Sensors and Materials, Vol. 38, No. 3 (2026) 1695–1715
MYU Tokyo

S & M 4401

Bandwidth-aware Multimodal Sensor Data Prioritization for Multi-unmanned Surface Vehicle Tracking Using Dynamic Marine Attention Network

Lu Wen,[1] Jiayi Wen,[2*] Xiaorong Zhang,[1] Yan Li,[1] and Qiang Wang[1]

[1]Qingdao Ocean Shipping Mariners College, Qingdao 266400, China
[2]INTI International University, Nilai 71800, Malaysia

(Received January 29, 2026; accepted March 19, 2026)

Keywords: marine sensing, sensor data prioritization, deep reinforcement learning (DRL), multimodal sensors, LiDAR/radar/sonar

The deployment of unmanned surface vehicles (USVs) for autonomous maritime operations requires robust target tracking based on multimodal marine sensing under stochastic ocean environments. However, conventional tracking methods suffer from high-dimensional sensor noise and limitations in onboard sensing resources and the computational burden of multiagent coordination. We developed the dynamic marine attention network (DyMAN), a sensing-oriented multiagent deep reinforcement learning framework integrated with an ensemble hard attention mechanism and game theory. By utilizing a centralized critic and distributed actor architecture, DyMAN mitigates multiagent nonstationarity while ensuring decentralized execution. To address the limitations of onboard hardware, a hard attention mechanism is integrated to prioritize informative sensor data, effectively filtering environmental noise from light detection and ranging and radio detection and ranging data, and improving the efficiency of multimodal sensor data processing by focusing computational resources on important targets. The Nash equilibrium is also integrated to ensure stable cooperative sensing and coordination and reduce decision conflicts among the USVs in the fleet. The experimental results demonstrate that DyMAN outperforms baseline algorithms, including multiagent deep deterministic policy gradient, deep deterministic policy gradient, and proximal policy optimization in terms of cumulative reward and convergence stability. DyMAN maintains a stabilized minimum interagent distance of approximately 40 units, which is a 50% improvement in formation stability compared with that in early training, and significantly reduces tracking error and energy consumption. The results of DyMAN provide a computational framework for distributed marine sensing nodes, enhancing the reliability and efficiency of multimodal sensing systems in complex maritime environments.

*Corresponding author: e-mail: wenjiayi202201@163.com
https://doi.org/10.18494/SAM6238

ISSN 0914-4935
https://myukk.org/

1. Introduction

Unmanned surface vehicles (USVs) are widely deployed as autonomous maritime platforms that support diverse oceanic operations.[1] Advances in automation and machine learning have contributed to the expansion of their capabilities in navigation,[2] path planning,[3] real-time sensing and sensor-based perception, and task execution with minimal human oversight.[4] As a result, USVs have become essential for marine environmental sensing and monitoring, maritime surveillance, search-and-rescue missions, and oceanographic research.[1] With global maritime activity intensifying, USVs are required to reliably track targets such as vessels, marine organisms, or drifting objects even under highly stochastic oceanic conditions based on multimodal sensing data.

Such target tracking requires robust guidance, precise motion control, and effective disturbance rejection,[5] which must be supported by adaptive decision-making.[6,7] However, marine environments are inherently uncertain and time-varying, with disturbances such as waves, currents, wind, and sensor measurement noise. Those disturbances degrade tracking accuracy and destabilize closed-loop control systems.[8,9] Therefore, USVs must autonomously conduct synchronized operations and avoid collisions. However, the real-time coordination for such operations is hindered by communication constraints and interference among multiple sensing nodes and platform behaviors, which lead to redundant actions or conflicting strategies.[8] Furthermore, abrupt changes in target behavior or environmental conditions demand the rapid adaptability of USVs, which traditional models lack owing to heavy reliance on fixed parameters.[10]

Conventional controllers used in the traditional USVs include proportional–integral–derivative (PID) controllers, linear quadratic regulators, and model predictive control (MPC). These controllers satisfactorily perform under stable and predictable conditions but struggle in nonlinear, disturbed environments.[10,11] MPC optimizes actions within defined constraints.[11] However, it imposes high computational burdens and remains highly sensitive to model mismatch. Fuzzy controllers are introduced to overcome these limitations by improving robustness.[12] Nevertheless, they fail to effectively operate in complex multi-USV communication and sensing environments, where scalability and communication are essential.[12] Therefore, learning-based methods such as reinforcement learning (RL) have been applied,[13] but multiagent RL (MARL)'s nonstationarity undermines simultaneous policy updates for diverse agents' cooperation. In addition, the deployment of these algorithms is constrained by noisy sensing environments and the limited bandwidth of onboard sensors, which restricts the availability of clean, high-frequency sensor data essential for reliable control.[14]

The tracking performance of USVs depends on well-organized control algorithms and, more broadly, on the synergy between marine sensing hardware and sensor data processing. Therefore, we developed a dynamic marine attention network (DyMAN) to overcome the sensing-to-control bottleneck by prioritizing task-relevant data streams.[15] DyMAN aligns with emerging sensor technologies that contain intelligent nodes that enable adaptive sampling and mission-aware data prioritization.[16] Such a capability requires resilient autonomous marine sensing networks that enhance perception and control to strengthen the link between algorithmic frameworks and physical sensing hardware in DyMAN.[17]

DyMAN introduces a game-theoretical multiagent framework to address the challenges of dynamic target tracking in marine environments. The network employs a centralized critic with distributed actors, which mitigates nonstationarity and enables decentralized execution.[18,19] It integrates a hard attention mechanism that focuses computational resources on targets and related variables, thereby improving robustness against sensor measurement noise and sensing bandwidth constraints.[16] DyMAN incorporates Nash equilibrium, a dynamic game-theoretic interaction principle, to enable agents to adapt to environmental variability and peer USV behaviors without requiring precise hydrodynamic models.[20,21]

2. Marine Sensing System and Materials Considerations

Autonomous target tracking in marine environments relies not only on decision-making algorithms but also on the reliability of onboard sensing systems. In practical deployments, USVs operate under dynamic and uncertain ocean conditions where sensor measurements may be affected by environmental disturbances, limited communication bandwidth, and constrained onboard computational resources. Therefore, an effective sensing framework must integrate multimodal sensor observations, efficient data processing, and robust hardware design. In this section, we describe the sensing architecture considered in this study, including the sensor suite and measurement model, system-level sensing constraints, and materials considerations for marine sensing nodes.

2.1 Sensor suite and measurement model

Autonomous USV operations rely on heterogeneous sensing systems that provide complementary environmental observations. In typical maritime applications, a USV is equipped with multiple sensors, including light detection and ranging (LiDAR), radio detection and ranging (radar), inertial measurement units (IMUs), and global navigation satellite system (GNSS) receivers. These sensors jointly provide spatial perception, motion estimation, and environmental awareness necessary for cooperative tracking tasks.

LiDAR sensors are commonly used to obtain high-resolution spatial measurements by emitting laser pulses and measuring the return time of reflected signals. The resulting point cloud data provides detailed geometric information about surrounding targets and obstacles. However, LiDAR measurements may be affected by environmental conditions such as sea spray, fog, and reflective water surfaces, which introduce noise and incomplete observations. Radar sensors provide complementary sensing capabilities through electromagnetic wave reflections. Compared with LiDAR, radar systems are robust to adverse weather conditions and can detect objects at long ranges. Although radar measurements typically have lower spatial resolutions than LiDAR data, their robustness in harsh environments makes them valuable for maritime sensing tasks. In addition to exteroceptive sensors, proprioceptive sensors such as IMUs and GNSS receivers provide motion and localization information for the USV platform. IMUs measure angular velocity and linear acceleration to estimate the dynamic state of the vehicle, while GNSS receivers provide global position estimates that support navigation and multiagent coordination.

The combined sensing observations can be represented by a measurement model that relates the true environmental state to sensor outputs. The target state at time t is defined as

$$\boldsymbol{x}_t = \left[p_x, p_y, v_x, v_y \right]^T, \tag{1}$$

where p_x and p_y represent the target position in the horizontal plane, and v_x and v_y are velocity components. The observation obtained from the sensing system is modeled as

$$\boldsymbol{z}_t = h\left(\boldsymbol{x}_t\right) + \boldsymbol{n}_t, \tag{2}$$

where $\boldsymbol{z}_t$ denotes the measurement (what the sensor actually reports), $h(\boldsymbol{x}_t)$ is the observation function that maps the true state into what the sensor can detect, and $\boldsymbol{n}_t$ denotes the noise added to the measurement, representing imperfections in sensing.

In marine environments, the noise originates from multiple sources such as wave-induced platform motion, wind disturbances, sensor interference, and environmental reflections. Sensor noise $\boldsymbol{n}_t$ is commonly modeled as a Gaussian distribution as

$$\boldsymbol{n}_t \sim N\left(0, \Sigma\right), \tag{3}$$

where Σ represents the covariance matrix describing measurement uncertainty.

For multimodal sensing systems, the observation vector is expressed as

$$\boldsymbol{z}_t = \left[\boldsymbol{z}_t^{LiDAR}, \boldsymbol{z}_t^{Radar}, \boldsymbol{z}_t^{IMU}, \boldsymbol{z}_t^{GNSS} \right], \tag{4}$$

where each component corresponds to the measurement from LiDAR, radar, IMU, and GNSS. The heterogeneous observations must be integrated to support reliable perception and target tracking.

2.2 Bandwidth and computational constraints on sensing nodes

In real-world maritime deployments, USVs operate under strict hardware limitations. Each vehicle functions as an autonomous sensing node that collects, processes, and exchanges sensor data with other agents. However, transmitting all raw sensor measurements across the network is often impractical owing to communication bandwidth limitations and processing delays.

High-resolution sensors such as LiDAR generate large volumes of data at high sampling rates. Processing these data streams in real time may exceed the computational capacity of onboard embedded processors. In addition, cooperative multi-USV missions require the sharing of sensing information among agents, which further increases communication overhead.

When the available communication bandwidth is represented as B and the total size of transmitted sensor data at time t is represented as D_t, their relationship is expressed as

$$D_t \leq B. \tag{5}$$

Similarly, onboard computational resources can be described by the processing capacity C and the required computational load for sensor processing L_t, which must satisfy

$$L_t \leq C. \tag{6}$$

Such a relationship implies that only a subset of sensor observations can be processed or transmitted at each time step. Efficient data selection and prioritization mechanisms are therefore necessary to maintain real-time perception performance.

2.3 Materials and packaging considerations for marine sensor nodes

Marine sensor nodes on USVs must operate reliably under saltwater corrosion, high humidity, temperature variations, and wave-induced vibration. Therefore, appropriate material selection and packaging are essential to protect sensing performance and extend service life. Sensor housings are mostly made from corrosion-resistant materials such as stainless steel, marine-grade aluminum alloys, or composite materials, combined with sealing structures, such as O-rings and gaskets, to prevent moisture ingress. For optical sensors such as LiDAR, protective windows should maintain optical transparency while resisting abrasion and salt deposition; tempered glass or optical-grade polymers are typically used. For radar sensors, dielectric radomes are required to minimize electromagnetic attenuation while maintaining mechanical strength, and composite dielectric materials are often adopted.

Thermal management is also important for stable sensor operation. Heat-conductive enclosure designs and an appropriate internal layout help dissipate heat generated by continuous sensing and computation. Overall, robust materials and packaging improve the durability and reliability of marine sensing nodes and support long-term multi-USV deployments in harsh ocean environments.

3. Problem Formulation

Game theory is applied in DyMAN to enhance decision-making in dynamic multitarget tracking with multiple USVs within a cooperative marine sensing network (Fig. 1). In the complex network of multi-USV operations, achieving Nash equilibrium is a fundamental challenge. Nash equilibrium is achieved when no individual USV improves its outcome by unilaterally changing its strategy, provided that the strategies of all other USVs remain fixed. The equilibrium strategy σ_i^* for each USV i is defined as

$$\sigma_i^* = \underset{\sigma_i}{\operatorname{argmax}}\, \pi_i\left(\sigma_i, \sigma_{-i}^*\right), \tag{7}$$

where $\arg\max_{\sigma_i}$ is the mathematical operator for the argument of the maximum to find the specific value of the possible strategy or action σ_i that results in the highest possible value for the

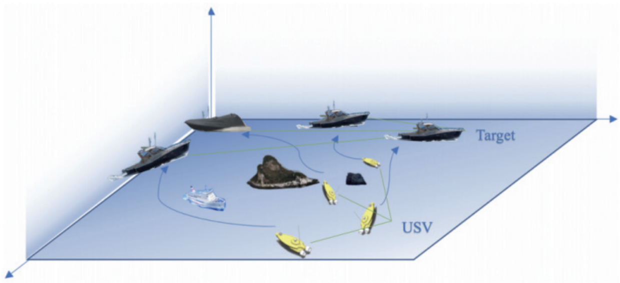

Fig. 1. (Color online) Multi-USV cooperative sensing and communication system.

function π_i (payoff or outcome function). The outcome function represents the expected cumulative payoff, which is a composite utility reflecting tracking accuracy, collision avoidance, and bandwidth conservation in DyMAN. This outcome is linked to the actions of neighboring agents. Therefore, the Nash equilibrium represents a steady state where each USV has optimized its sensor modality selection and movement trajectory in a way that is robust to the dynamic sensing strategies of the rest of the fleet. In sensing in marine environments, the payoff is related to sensing reliability, target tracking accuracy, and the efficient utilization of sensor observations. σ_{-i}^* is the strategy of others. The asterisk indicates that these other agents are also playing their equilibrium strategies.

Achieving Nash equilibrium in multi-USV operations is complicated by incomplete information and the dynamic nature of the maritime environment. In addition, the uncertainty and noise associated with multimodal sensor measurements further complicate cooperative decision-making. These factors make it difficult for individual vehicles to anticipate the strategies of others and to maintain stable cooperation. To address this challenge, cooperative game theory is used for equitable resource and reward distribution. Game theory is a mathematical framework used to model strategic interactions between multiple decision-makers, where the outcome for each participant depends on their actions as well as the actions of others. In this study, game theory was applied to address the challenge of multiagent coordination. Specifically, since multiple USVs share limited communication bandwidth and sensing resources, their sensor prioritization strategies are interdependent. By modeling this interaction as a cooperative game, it is possible to achieve Nash equilibrium, where each USV's sensor selection is optimized relative to the strategies of the rest of the fleet, ensuring stable and efficient collective tracking.(22)

In distributed sensing systems, these resources include sensing coverage, communication bandwidth, and computational capacity among multiple USVs. In particular, the Shapley value is used to enable the coalition and allocate the gains from cooperation in proportion to each USV's contribution. By ensuring that every participating vehicle's role is valued, the Shapley value is used to ensure collective performance and sustain collaboration among autonomous agents. The total payoff that any coalition S of players guarantees is given as

$$\sum_{i \in S} x_i \geq v(S), \forall S \subseteq N, \tag{8}$$

where $\sum_{i \in S} x_i$ is the sum of all individual gains allocated to the members of coalition S, S is the specific coalition or subset of USVs within the set of all participating autonomous agents N, x_i is the individual gain or payoff allocated to USV i, and $v(S)$ is the function representing the total value or payoff that coalition S achieves. Equation (8) states that the sum of the payoffs x_i for all players in S of N must be as high as the $v(S)$ value, which is the total payoff that the group S ensures through communication. In practical marine sensing scenarios, this cooperative objective enables multiple USVs to share sensing information and improve overall situational awareness.

The objective of each USV is, then, to track targets in communication with other USVs while efficiently utilizing sensing data, communication bandwidth, and onboard computational resources.

4. DyMAN Algorithm

On the basis of the principles of centralized training and decentralized execution and classical game theory, we introduced DyMAN and the multimodal bandwidth-aware reward structure. Through the integration of a hard-attention mechanism, sensor modalities based on real-time bandwidth constraints are modeled as a cooperative game among USVs for maritime target tracking.

The development of DyMAN begins with the design of an actor–critic architecture, which is the foundation for stable decision-making. In RL, the actor–critic architecture combines policy-based learning (the actor) and value-based learning (the critic). The critic acts as a stable baseline, significantly smoothing the learning process. The actor–critic architectures naturally handle continuous action spaces.[23] In the context of distributed marine sensing systems, the DyMAN framework processes multimodal sensing observations collected by multiple USVs and transforms them into cooperative decision policies for target tracking tasks.

Building on this structure, strategies are optimized and objectives are adjusted to enhance adaptability in dynamic environments. A hard attention mechanism is then integrated to concentrate computational resources on the most critical targets and state variables derived from sensor observations to improve the efficiency of tracking under sensor noise, sensing bandwidth, and onboard computation constraints. Finally, game-theory-based strategies are applied to ensure rigorous coordination and equitable resource allocation among USVs, culminating in the strategic resolution of the Nash equilibrium within a cooperative marine sensing network.

4.1 Actor–critic and partial observability

In the DyMAN algorithm, an actor–critic architecture is employed to coordinate policy and value functions (μ_{θ_1}) in each decision cycle. The architecture is integrated into the DyMAN algorithm to address partial observability, synchronize decision-making across multi-USV environments, and enhance both local decision quality and overall system efficiency in cooperative operations. Each USV operates as an autonomous sensing node that collects local sensor observations while exchanging summarized sensing information with neighboring agents. The actor network is responsible for action selection, processing the local state s_i of an individual USV while simultaneously incorporating data from the rest of the fleet through cooperative sensing and communication. This dual-input approach ensures that the resulting actions are adaptive to local conditions and comprehensive regarding collective fleet dynamics. The selection process is presented as

$$\mu_{\theta_1}\left(s_i,\vec{s}\right)\rightarrow a_i, \tag{9}$$

where μ_{θ_1} represents the policy function, parameterized by θ_1, which dictates the mapping from observed states to a specific action, s_i represents local sensing observations collected by USV and the immediate environment, $\vec{s}$ represents the aggregated state information collected from other USVs in the fleet, facilitating a global perspective for decentralized agents, and a_i is the resulting action selected by USV i for the current decision cycle.

In cooperative marine sensing environments, each USV has access only to partial observations owing to sensing range limitations, environmental disturbances, and communication delays. In multi-USV sensing systems managed by the DyMAN algorithm, to ensure partial observability, the critic network evaluates the immediate and long-term effects of actions taken by individual USVs on the overall efficiency of the fleet. This multilevel feedback mechanism is essential, where each USV receives only limited sensor observations from the environment, ensuring that decision-making remains coordinated and effective despite incomplete data.

$$Q_{\phi_j}\left(s_1,\ldots,s_N,a_1,\ldots,a_N\right)\rightarrow Q_i \tag{10}$$

Here, Q_{ϕ_j} is the global action-value function (or critic network), evaluating the overall quality of the joint actions taken by the entire fleet. (s_1, ..., s_N) are the global states of all individual USVs in the system constructed from distributed sensing observations. While individual agents have limited environmental information, the critic uses the full set of states to ensure coordinated decision-making. (a_1, ..., a_N) are the joint actions performed by all USVs in the fleet. Q_i is the localized feedback or value derived from the global function for USV i. It allows the training of the critic network to translate collective efficiency into individual guidance, ensuring that each agent's decisions remain effective despite having incomplete sensing data.

In the training of the critic network, the global action–value function is used to derive localized behavioral strategies. This process strengthens the system's ability to manage partial observability and enhances the robustness of policies in distributed marine sensing operations.

4.2 Hard attention mechanism

In response to the complexity of dynamic multitarget tracking in multi-USV operations, the DyMAN algorithm integrates a hard attention mechanism. This mechanism is designed to prioritize informative sensor observations collected from distributed sensing nodes while suppressing redundant or noisy sensing data. By maintaining the most relevant inputs, it enhances the accuracy of target recognition and accelerates the decision-making processes of USVs.

In cooperative marine sensing systems, multiple USVs simultaneously collect heterogeneous sensor data such as LiDAR measurements, radar detections, and navigation states. Processing all sensing information in real time may exceed onboard computational and communication capacities. Therefore, the hard attention mechanism selectively focuses on the most informative sensing states for cooperative tracking.

The hard attention mechanism is implemented using a probabilistic selection function $g: S \rightarrow P(U)$ to map the current states to a probability distribution over all USVs U. In each decision cycle, a subset of USVs, $U' \subseteq U$, is selected for focused processing based on the probabilities calculated using g.

$$U' = \left\{ i \in U : \xi_i < g\left(s, i\right) \right\} \tag{11}$$

Here, U' is the selected subset of USVs for focused processing and attention during the current decision cycle. ξ_i is the random variable sampled from a uniform distribution between 0 and 1. This stochastic element ensures the exploration of different strategies during the training phase and helps the algorithm avoid becoming trapped in local optima. $g(s, i)$ is the selection function that calculates the probability of USV i based on the current environmental state s derived from sensor observations.

The selection function g is parameterized by a neural network with parameters that are optimized to maximize the expected cumulative reward of USVs.

$$\theta^* = \text{argmax}_\theta E_{s\sim p, a\sim \pi}\left[R\left(s, a\right) \right] \tag{12}$$

Here, θ^* is the optimal set of parameters for the neural network that defines the selection function and argmax_θ is a mathematical operator to find the values of θ that result in the highest value for the following expression. $E_{s\sim p,a\sim \pi}$ is the expected average outcome calculated over the distribution of states s (following the transition probability p) and actions a (following the policy π). $R(s,a)$ is the reward function, which represents the utility or payoff associated with taking action. This reward reflects the decisions of the selected subset of vehicles U'.

Attention weights are updated using a gradient descent approach, as illustrated in Fig. 2. Through iterative training, the attention mechanism learns to allocate computational resources to the most informative sensing nodes and observations, thereby improving the efficiency of distributed marine sensing and cooperative target tracking.

The update rule is defined as

$$\Delta \psi_i = \frac{\partial L}{\partial \psi_i}, \tag{13}$$

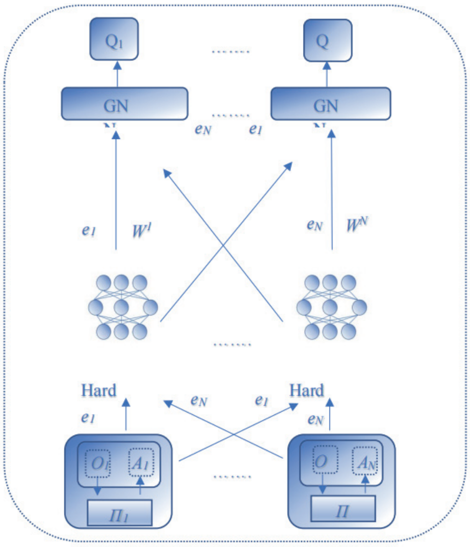

Fig. 2. (Color online) Attention network in hard attention mechanism.

where $\Delta\psi_i$ represents the change in attention weight, L is the learning rate, and $\partial L / \partial \psi_i$ is the partial derivative of L with respect to the weights ψ_i. The weights are integrated into the decision models of USVs, affecting action selection [Eq. (14)].

$$a_i(t) = \mu_{\theta_i}(s_i, t), w_{ij}(t) \tag{14}$$

Here, $a_i(t)$ is the action chosen by USV i at time t, $s_i(t)$ is the state of USV i at t, and $w_{ij}(t)$ is the attention weight at t, which adjusts sensitivity based on environmental parameters and the collective state to provide focused responses to significant targets. This method effectively adjusts the sensitivity of attention based on environmental parameters and the collective state, providing a focused response to significant targets.

4.3 Game-theory-based optimization

To enhance operational efficiency and reduce energy consumption in complex marine environments, the DyMAN algorithm incorporates game theory to select strategic interactions and allocate resources among multiple USVs. In distributed marine sensing missions, each USV functions as an autonomous sensing node that collects environmental observations and cooperates with other agents to achieve reliable target perception. Each USV is treated as an independent player in a strategic game to balance task execution rewards with energy usage, formalized through the following utility function:

$$\max_i E_{r_i, s_i'}\left[\sum_{t=0}^{T} \gamma t r_i(t) - e_i(t)\right], \tag{15}$$

where, $r_i(t)$ is the reward for task execution at t, which reflects the effectiveness of sensing-driven target tracking, $e_i(t)$ denotes the energy consumption of USV i at time t, and γ represents the discount factor for future rewards and serves as a balance between energy costs and task rewards during long-term sensing operations.

This strategic model achieves Nash equilibrium, ensuring that no USV can improve its outcome by unilaterally changing its strategy. The equilibrium is maintained through the following policy update rule:

$$\Delta\theta_i = \alpha \nabla_{\theta_t} J_i\left(\theta_t, \theta_{-t}\right), \tag{16}$$

where, α is the learning rate, $J_i(\theta_t, \theta_{-t})$ is the expected utility of USV i, given its current strategy θ_t and the strategies of all other agents θ_{-t}, considering the strategies of others, and ∇_{θ_t} denotes the gradient of the utility function with respect to the strategy parameters. Through iterative policy updates, each sensing node adapts its strategy to maximize cooperative sensing performance while minimizing energy consumption.

This game-theoretical framework facilitates effective communication among USVs by dynamically adjusting strategies in response to evolving operational contexts and the actions of other USVs within a cooperative marine sensing network.

4.4 Adaptive reward function

In multi-USV systems, an adaptive reward function optimizes both individual and collective behaviors. For sensing-driven tracking tasks, the reward function evaluates how well each USV utilizes sensing information to maintain reliable perception and cooperative tracking.

$$R\left(s, a, s'\right) = R_t + \lambda R_d + \gamma R_c + \delta R_a \tag{17}$$

Here, $R(s, a, s')$ is the total reward when a USV takes action, R_t is the tracking reward, R_d is the distance reward, R_c is the coordination reward, R_a is the attention reward, and λ, γ, and δ are scalar weights used to tune the importance of each reward component. λ is used to control the effect of distance-based behavior, γ is used to adjust the emphasis on coordination, and δ is used to scale the impact of attention-driven decisions on sensing data selection.

R_t is calculated using Eq. (18), where N_t is the number of targets successfully tracked by the USVs at any given time and N is the total number of available targets. This ratio, multiplied by 100, scales the reward to a percentage, making it intuitive and directly tied to performance metrics.

$$R_t = \frac{N_t}{N} \times 100 \tag{18}$$

A higher value indicates better alignment or proximity to the target.

R_d denotes the distance reward for spatial efficiency, such as minimizing unnecessary movement or maintaining optimal formation distances between USVs. R_d leads to the minimization of operational costs and enhances fuel efficiency by penalizing excessive distances traveled by USVs.

$$R_d = -\sum_{i=1}^{N}\left(\frac{d_i}{D}\right) \tag{19}$$

Here, d_i is the distance traveled by the ith USV to its assigned target and D is a normalization factor typically set as the maximum expected operating range, ensuring that the penalty is proportional and consistent across various operational scales.

R_c is the coordination reward to promote cooperative behavior among USVs, rewarding synchronized actions, conflict avoidance, or shared goal achievement. R_c is used to foster teamwork among USVs, preventing collisions and ensuring a collaborative approach to target tracking.

$$R_c = \sum_{i=1}^{N}\max\left(0, d_{min,i} - d_{threshold}\right) \tag{20}$$

Here, $d_{min,i}$ is the minimum distance maintained by the ith USV from its nearest neighbor, and $d_{threshold}$ is the safe operational distance threshold. The reward becomes negative if USVs come too close, thus penalizing potentially risky maneuvers that could lead to operational hazards.

R_a is the attention reward that reflects the effectiveness of the attention mechanism, rewarding the agent for focusing on high-utility features or critical targets. R_a is used for a proactive collision avoidance strategy, which is crucial for maintaining safety and operational integrity.

$$R_a = \sum_{i=1}^{N}\max\left(0, d_{safe,i} - d_{min,i}\right) \tag{21}$$

Here, $d_{safe,i}$ is the predefined safe distance threshold and $d_{min,i}$ is the actual minimum distance between the ith USV and any other USV or obstacle. A positive reward is given for maintaining or exceeding the safe distance, incentivizing the avoidance of close encounters. The framework flow of adaptive rewarding is shown in Algorithm 1.

Algorithm 1 presents the proposed training procedure for the DyMAN framework. While the structure is based on the centralized training and decentralized execution paradigm,[24] the attention-based observation filtering and the game-theoretic weight adjustment were adopted to optimize multi-USV sensing under bandwidth constraints.

Algorithm 1
Training Algorithm for the DYMAN Framework Enhanced with Hard Attention Mechanism and Game Theory Integration.

```
Input: Actor architecture A, Critic architecture C, Attention mechanism T
Input: Batch size B, Number of epochs E, Steps per episode S
Input: Action noise δ, Number of USVs N
Input: Actor weights θ_a, Critic weights θ_c, Attention weights θ_t
Input: Replay buffer capacity D_cap, Exploration steps η, Attention threshold ε, Discount factor γ
Input: Game theory trade-off factor λ

Initialize: Observations O_n
Initialize: Actor network A(θ_a), Critic network C(θ_c)
Initialize: Target Actor network A'(θ_a'), Target Critic network C'(θ_c')
Initialize: Attention network T(θ_t), Replay buffer D (capacity D_cap), Exploration counter η

for episode = 1 to E do
  for step = 1 to S do
    if step < η then
      Each USV i randomly selects an action a_i
    else
      Each USV i selects an action a_i according to:
        if T(O_i, θ_t) > ε then
          a_i = A(O_i, θ_a)
    end if

    Execute actions a_n, observe reward r_n and new state S_n
    Store transition (O_n, a_n, r_n, S_n) in D

    if step > B then
      Sample a batch of transitions of size B from D
      For each sampled transition i:
        Compute target Q-value:
          y_i = r_i + γ · C'(s_i', A'(s_i', θ_a'), θ_c')

        Update Critic C to minimize loss:
          L = (1/2) Σ |y_i − C(s_i, a_i, θ_c)|^2

        Update Actor A to maximize expected return:
          maximize Σ C(s_i, A(s_i, θ_a), θ_c)

        Update Attention network T according to attention scores
        Apply game theory updates to adjust θ_t towards Nash equilibrium

      Soft update target networks A' and C' at rate τ:
        θ_a' = τ · θ_a + (1 − τ) · θ_a'
        θ_c' = τ · θ_c + (1 − τ) · θ_c'
    end if
  end for
end for
```

5. Results and Discussion

5.1 Configuration of simulation

The simulation environment was designed to replicate complex oceanic conditions characterized by time-varying disturbances and maneuvering targets. In particular, the simulation includes sensing uncertainty and environmental disturbances that affect marine sensor measurements. We utilized a dynamic fleet configuration to rigorously test the scalability and adaptability of the DyMAN algorithm within a cooperative marine sensing framework (Table 1).

The sensing environment includes simulated multimodal observations representing typical marine sensing systems such as radar detection, LiDAR perception, and navigation states. Environmental disturbances introduce sensing noise and measurement uncertainty, allowing the evaluation of DyMAN under realistic sensing conditions.

5.2 Learning performance

Figure 3 illustrates the comparative reward performance of the DyMAN algorithm against baseline methods. The learning efficiency and stability of DyMAN were evaluated using cumulative reward metrics and training loss curves derived from sensing-driven tracking performance. As shown in Fig. 3, DyMAN outperforms baseline algorithms such as multiagent deep deterministic policy gradient (MADDPG), deep deterministic policy gradient (DDPG), and proximal policy optimization (PPO), achieving higher total rewards by reducing fluctuations and maintaining robust stability throughout training. The model's convergence rate shows that the integration of attention mechanisms significantly optimizes strategy adaptation in dynamic multi-USV environments by prioritizing informative sensing observations and suppressing noisy measurements.

DyMAN achieves a notable reduction in average total distance and demonstrates superior stability during the later stages of training. This improvement is primarily attributed to the ensemble hard attention mechanism, which directs USVs' computational resources toward high-value environmental features and critical target states derived from sensor observations. By filtering out nonessential sensing data, the mechanism accelerates the learning process and enhances policy responsiveness under complex operational conditions.

Figure 4 illustrates the training loss trajectory of the DyMAN algorithm. A smooth and consistent downward trend indicates the effective optimization of joint policies across the

Table 1
Parameters for experiments.

Parameter	Value/description
Number of targets	4–7 (dynamically adjusted)
Number of USVs	4–7 (dynamically adjusted)
Simulation area	1000×1000 m^2
Learning rate (actor)	0.001 (adaptive)
Learning rate (critic)	0.0001 (adaptive)
Discount factor (γ)	0.95
Batch size	128
Actor network architecture	3 layers, 300 neurons each
Critic network architecture	4 layers, 300 neurons each
Attention network architecture	2 layers, 200 neurons each
Total training episodes	1000
Steps per episode	500
Reward function modifiers	Target dynamics and environmental noise
Environmental noise handling	Sensor noise mitigation through integrated adaptive filtering

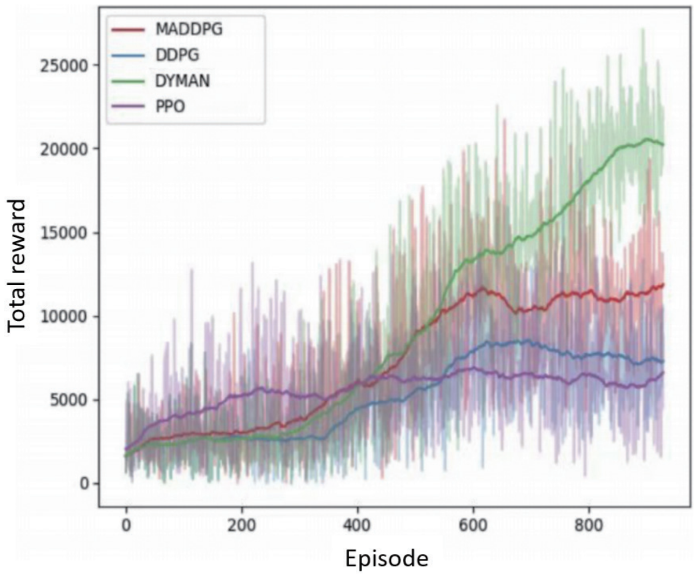

Fig. 3. (Color online) Reward comparison over episodes for various algorithms.

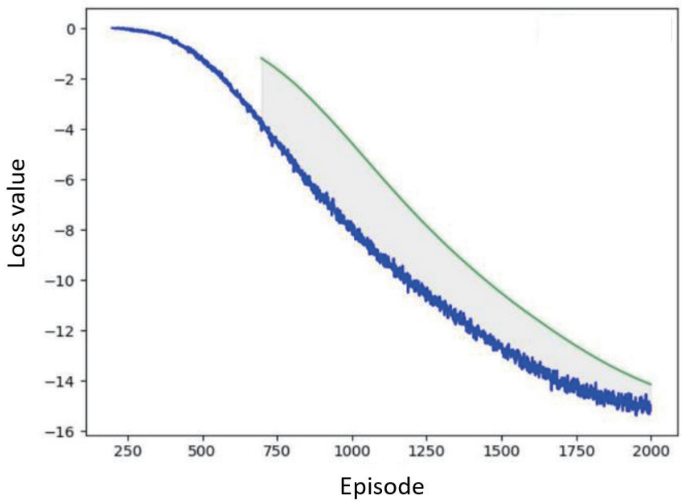

Fig. 4. (Color online) Algorithmic loss metrics over iterative episodes (green line: smoothed loss curve).

multi-USV system. This stability reflects the algorithm's ability to manage the nonstationarity inherent in multiagent environments while maintaining reliable sensing-driven perception. The integration of a centralized critic ensures accurate gradient computations, while the ensemble hard attention mechanism facilitates precise parameter updates by directing computational focus toward critical sensor-derived state variables. Moreover, the incorporation of game-theoretic principles enables managing strategic interactions among USVs. By simulating each USV as an independent player within a cooperative setting, DyMAN promotes convergence toward Nash equilibrium, reducing decision conflicts and eliminating redundant strategies. Such dynamic adjustment of policies, combined with continuous performance evaluation, results in a decline in training loss and significantly enhances the collaborative efficiency of the fleet in distributed marine sensing missions.

5.3 Tracking accuracy and navigational efficiency

Tracking accuracy was evaluated by measuring the progressive distance between the USV fleet and dynamic targets. As shown in Fig. 5, DyMAN reduces tracking error and positioning variance compared with traditional multiagent deep RL. This improvement is attributed to the hard attention mechanism, which enables USVs to generate efficient trajectories by filtering out sensory noise and focusing on trajectory-relevant cues. By concentrating on essential environmental features, DyMAN minimizes unnecessary detours and positioning errors, thereby improving overall tracking performance. In addition, DyMAN's continual learning capability allows rapid adaptation to target motion changes, enhancing action prediction and adjustment accuracy. The integration of game-theoretic principles further strengthens coordination among multiple USVs during joint tracking, ensuring efficient fleet-wide performance.

Energy consumption analysis results reveal a sharp initial decline, reflecting the agents' rapid ability to eliminate erratic movements and redundant control actions. As shown in Fig. 6, energy consumption drops significantly at the beginning of training, indicating early policy improvements that reduce unnecessary actions and energy waste. Although fluctuations occur in the raw curve, the overall trend gradually stabilizes with a slight increase. In episodes 750–850 (the zoomed area), energy consumption still varies, but the smoothed curve demonstrates significant stabilization. This suggests that DyMAN achieves consistent coordination and refined propulsion management as training progresses. Improved energy efficiency is critical for long-duration missions and resource-constrained deployments.

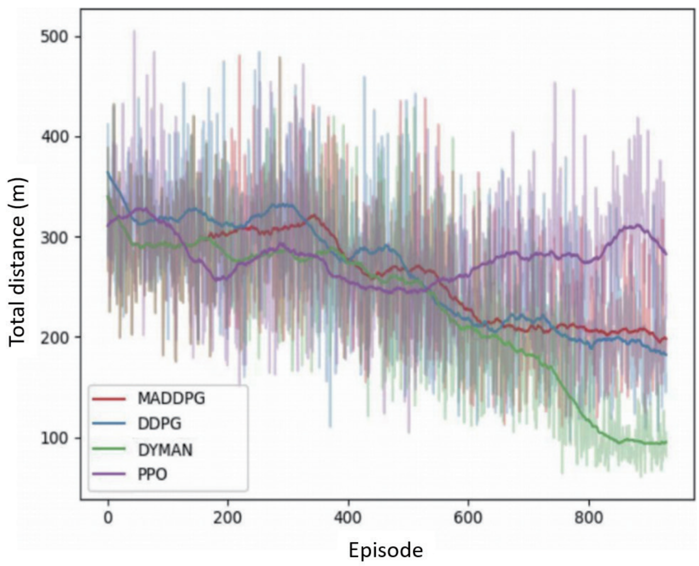

Fig. 5. (Color online) Progressive distance-to-target metrics: a comparative study of various algorithms.

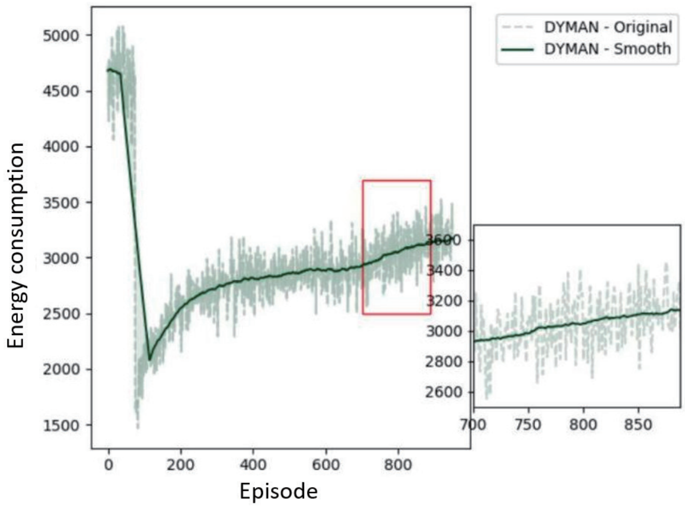

Fig. 6. (Color online) Energy consumption.

The collision rate exhibits a general downward trend as training progresses, validating DyMAN's ability to optimize collision avoidance strategies (Fig. 7). The DyMAN algorithm progressively reduces collisions by focusing on critical state features and incorporating global fleet information during interactions. In the zoomed interval between episodes 800 and 900, minor fluctuations are observed, but the overall trend remains downward. This result highlights DyMAN's capability to enhance multi-USV coordination and ensure safe operations in dynamic environments through effective decision-making and efficient interaction management.

The minimum interagent distance serves as an important indicator of operational safety and coordination. Early in training, the minimum distance varies widely, with some episodes exceeding 80 units, reflecting unstable or inconsistent behaviors. As training progresses, the metric decreases and stabilizes around 40 units, particularly after episode 200 (Fig. 6). This stabilization indicates that DyMAN enables the fleet to maintain tighter formations while preserving safe separation thresholds. The observed trend demonstrates improved spatial awareness and coordination among USVs, ensuring both operational efficiency and safety in complex multiagent environments.

5.4 Multiagent coordination and safety

The operational versatility of DyMAN is demonstrated through dynamic tracking renderings (Fig. 8), which show the algorithm's performance across distinct mission phases under cooperative marine sensing conditions.

Figures 8(a) and 8(b) present the fleet converging on targets, with the attention mechanism monitoring sensing and communication constraints to ensure the efficient allocation of resources. The game-theoretic framework supports interaction-aware decision-making, enabling coordinated capture and collection tasks while maintaining shared situational awareness from distributed sensor observations. As shown in Fig. 8(c), DyMAN rapidly transitions to alternative

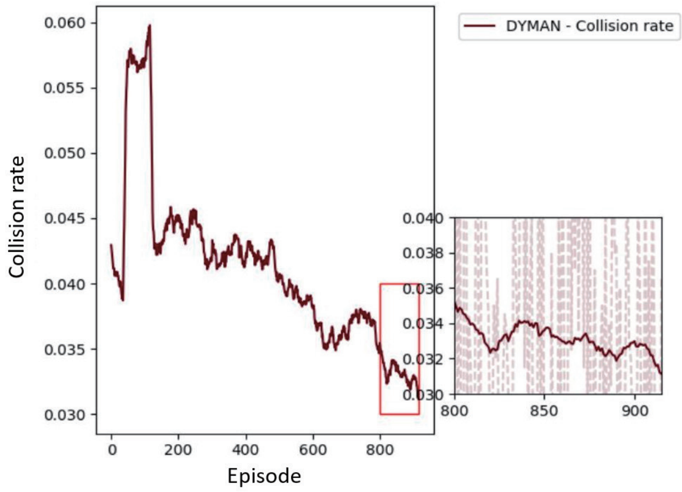

Fig. 7. (Color online) Collision rate.

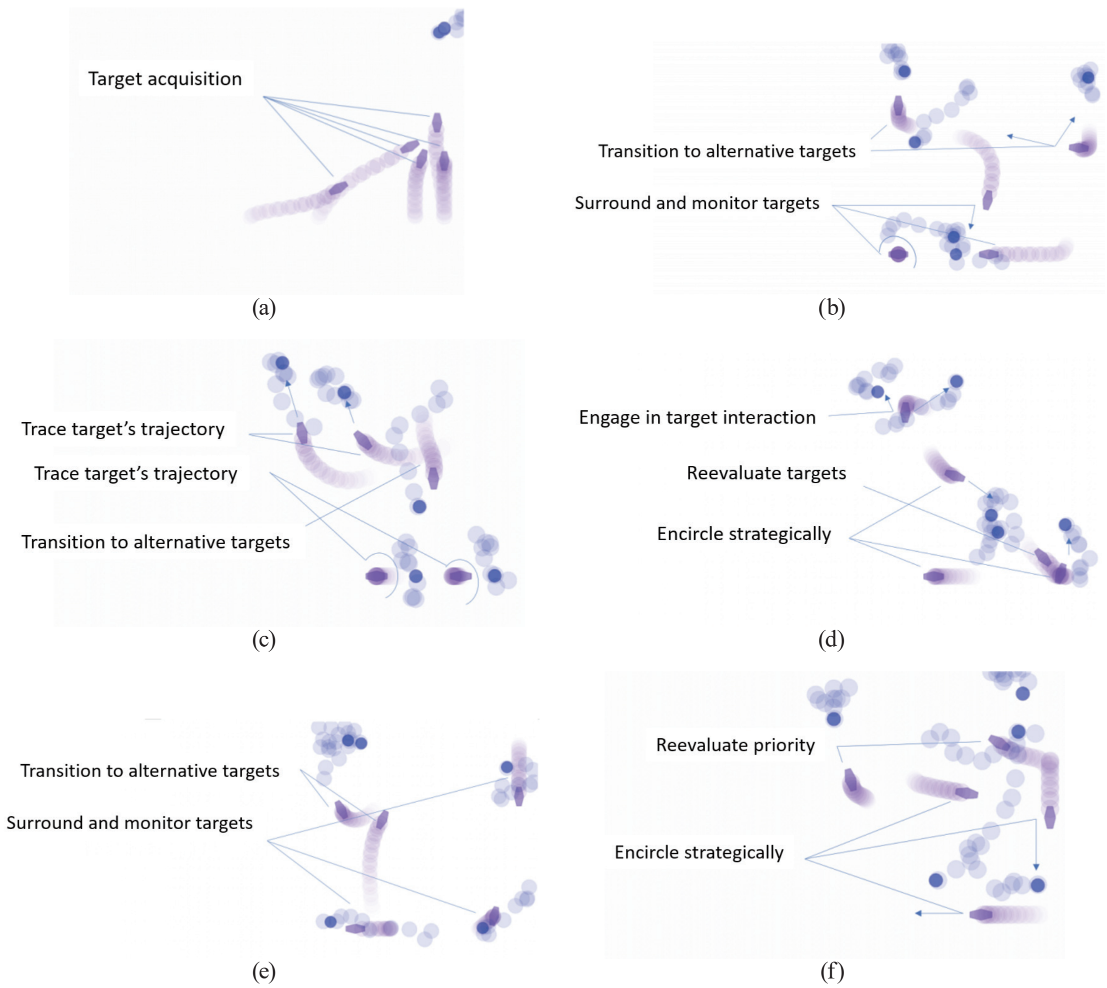

Fig. 8. (Color online) Dynamic tracking renderings. (a) Target acquisition; (b) Transition to alternative targets and surrounding; (c) Trajectory tracking and target transition; (d) Strategic encirclement and interaction; (e) Surrounding and monitoring; (f) Priority reevaluation and encirclement.

targets when mission priorities shift. The attention mechanism evaluates target importance on the basis of sensor-derived cues and reallocates computational resources, allowing fast route updates and effective replanning.

DyMAN predicts target motion on the basis of current position and velocity, focusing computational efforts on trajectory-relevant cues extracted from noisy sensor measurements. The attention mechanism filters sensor measurement noise and redundant sensing information, while the game-theory component guides USVs in adjusting positions for efficient tracking and reduced energy use. During close-range monitoring or collection tasks, the framework evaluates each USV's actions relative to fleet positioning, ensuring safe and effective interaction without interference. As shown in Fig. 9, the minimum interagent distance gradually decreases and stabilizes during training, indicating improved coordination and safety among USVs.

Figures 8(d)–8(f) illustrate DyMAN's ability to encircle dynamic targets. On the basis of game-theoretic principles, the algorithm predicts target behaviors and selects strategies that maximize control while limiting the target's freedom of movement under sensing uncertainty and environmental disturbances. Through refined control commands, USVs form tight formations, maintain continuous monitoring, and coordinate effectively with one another. The attention mechanism further enhances this process by concentrating resources on critical targets and prioritizing informative sensor observations during encirclement and close-range tracking, while game-theoretic optimization ensures the precise adjustment of actions and positions across the fleet.

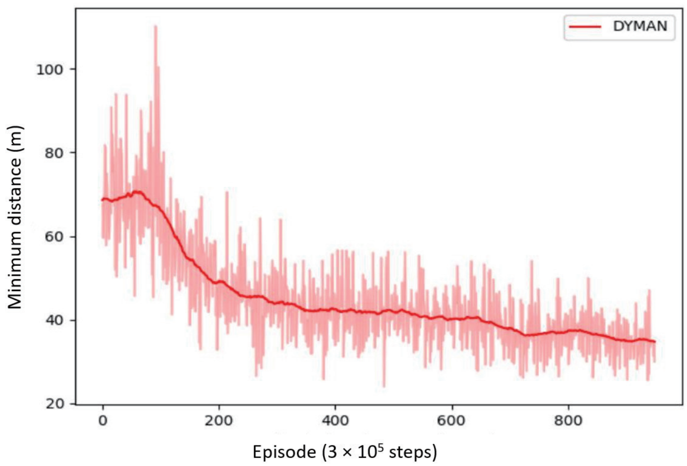

Fig. 9. (Color online) Minimum distance.

6. Conclusion

In this paper, DyMAN was developed as a sensing-oriented cooperative framework for multi-USV dynamic target tracking in uncertain maritime environments. By integrating MARL with game-theoretic coordination and an ensemble hard attention mechanism, the proposed approach addresses key challenges in distributed marine sensing, including coordination stability, computational efficiency, communication limitations, and measurement noise in multimodal sensor observations. Simulation results demonstrate that DyMAN achieves superior learning efficiency and stable convergence compared with baseline methods. The USV fleet exhibits rapid convergence, reduced collision rates, and stabilized formation proximity at a safe threshold of approximately 40 units, ensuring reliable operation in complex maritime scenarios. The integration of Nash equilibrium and the Shapley value provides a principled framework for cooperative strategy formation, resource allocation, and conflict resolution, enabling coordinated maneuvers with optimized energy expenditure and continuous target monitoring. In addition, the hard attention mechanism effectively prioritizes informative sensing observations while suppressing redundant or noisy measurements, thereby reducing computational burden on onboard sensing units and mitigating the effects of noise inherent in maritime sensing hardware such as LiDAR, radar, and sound navigation and ranging (sonar). By linking high-level decision-making with low-level sensor data processing, DyMAN supports the development of intelligent distributed sensing nodes capable of operating under bandwidth and resource constraints while enabling the adaptive fusion of multimodal sensor inputs. These findings highlight the potential of attention-based architectures for advancing intelligent marine sensing systems that dynamically prioritize critical information in complex environments. Future work will focus on hardware-in-the-loop validation using physical USV platforms equipped with multimodal sensor arrays and on the development of adaptive sensing strategies, such as dynamic sampling and bandwidth-aware data prioritization, to further enhance energy efficiency and extend the operational lifespan of marine sensing systems.

References

1 R. Patterson, E. Lawson, V. Udyawer, G. Brassington, R. Groom, and H. Campbell: Front. Mar. Sci. **8** (2022) 2021. https://doi.org/10.3389/fmars.2021.736984
2 N. Wang and H. Xu: IEEE Trans. Veh. Technol. **69** (2020) 6928. https://doi.org/10.1109/TVT.2020.2991220
3 N. Wang, X. Jin, and M. J. Er: Ocean Eng. **184** (2019) 1. https://doi.org/10.1016/j.oceaneng.2019.05.017
4 P. Stone, G. Kaminka, S. Kraus, and J. Rosenschein: Proc. AAAI Conf. Artif. Intell. **24** (2010) 1504. https://doi.org/10.1609/aaai.v24i1.7529
5 Z. Peng, K. Cui, H. Li, N. Gu, L. Liu, and D. Wang: IEEE Trans. Indus. Inform. **20** (2024) 2387. https://doi.org/10.1109/TII.2023.3288217
6 R. S. Sutton and A. G. Barto: Reinforcement Learning: An Introduction (MIT Press, Boston, 2018) 2nd ed., pp. 23–45.
7 V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis: Nature **518** (2015) 529. https://doi.org/10.1038/nature14236
8 H. R Karimi and Y. Lu: Contr. Eng. Prac. **111** (2021) 104785. https://doi.org/10.1016/j.conengprac.2021.104785
9 W. Wu and S. Tong: Ocean Eng. **281** (2023) 114813. https://doi.org/10.1016/j.oceaneng.2023.114813
10 Z. Dong, Z. Zhang, S. Qi, H. Zhang, J. Li, and Y. Liu: Ocean Eng. **270** (2023) 113633. https://doi.org/10.1016/j.oceaneng.2023.113633

11 Y. Ma, Z. Liu, T. Wang, S. Song, J. Xiang, and X. Zhang: Ocean Eng. **295** (2024) 116845. https://doi.org/10.1016/j.oceaneng.2024.116845
12 W. Song and S. Tong: Ocean Eng. **287** (2023) 115700. https://doi.org/10.1016/j.oceaneng.2023.115700
13 H. Nguyen, S. Thudumu, H. Du, R. Vasa, and K. Mouzakis: Mach. Learn. Appl. **22** (2025) 100745. https://doi.org/10.1016/j.mlwa.2025.100745
14 D. P. Kingma and J. Ba: arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980
15 H. Qian, M. Wang, M. Zhu, and H. Wang: Sensors **25** (2025) 6033. https://doi.org/10.3390/s25196033.
16 L. Zhou, S. Leng, Q. Liu, H. Chai, and J. Zhou: IEEE Internet Things J. **9** (2022) 15066. https://doi.org/10.1109/JIOT.2021.3125530
17 S. Ryu and S. Han: Korean J. Remote Sens. **41** (2025) 1225. https://doi.org/10.7780/kjrs.2025.41.6.21
18 J. Foerster, N. Nardelli, G. Farquhar, T. Afouras, P. H. S. Torr, P. Kohli, and S. Whiteson: Proc. 34th Int. Conf. Machine Learning (ICML, 2017) 1146. https://dl.acm.org/doi/10.5555/3305381.3305500
19 X. Wang, L. Ke, G. Zhang, and D. Zhu: Inf. Sci. **669** (2024) 120560. https://doi.org/10.1016/j.ins.2024.120560
20 W. Gan, X. Qu, D. Song, and P. Yao: IEEE Trans. Automat. Sci. Eng. **21** (2024) 5895. https://doi.org/10.1109/TASE.2023.3319510
21 X. Qu, W. Gan, D. Song, and L. Zhou: Ocean Eng. **273** (2023) 114016. https://doi.org/10.1016/j.oceaneng.2023.114016
22 W.-J. Liu, X.-Yu Yao, M. Hua, J. H. Park, and M.-F. Ge: Neurocomputing **650** (2025) 130821. https://doi.org/10.1016/j.neucom.2025.130821
23 E. Mohammadi, D. Ortiz-Arroyo, A. A. Hansen, M. Stokholm-Bjerregaard, S. Gros, A. S. Anand, and P. Durdevic: Expert Syst. Appl. **277** (2025) 127180. https://doi.org/10.1016/j.eswa.2025.127180
24 R. Lowe, Y. Wu, A. Tamar, J. Harb, O. P. Abbeel, and I. Mordatch: Proc. 31st Int. Con. Neural Information Processing Systems (NIPS, 2017) 6382. https://dl.acm.org/doi/10.5555/3295222.3295385