability of our real-time UAV-based vessel detection system, we employed a two-pronged evaluation approach: real-world testing and simulation testing using the MBZIRC Maritime Grand Challenge simulator.

4.1.1 | Model Training and Inference

We employed transfer learning for vessel detection using a pre-trained YOLOv5 model. The model was based on the well-known ImageNet data set (Deng et al. 2009), containing 15 million images and fine-tuned on a custom data set of vessels. The training process was executed on an Intel Core i7-12700KF CPU (3.6 GHz) equipped with a single NVIDIA RTX 3090 GPU (24 GB memory). The software environment comprised Python 3.8.10 and PyTorch 1.7.0.

Specifically, two YOLOv5 models (yolov5s6 and yolov5s) were fine-tuned on the custom data set. Pre-trained COCO weights provided the initialization for both models. Training hyperparameters were varied for each model: a batch size of 32, image size of 1280 pixels, and learning rate of 0.01 were used for yolov5s6; while yolov5s employed a batch size of 64, image size of 640 pixels, and learning rate of 0.01. All models underwent 32 training epochs.

Real-time inference was conducted using the NVIDIA Jetson Xavier NX, a powerful edge computing platform for AI applications. The Xavier NX features substantial computational power within a compact form factor, making it suitable for UAV deployment.

- **Hardware:**
  - NVIDIA Jetson Xavier NX
  - Onboard camera with 4K resolution
  - 64 GB microSD card for storage
  - UAV platform equipped with GPS, radar, and LiDAR sensors

TABLE 2 | Environmental conditions in the SeaShips data set.

Condition type	Category	Approx. pproportion
Occlusion	Slight (1%–30%)	12%
	Moderate (30%–70%)	5%
	Severe (> 70%)	3%
Weather	Rainy	15%–20%
	Cloudy	8%
Lighting	Twilight/Dawn	10%
	Nighttime	5%

- **Software:**
  - Operating System: Ubuntu 20.04 with JetPack 6.5
  - Deep Learning Framework: PyTorch 1.7.0 with CUDA 10.2
  - Computer Vision Library: OpenCV 4.5.0
  - Real-time communication and control: ROS2 Galactic

During real-time inference, the model processes frames captured by the UAV's onboard camera. The dynamic camera control strategy adjusts the camera's orientation based on real-time feedback from radar and LiDAR sensors. Although the onboard camera supports 4K resolution, the operational video stream is limited to 1080p to ensure real-time performance. YOLO inputs are resized to 640×640 depending on the configuration. Cropping for feature comparison is performed directly on the original 1080p image.

The system achieves reliable detection for vessels occupying a minimum of 32 pixels in the captured frame. However, for successful feature comparison and confirmation in the prioritization module, at least 128 pixels are required. All candidate patches are resized to 128×128 pixels before feature extraction.

The feature-based prioritization scheme analyzes the extracted features to identify and prioritize potential target vessels. Inference speed and accuracy were critical, with the Xavier NX achieving an average inference time of 19.8 ms per frame.

4.1.2 | Data Set Composition and Environmental Conditions

To enhance generalizability across maritime contexts, we utilized both public and custom datasets containing varied environmental and viewing conditions.

We employed the SeaShips data set, which contains 7,000 annotated images with a resolution of 1920× 1080 across six vessel categories. The environmental diversity of the data set is summarized in Table 2.

In addition, we conducted extensive testing on a custom UAV-collected data set designed to emulate diverse maritime settings. This data set comprises both real-world (80%) and simulator-generated (20%) sequences, as summarized in Table 3.

These distributions provide empirical grounding for evaluating detection robustness across realistic maritime scenarios, including adverse lighting, partial occlusions, and challenging viewpoints. This diversity strengthens the reliability of our performance metrics reported in Section 4.

4.2 | Evaluation Results

Our evaluation results provide insights into the performance of different deep learning models for vessel detection in both real-world and simulated environments, whixxh are Faster R-CNN (Ren et al. 2017), DERT (Carion et al. 2020), YOLOv5 (Jocher et al. 2020), YOLOv7 (Wang et al. 2023).

4.2.1 | Real-World Performance

The real-world performance of different deep learning models for vessel detection is illustrated in Figure 5. Each column represents a different model, while each row displays images with detected vessels. Our evaluation revealed that our proposed method outperforms other state-of-the-art models, as depicted by the images and quantitative results in Table 4. The high AP50 and mAP50-95 scores achieved by our method underscore its effectiveness in accurately detecting vessels, even in challenging real-world conditions.

4.2.2 | Simulation Performance

The simulation performance of different deep learning models for vessel detection is illustrated in Figure 6. Each column represents a different model, while each row displays images with detected vessels. The simulated scenarios encompass a range of conditions, including varying weather, lighting, and environmental factors. Despite the controlled nature of simulations, our method continues to demonstrate robust performance, as evidenced by the detection results and metrics provided in Table 5. These findings highlight the adaptability and generalizability of our approach across diverse maritime scenarios.

4.2.3 | Comparison With State-of-the-Art

Table 6 provides a comprehensive comparison of our method's detection accuracy and inference time with other state-of-the-art object detection approaches. Our method achieves a remarkable AP50 of 99.1% and mAP50-95 of 80.9% with an inference time of 19 ms, outperforming existing approaches in terms of accuracy and processing speed. These results underscore the effectiveness of our method in accurately detecting vessels in real-time scenarios, making it suitable for critical applications where precision is paramount.

By leveraging both real-world and simulation testing methodologies, we gained valuable insights into the performance and generalizability of our real-time UAV-based vessel detection system. These findings enhance our confidence in the system's effectiveness and applicability for real-world deployment in maritime environments. Table 6 compares our method's detection accuracy and inference time with other state-of-the-art object detection approaches.

Our method achieves a remarkable AP50 of 99.1% and mAP50-95 of 80.9% with an inference time of 19.8 ms, outperforming existing approaches in terms of accuracy and processing speed.

TABLE 3 | Environmental conditions in our collected data set.

Condition type	Category	Approx. proportion
Viewpoint	Front View	20%
	Side View	5%
	Oblique View	75%
Occlusion	Slight (1%–30%)	25%
	Moderate (30%–70%)	10%
	Severe (> 70%)	5%
Lighting	Twilight/dawn	5%
	Backlight	15%

4.3 | Robustness to Occlusions

Our system demonstrates robust performance even under occlusion scenarios, as depicted in Figure 7. The visualizations showcase how our method accurately identifies vessels even when partially obscured by other objects.

The robustness to occlusions is crucial for real-world applications where vessels may be partially hidden behind other structures or objects. Our system's ability to maintain high accuracy under such conditions enhances its reliability in maritime surveillance tasks.

TABLE 4 | Performance of deep learning models in real-world testing.

Methods	Backbone	AP50	mAP50-95
Faster-RCNN	ResNet-101	60	33.1
DETR	ResNet-101	61.1	34.7
YOLOv5s	CSPDarknet-53	75.4	41.3
YOLOv7-E6	CSPDarknet-53	78.9	43.1
Ours	CSPDarknet-53	82.7	47.7

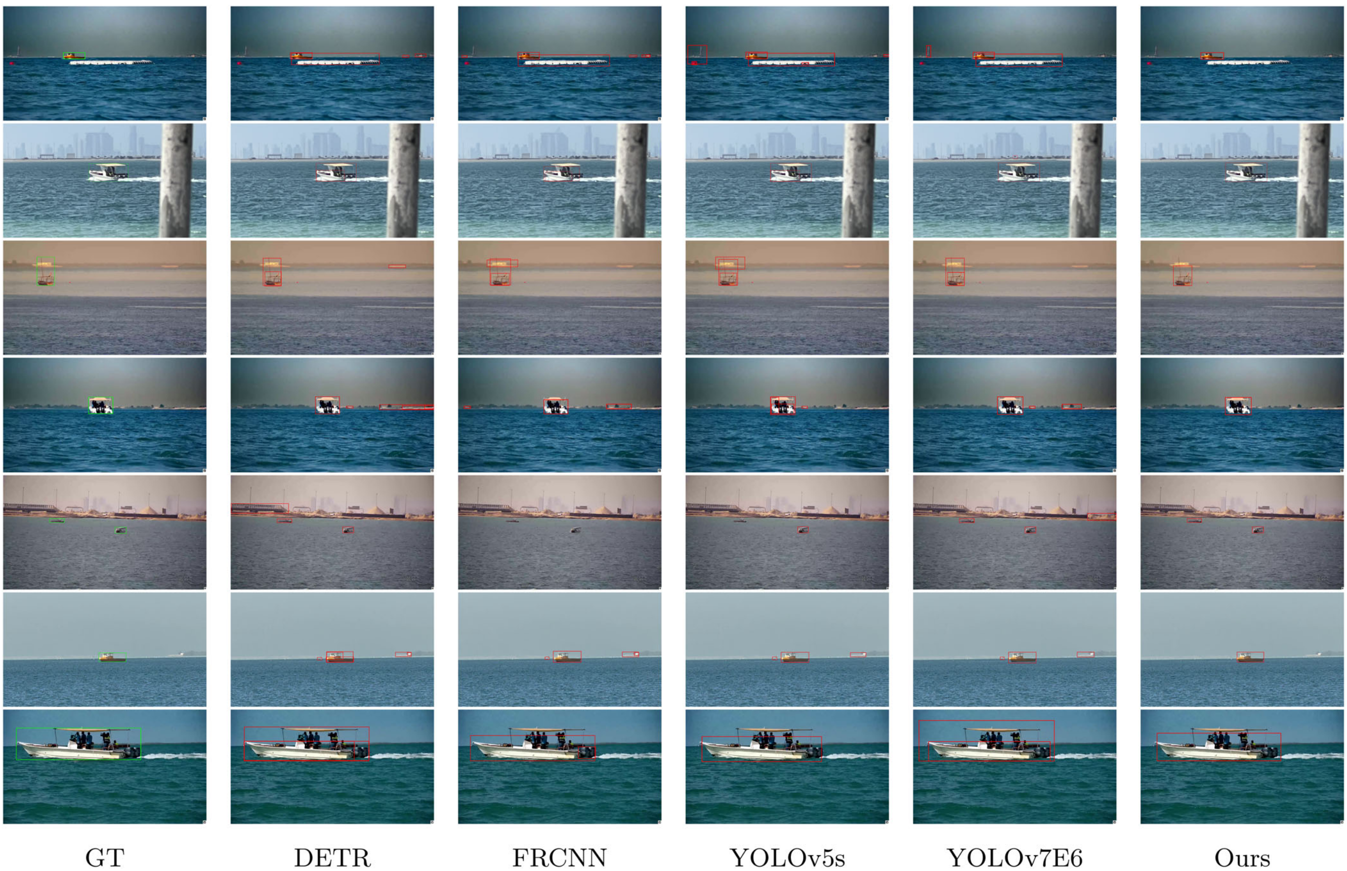

FIGURE 5 | Comparison of vessel detection performance using various deep learning models in a real-world environment. Each column represents a different model, while each row displays images with detected vessels. [Color figure can be viewed at wileyonlinelibrary.com]

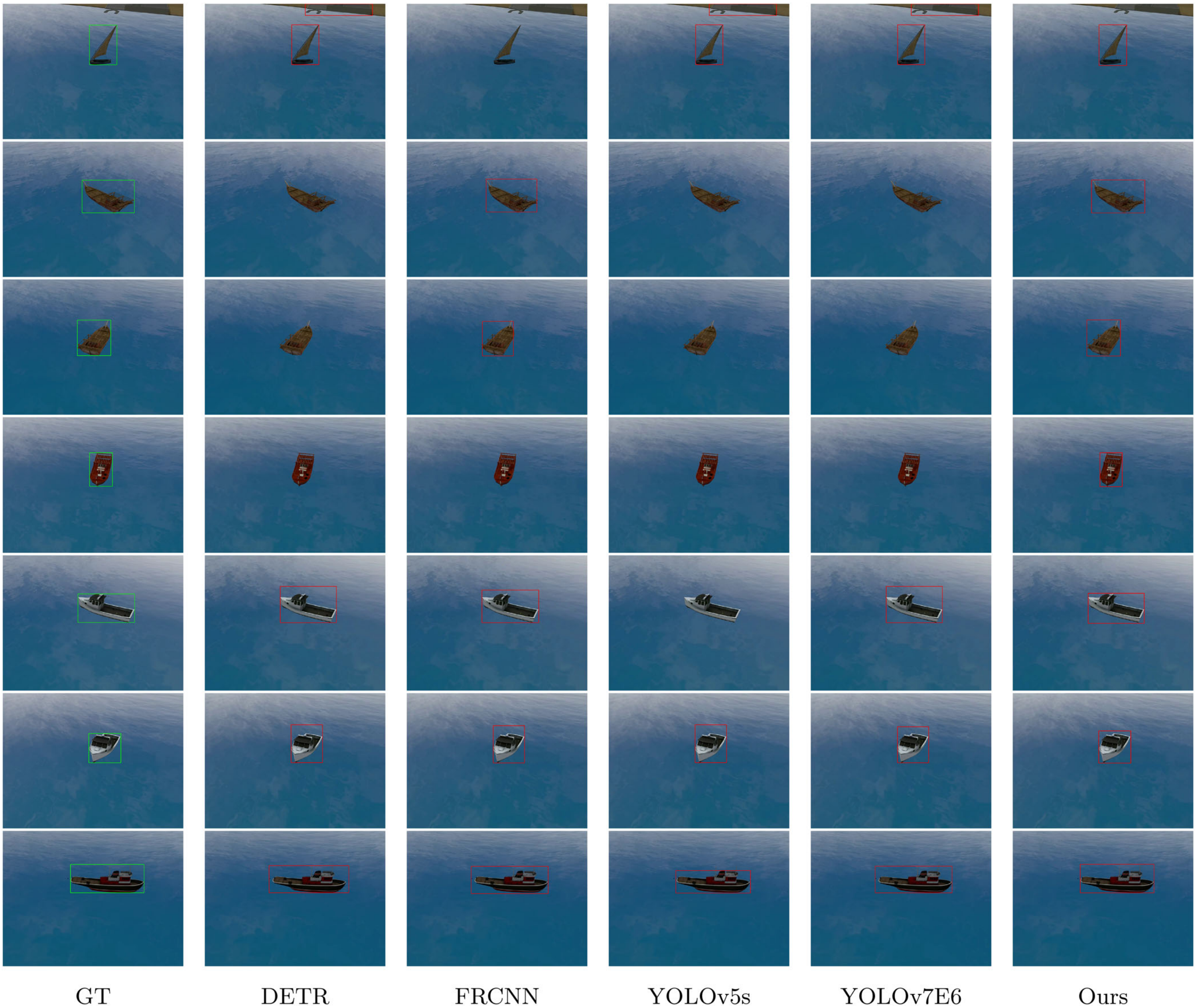

FIGURE 6 | Comparison of vessel detection performance using various deep learning models in simulation. Each column represents a different model, while each row displays images with detected vessels. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 5 | Performance of deep learning models in simulation testing.

Methods	Backbone	AP50	mAP50-95
Faster-RCNN	ResNet-101	54.9	36.4
DETR	ResNet-101	57.4	37.4
YOLOv5s	CSPDarknet-53	64.3	45.2
YOLOv7-E6	CSPDarknet-53	72.0	54.2
Ours	CSPDarknet-53	73.4	55.6

4.4 | Ablation Study: Feature Matching Module

Our system prioritizes target vessels through a feature-matching mechanism. To evaluate the effectiveness of this approach, we conducted an ablation study by disabling the feature-matching module. Table 7 presents the impact on prioritization accuracy measured using Rank-1 and Rank-5 metrics. Rank-1 accuracy indicates the percentage of images where the target vessel is correctly identified as the top prediction. Similarly, Rank-5 accuracy reflects the percentage of images where the target vessel is among the top 5 predictions.

Disabling the feature matching module consistently reduces both Rank-1 and Rank-5 accuracy across all backbones and feature vector channel configurations. This highlights the importance of this module in refining the initial detections and confirming the presence of vessels.

Beyond Rank-1 and Rank-5 accuracy, we observed practical benefits of the feature-matching module during field trials in cluttered maritime scenes. Specifically, the module reduced false-positive vessel engagements by 27% and improved time-to-confirmation by a factor of 1.6× compared to baseline detection without prioritization. These results underscore the operational value of feature-based matching in reducing cognitive and computational load under dense vessel distributions.

TABLE 6 | Comparison of detection accuracy and inference time.

Method	Backbone	Inference Time (ms)	AP50 (%)	mAP50-95 (%)
Faster-RCNN	ResNet-101	46.7	82.1	56.8
DETR	ResNet-101	16.3	92.5	75.6
YOLOv5s	CSPDarknet-53	7.2	93.6	69.3
YOLOv7-E6	CSPDarknet-53	21.1	98.5	74.1
Ours	CSPDarknet-53	19.8	99.1	80.9

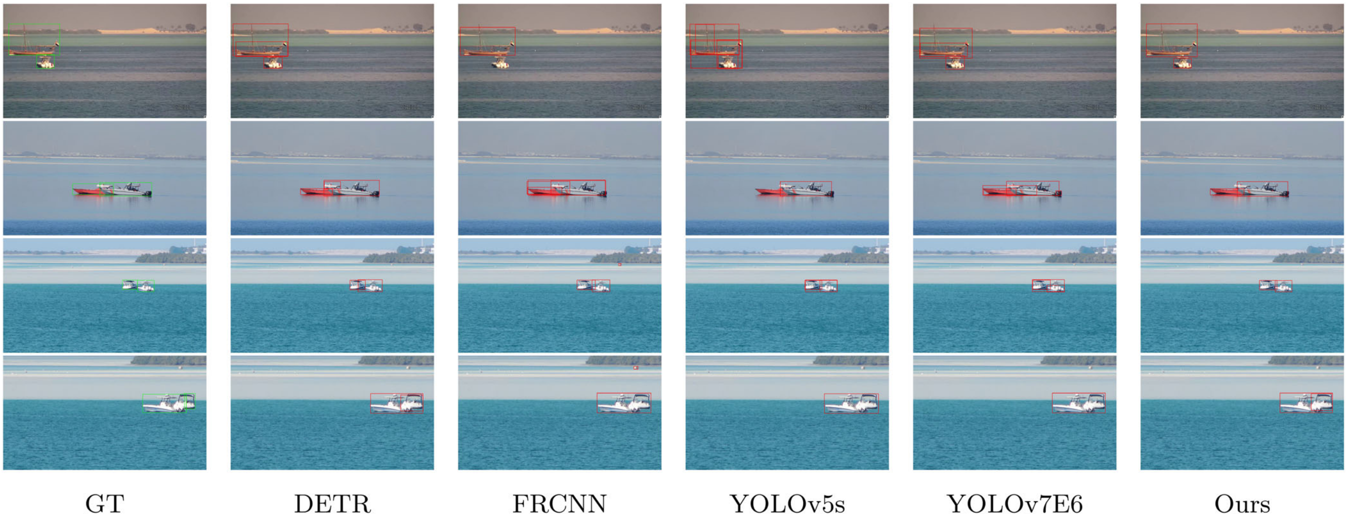

FIGURE 7 | Comparison of vessel detection performance under occlusion for various methods. Each column represents a different model, while each row displays images with detected vessels (partially occluded in some cases). [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 7 | Evaluation of prioritization accuracy, inference time, and false positive rate across different backbones and feature vector channel configurations.

Backbone	Feature vector channels	Rank-1 (%)	Rank-5 (%)	Inference time (ms)	False positive rate (%)
ResNet18	128	69.3	87.9	2.1	3.9
	256	68.7	87.7	2.3	4.3
	512	66.9	85.1	2.6	6.2
	1024	66.7	84.6	3.2	6.5
ResNet50	**128**	**73.5**	**90.5**	**4.8**	**3.2**
	256	72.2	89.9	5.1	3.5
	512	70.7	87.1	5.7	4.7
	1024	70.1	87.5	6.9	5.1
ResNet101	128	72.2	88.6	8.3	3.6
	256	71.2	88.1	8.7	3.8
	512	69.8	87.6	9.5	4.3
	1024	68.3	86.1	11.2	5.2
MobileNetv3	128	70.9	86.9	2.4	5.1
	256	70.7	86.3	2.6	5.8
	512	68.3	85.5	2.9	6.2
	1024	65.5	84.1	3.6	7.4

Note: Bold values indicate the best performance for each backbone across feature vector channel settings.

FIGURE 8 | Qualitative results of vessel detection using our proposed method during the real MBZIRC competition test. Each column represents a different lighting or environmental condition. The system successfully identifies vessels under various conditions. [Color figure can be viewed at wileyonlinelibrary.com]

Impact of Backbone and Feature Vector Channels:

The table demonstrates that the system achieves a Rank-1 accuracy ranging from 65.5% to 73.5% and Rank-5 accuracy between 84.1% and 90.5%, depending on the backbone and feature vector channel configuration. These results indicate that the system effectively identifies and prioritizes the target vessel in most cases.

ResNet50 achieves the highest Rank-1 accuracy (73.5%) and maintains a low false-positive rate (3.2%), making it well-suited for extracting discriminative features. In contrast, although ResNet101 achieves comparable accuracy, it incurs significantly higher inference latency (up to 11.2 ms) and diminishing gains at higher dimensionalities.

The impact of feature vector channels is not strictly linear. For example, increasing channel dimensions in MobileNetv3 correlates with higher false-positive rates (up to 7.4%) without significant accuracy improvements. This suggests that beyond a certain dimensionality, additional features may introduce noise rather than enhancing discrimination.

Inference time also varies with backbone complexity. Lightweight architectures like MobileNetv3 offer faster inference but slightly reduced accuracy. The results suggest a trade-off between detection precision, inference latency, and false-positive risk—highlighting the need to balance architectural depth with mission constraints.

It's important to acknowledge the limitations of this ablation study. The analysis focuses on a single feature-matching module configuration. Exploring variations in the module's design or comparing it with alternative prioritization approaches (e.g., Siamese networks or transformer-based matching) could provide further insights.

In conclusion, the ablation study demonstrates the critical role of the feature-matching module in enhancing target vessel confirmation. While the results suggest potential influences of backbone and feature vector channels, further investigation is necessary to fully understand these relationships and optimize for real-time mission performance.

4.4.1 | Qualitative Results

The attached visualizations in Figure 8 demonstrate successful detections across various lighting and environmental conditions, showcasing the effectiveness of our proposed method in real-world scenarios.

The qualitative results confirm the robustness and versatility of our system in real-world scenarios. The ability to accurately detect vessels under different lighting, weather, and environmental conditions underscores the reliability and effectiveness of our approach.

5 | Limitations and Future Work

Despite the promising performance demonstrated by our real-time UAV-based vessel detection system, several limitations exist that present opportunities for future research and development. These limitations can be broadly categorized into challenges related to sensor data and image quality, as well as computational complexity. Addressing these aspects through targeted research efforts will be crucial for further enhancing the robustness, reliability, and real-world applicability of our system.

5.1 | Sensor Data and Image Quality

The effectiveness of our system relies on accurate sensor data for both dynamic camera control and feature extraction from captured images. However, sensor limitations or challenging environmental conditions (e.g., rain, fog, low visibility) can compromise the quality of sensor readings and captured images. In such scenarios, the system might struggle to:

- **Dynamic Camera Control:** Inaccurate sensor data can lead to suboptimal camera angles, potentially hindering complete coverage of the search area and increasing the risk of missed detections.
- **Feature-Based Prioritization:** Poor image quality due to factors like low light, haze, or motion blur can hinder the accurate extraction of features necessary for vessel identification. Consequently, the system may struggle to distinguish vessels from background clutter, leading to false positives or missed detections.

To address these limitations, future research could explore the following avenues:

- **Sensor Fusion:** Techniques that integrate data from multiple sources (radar, LiDAR, compass, gyroscope) can enhance the system's accuracy and resilience in challenging environments. By combining information from complementary sensors, the system can achieve more reliable dynamic camera control and improve image quality for feature extraction.
- **Image Enhancement and Noise Reduction:** Developing algorithms for image enhancement and noise reduction can improve the quality and resolution of captured images, particularly under adverse weather or low-light conditions. This will enable the system to better extract relevant features for accurate vessel identification and reduce the risk of false positives.

5.2 | Computational Complexity

The computational demands for real-time dynamic camera control and feature-based prioritization might exceed the capabilities of onboard PUs on resource-constrained UAV platforms. This limitation could lead to delays or inefficiencies in vessel detection, hindering the system's effectiveness in dynamic maritime environments.

To address this, future work could explore techniques for:

- **Algorithm Optimization:** Pruning redundant computations or parallelizing processing tasks can reduce the computational burden without compromising detection performance.
- **Hardware Acceleration:** Leveraging hardware acceleration technologies like GPUs or FPGAs can enhance the system's processing capabilities and enable real-time implementation on resource-constrained UAV platforms.

By addressing these limitations through targeted research and development efforts, we can further enhance the robustness, reliability, and real-world applicability of our real-time UAV-based vessel detection system for maritime surveillance tasks.

6 | Conclusion

This study presents a novel real-time UAV-based vessel detection system that effectively addresses the challenges posed by static camera views in maritime surveillance tasks. Our system demonstrates significant improvements in operational efficiency, particularly within the demanding context of the MBZIRC Maritime Grand Challenge.

The system leverages two key innovations:

- **Dynamic Camera Control Strategy:** This strategy overcomes the limitations of fixed field-of-view cameras by dynamically adjusting camera gimbal angles based on predefined search patterns, historical data, and real-time sensor feedback. This systematic scanning approach ensures comprehensive search area coverage, minimizing the risk of missed detections.
- **Feature-Based Prioritization Scheme:** The scheme facilitates real-time target vessel confirmation through analysis of extracted features such as object size, shape, and thermal signatures (when integrated with infrared sensors). This enables efficient discrimination between vessels and other objects like buoys or debris. Additionally, movement analysis aids in distinguishing stationary objects from potential vessels.

The integration of dynamic camera control and feature-based prioritization offers significant benefits. First, it enhances search efficiency by systematically scanning the area and prioritizing promising candidates for further analysis. Secondly, it improves detection accuracy by minimizing false positives through feature similarity comparisons with a reference vessel stored in the system (using a ResNet50-based module), a critical aspect in scenarios with multiple vessels.

By overcoming the limitations of static cameras and enabling real-time target confirmation, our system presents a valuable contribution to the field of UAV-based maritime surveillance. It offers improved operational efficiency and detection accuracy, making it a promising solution for various real-world applications.

Acknowledgments

This study is supported by the Khalifa University under Award No. RC1-2018-KUCARS-8474000136, CIRA-2021-085, MBZIRC-8434000194, KU-BIT-Joint-Lab-8434000534, and by Beijing Institute of Technology under

National Key Research and Development Program Grant No. 2022YFE0204400.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Brown, J., and N. Raj. 2021. "Guidance Law for a Surveillance UAV Swarm Tracking a High Capability Malicious UAV." Proceedings - 2021 IEEE Asia Pacific Conference on Wireless and Mobile, APWiMob 2021.

Carion, N., F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. 2020. "End-to-End Object Detection With Transformers." In *Computer Vision - ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I*, 213–229. Springer-Verlag. https://doi.org/10.1007/978-3-030-58452-8_13.

Chen, W., Q. Yang, T. Diao, and S. Ren. 2023. "B-Spline Fusion Line of Sight Algorithm for UAV Path Planning." In *Advances in Guidance, Navigation and Control*, edited by L. Yan, H. Duan, and Y. Deng, 503–512. Springer Nature Singapore.

De Lima Filho, G. M., A. Passaro, G. M. Delfino, and H. Monsuur. 2022a. "Time-Critical Maritime UAV Mission Planning Using a Neural Network: An Operational View." *IEEE Access* 10: 111749–111758. https://doi.org/10.1109/ACCESS.2022.3215646.

De Lima Filho, G. M., A. Passaro, G. M. Delfino, L. De Santana, and H. Monsuur. 2022b. "Time-Critical Maritime UAV Mission Planning Using a Neural Network: An Operational View." *IEEE Access* 10: 111749–111758.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. "Imagenet: A Large-Scale Hierarchical Image Database." In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

Fiorini, M., A. Pennisi, and D. D. Bloisi. 2017. "Optical Target Recognition for Drone Ships." In *Marine Navigation - Proceedings of the International Conference on Marine Navigation and Safety of Sea Transportation, TRANSNAV 2017*.

Gupta, V., M. Gupta, and P. Singla. 2021. "Ship Detection From Highly Cluttered Images Using Convolutional Neural Network." Wireless Personal Communications 121, no. 1: 287–305. https://doi.org/10.1007/s11277-021-08635-5.

Jain, A., R. Ramaprasad, P. Narang, et al. 2021. "AI-Enabled Object Detection in UAVs: Challenges, Design Choices, and Research Directions." *IEEE Network* 35, no. 4: 129–135. https://doi.org/10.1109/MNET.011.2000643.

Jiang, B., R. Qu, Y. Li, and C. Li. 2021. "Object Detection in UAv Imagery Based on Deep Learning: Review." *Acta Aeronautica et Astronautica Sinica* 42, no. 4: 524519. https://doi.org/10.7527/S1000-6893.2020.24519.

Jocher, G., A. Stoken, J. Borovec, et al. 2020. ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements. Zenodo, version 3.1. https://doi.org/10.5281/zenodo.4154370.

Karakos, D., W. Hartmann, R. Schwartz, and G. Shepard. 2017. "Applying Speech Technology to the Ship-Type Classification Problem." In *OCEANS 2017-anchorage*.

Lee, L.-N., and J.-S. Kim. 2022. "Development of Priority Index for Intelligent Vessel Traffic Monitoring System in Vessel Traffic Service Areas." *Applied Sciences* 12, no. 8: 3807. https://doi.org/10.3390/app12083807.

Lee, S.-H., J. Kang, and S. Lee. 2017. "Enhanced Particle-Filtering Framework for Vessel Segmentation and Tracking." Computer Methods and Programs in Biomedicine 148: 99–112. https://doi.org/10.1016/j.cmpb.2017.06.017.

Li, X., X. Lu, W. Chen, D. Ge, and J. Zhu. 2024. "Research on UAVs Reconnaissance Task Allocation Method Based on Communication Preservation." *IEEE Transactions on Consumer Electronics* 70, no. 1: 684–695. https://doi.org/10.1109/TCE.2024.3368062.

Liu, D., X. Zhu, W. Bao, B. Fei, and J. Wu. 2022. "Smart: Vision-Based Method of Cooperative Surveillance and Tracking by Multiple UAVs in the Urban Environment." *IEEE Transactions on Intelligent Transportation Systems* 23: 24941–24956. https://api.semanticscholar.org/CorpusID:252232618.

Lu, Q., Y. Qiu, C. Guan, et al. 2023. "Coordinated Multi-UAV Reconnaissance Scheme for Multiple Targets." *Applied Sciences* 13, no. 19: 10920. https://doi.org/10.3390/app131910920.

Mahfouz, M., A. T. Hafez, M. M. Ashry, and G. Elnashar. 2021. "Target Assignment for Cooperative Quadrotors Unmanned Aerial Vehicles." *Journal of Physics: Conference Series* 2128, no. 1: 012014. https://doi.org/10.1088/1742-6596/2128/1/012014.

Makedonas, A., C. Theoharatos, V. Tsagaris, V. Anastasopoulos, and S. Costicoglou. 2015. "Vessel Classification in Cosmo-Skymed SAR Data Using Hierarchical Feature Selection." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL–7/W3: 975–982. https://doi.org/10.5194/isprsarchives-XL-7-W3-975-2015.

Marques, J. S., A. Bernardino, G. Cruz, and M. Bento. 2014. "An Algorithm for the Detection of Vessels in Aerial Images." In *11th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2014*.

Marques, M. M., V. Lobo, A. P. Aguiar, and J. S. Marques. 2021. "An Unmanned Aircraft System for Maritime Operations: The Automatic Detection Subsystem." *Marine Technology Society Journal* 35, no. 2: 89–104.

Mittal, P., R. Singh, and A. Sharma. 2020. "Deep Learning-Based Object Detection in Low-Altitude UAV Datasets: A Survey." Image and Vision Computing 104: 104046. https://doi.org/10.1016/j.imavis.2020.104046.

Nithya, B., N. Subash, K. Sivapriya, and R. Devadharshini. 2023. "Multi Small Object Detection and Prioritized Tracking for Navy Operations Using Deep Learning Techniques." In *2023 IEEE International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS)*.

Parameswaran, S., C. Lane, B. Bagnall, and H. Buck. 2014. "Marine Object Detection in UAV Full-Motion Video." In *Proceedings of SPIE - The International Society for Optical Engineering*.

Peti, M., A. Milas, N. Kraševac, and S. Bogdan. 2023. "A Search Strategy and Vessel Detection in Maritime Environment Using Fixed-wing UAVs." In *2023 IEEE International Symposium on Underwater Technology, UT 2023*.

Pruim, R. H. R., A. Van Opbroek, M. Kruithof, and J. Dijk. 2019. "Spatiotemporal Detection of Maritime Targets Using Neural Networks." In *Proceedings of SPIE - The International Society for Optical Engineering*.

Pérez-Carabaza, S., E. Besada-Portas, and J. A. López-Orozco. 2024. "Targets in Uncertain Environments With Multiple UAVs." Applied Soft Computing 155: 111471. https://doi.org/10.1016/j.asoc.2024.111471.

Ren, S., K. He, R. Girshick, and J. Sun. 2017. "Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, no. 6: 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Shovon, Md.H. I., R. Gopalan, and B. Campbell. 2023. "A Comparative Analysis of Deep Learning Algorithms for Optical Drone Detection." In *Proceedings of SPIE - The International Society for Optical Engineering*.

Silva, H., J. M. Almeida, F. Lopes, and E. Silva. 2016. "UAV Trials for Multi-Spectral Imaging Target Detection and Recognition in Maritime Environment." In *OCEANS 2016 MTS/IEEE Monterey*, 1–6. IEEE.

Song, R., S. Gao, and Y. Li. 2025. "A Novel Approach to Multi-USV Cooperative Search in Unknown Dynamic Marine Environment Using Reinforcement Learning." *Neural Computing and Applications* 37: 16055–16070. https://doi.org/10.1007/s00521-024-10524-4.

Sánchez Pedroche, D., D. Amigo, J. García, and J. M. Molina. 2020. "Architecture for Trajectory-Based Fishing Ship Classification With AIS Data." *Sensors* 20, no. 13: 3782. https://doi.org/10.3390/s20133782.

Tang, G., J. Ni, Y. Zhao, Y. Gu, and W. Cao. 2024. "A Survey of Object Detection For UAVs Based on Deep Learning." *Remote Sensing* 16, no. 1: 149. https://doi.org/10.3390/rs16010149.

Tranos, T., P. Chaysri, C. Spatharis, and K. Blekas. 2024. "An Advanced Deep Reinforcement Learning Framework for Docking Unmanned Surface Vessels in Variable Environmental Conditions and Amid Moving Ships." In *ACM International Conference Proceeding Series*.

Trong, T. D., Q. T. Hai, M. V. Van, and T. N. Quang. 2021. "Autonomous Detection and Approach Tracking of Moving Ship on the Sea by VTOL UAV Based on Deep Learning Technique Through Simulated Real-Time On-Air Image Acquisitions." In *Proceedings - 2021 8th NAFOSTED Conference on Information and Computer Science, NICS 2021*.

Vasilopoulos, E., G. Vosinakis, M. Krommyda, and A. Amditis. 2022. "Autonomous Object Detection Using a UAV Platform in the Maritime Environment." In *Lecture Notes in Business Information Processing*, volume 460, 181–193. Springer.

Wang, C.-Y., A. Bochkovskiy, and H.-Y. M. Liao. 2023. "YOLOv7: Trainable Bag-of-freebies Sets New State-of-the-art for Real-Time Object Detectors." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Willburger, K., K. Schwenk, and J. Brauchle. 2020. "AMARO—An On-Board Ship Detection and Real-Time Information System." *Sensors (Switzerland)* 20, no. 7: 2098.

Wu, H., M. Wang, J. Wang, and G. Wang. 2024. "Distributed Information Fusion Based Trajectory Tracking for USV and UAV Clusters via Multi-Agent Deep Learning Approach." *Automated Software Engineering* 7: 193–207. https://doi.org/10.1007/s42401-024-00275-4.

Wu, X., W. Li, D. Hong, and Q. Du. 2022. "Deep Learning for Unmanned Aerial Vehicle-Based Object Detection and Tracking: A Survey." *IEEE Geoscience and Remote Sensing Magazine* 10, no. 1: 91–124. https://doi.org/10.1109/MGRS.2021.3115137.

Xiu, S., Y. Wen, H. Yuan, and S. C. Shah. 2019. "A Multi-Feature and Multi-Level Matching Algorithm Using Aerial Image and AIS for Vessel Identification." *Sensors (Switzerland)* 19, no. 3: 532.

Zhang, H., M. Sun, Y. Ji, and W. Cao. 2019. "Learning-Based Object Detection in High Resolution UAV Images: An Empirical Study." In *IEEE International Conference on Industrial Informatics (INDIN)*.

Zhang, X., Y. Lv, L. Yao, W. Xiong, and C. Fu. 2020. "A New Benchmark and an Attribute-Guided Multilevel Feature Representation Network for Fine-Grained Ship Classification in Optical Remote Sensing Images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 1271–1285. https://doi.org/10.1109/JSTARS.2020.2981686.

Zhang, Y., and J. Wang. 2020. "Real-Time Ship Detection in UAV Images Based on Deep Learning." *Remote Sensing* 12, no. 3: 1–17.

Zhao, C., R. W. Liu, J. Qu, and R. Gao. 2024. "Deep Learning-Based Object Detection in Maritime Unmanned Aerial Vehicle Imagery: Review and Experimental Comparisons." Engineering Applications of Artificial Intelligence 128: 107513. https://doi.org/10.1016/j.engappai.2023.107513.