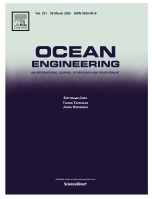
Ocean Engineering 347 (2026) 123947

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



Research paper

SeaTraNet: A local-global feature fusion network for abnormal behavior recognition of single trawler

Lingkai Kong [a], Zhuhua Hu [a,*], Yaochi Zhao [b], Wei Wu [a], Yanming Gu [a]

[a] School of Information and Communication Engineering, Hainan University, 570228, Haikou, China
[b] School of Cyberspace Security, Hainan University, 570228, Haikou, China

ARTICLE INFO

Keywords:
Single-trawler
AIS trajectory
Dual-path
Anomaly behavior recognition

ABSTRACT

To recognize single-trawler anomalous behaviors during fishing moratoriums, SeaTraNet is proposed as a novel trajectory anomaly recognition model that integrates both local and global features. SeaTraNet adopts a dual-path architecture: the local path employs Dual Residual Gated Convolution (DResGConv) to capture short-term abrupt changes, while the global path leverages a AIS-Encoder to model long-range behavioral patterns, enabling effective characterization of trawler temporal dynamics. A dedicated dataset, TrawlTag, is also constructed. A small set of representative samples is first manually annotated, after which a dual-modal Random Forest classifier expands the dataset by generating high-confidence pseudo-labeled samples, followed by rapid manual verification, significantly reducing annotation costs. Experiments conducted on TrawlTag and two additional real-world datasets demonstrate the effectiveness of the proposed method, with SeaTraNet outperforming eight representative baseline models.

1. Introduction

Frequent fishing activities pose potential risks ranging from national security threats and maritime safety incidents to economic disruptions and ecological degradation (Shahir et al., 2019). To safeguard the marine environment and fishery resources while promoting ecological sustainability, Hainan Province enforces an annual fishing moratorium, during the closed season, all fishing activities except those involving angling gear are officially suspended. Single-trawler, which operate by towing nets through the water column or along the seabed, are prohibited from fishing during the moratorium (He et al., 2021). However, due to certain unavoidable factors, many of these vessels continue to engage in illicit fishing activities during the ban period. Trawling operations of the single-trawler can cause severe disruption to the marine ecosystems and in conflict with current ecological protection policies. However, due to their complex behavioral patterns, single-trawler vessels are difficult to be recognized in a timely manner, making the effective monitoring and management of their activities a critical challenge for advancing maritime governance and environmental conservation.

Currently, the research on recognizing abnormal behaviors of fishing vessels is generally showing a trend of continuous deepening and rapid development. Early research mainly relied on manual recognition and supervision, although this method is intuitive, it is time-consuming and labor-intensive. With improvements in data acquisition and computing power, researchers have gradually begun to recognize abnormal behavior through domain expert experience, rule-based reasoning, and structured knowledge, or by relying on large-scale trajectory data and remote sensing observations to uncover potential abnormal patterns. However, these methods have significant limitations, they heavily rely on prior knowledge and rules (Nie et al., 2025), making it difficult to maintain universality in the complex and ever-changing maritime environment, and manual feature extraction performs poorly when faced with complex ship trajectory data (Jieyang et al., 2023). In response to these limitations, researchers have increasingly turned to deep learning techniques, which can automatically extract and represent features from large-scale trajectory data (Shaheen et al., 2016), these techniques have gradually been applied to abnormal behavior recognition. Nevertheless, most deep learning methods still focus on general ship behavior, often neglecting the unique operating patterns of specific types of fishing vessels. For example, single-trawler fishing vessels often exhibit behaviors such as prolonged offshore operations, repeated entry into restricted areas, and frequent and drastic changes in course during closed fishing seasons or specific operating areas. These behaviors exhibit both significant local mutations and global trajectory trends, making the unified modeling of short-term behavioral fluctuations and long term behavioral patterns a major challenge. Furthermore, high quality labeled

* Corresponding author.
E-mail addresses: konglingkai2001@163.com (L. Kong), eagler_hu@hainanu.edu.cn (Z. Hu), zhyc@hainanu.edu.cn (Y. Zhao), 21110810000025@hainanu.edu.cn (W. Wu), burry_gu@163.com (Y. Gu).

https://doi.org/10.1016/j.oceaneng.2025.123947
Received 19 July 2025; Received in revised form 27 November 2025; Accepted 7 December 2025
Available online 20 December 2025

trajectory data is extremely scarce, limiting the model's training and generalization capabilities. To address the aforementioned challenges, this study proposes SeaTraNet, a deep learning method for recognizing abnormal behaviors of single trawlers by jointly capturing local behavioral mutations and long range trajectory patterns. In parallel, a semi-supervised dataset TrawlTag, is constructed to alleviate the scarcity of labeled trajectory data and to provide high confidence annotations tailored to single-trawler operations. The main contributions are as follows:

- To address the issue of scarce labeled data in the recognition of abnormal behaviors of single trawler vessels, this study constructs the TrawlTag dataset to maximize data quality under limited cost constraints. A small set of representative trajectories is first meticulously annotated to form a high confidence seed label set. Subsequently, a dual-modal Random Forest fusion model is employed to generate large scale pseudo-labeled samples based on kinematic and behavioral features, thereby expanding the scale and diversity of the training data. Finally, the pseudo labels undergo rapid manual verification, with particular focus on boundary cases and potential misclassifications, ensuring the overall reliability and consistency of the annotations.
- To tackle the core challenge of jointly modeling short term local anomalies and long range global patterns in single trawler trajectories, this paper proposes a novel dual path anomaly recognition method, SeaTraNet. The dual path architecture integrates the DResGConv module with a AIS-Encoder, capturing short term local mutations and long term global dependencies respectively, while incorporating a gating mechanism at the feature fusion stage to achieve adaptive integration of the two feature types. This design substantially enhances the representation and discrimination of complex, non-stationary behavioral patterns.
- Comprehensive experiments are conducted on real AIS datasets, the quantitative results show that SeaTraNet outperforms the existing methods in overall performance, These results demonstrates good practicality and promotion potential, which is suitable for recognizing the behavior of abnormal fishing vessels.

The study is structured as follows: Section 2 systematically reviews the research foundation of recognizing abnormal behaviors of ships. Section 3 provides a formal description of the research problem, Section 4 describes in detail the modeling methodology and key techniques used, Section 5 shows and analyzes the experimental results, Section 6 explores the misrecognition trajectories in the process of model recognition, and Section 7 summarizes the full text of the work and looks ahead to the future directions of research.

2. Related work

In recent years, various methods have been proposed for recognizing abnormal behaviors in maritime vessels, particularly focusing on hybrid knowledge-driven and data-driven methods, traditional machine learning methods, and deep learning methods.

Hybrid knowledge-driven and data-driven methods: Knowledge-driven methods recognize abnormal behaviors by leveraging domain expertise, predefined rules, logical inference, and structured knowledge representations such as knowledge graphs (Wan et al., 2024). Data-driven methods recognize abnormal behaviors by mining patterns from large-scale data sources, including AIS trajectories and remote sensing imagery (Ribeiro et al., 2023). Hybrid methods that integrate both knowledge-driven and data-driven strategies aim to balance structured domain insights with data adaptability, thereby combining the advantages of both methods. For instance, Wan et al. (2023) conducted a graph-based feature analysis of vessel violations within a jurisdiction, leveraging a maritime knowledge graph. By analyzing features associated with certain anomalous vessel behaviors, their method effectively recognized cases of vessel identity fraud and other irregular activities. Rybicki et al. (2024) integrated kinematic features with seasonal characteristics within a unified model, utilizing spatiotemporal clustering of historical data to recognize vessel groups. They further evaluated the proximity of these groups to recent spatiotemporal clusters to determine whether their behaviors conformed to expected patterns or constituted anomalies. Mazzarella et al. (2015) proposed an architecture that employs a knowledge-based Automatic Recognition System (AIS) position projection step to facilitate the fusion of Synthetic Aperture Radar (SAR) ship recognition data with self-reported AIS data.

Traditional machine learning methods: Traditional machine learning based methods typically rely on manual data preprocessing and feature engineering where features relevant to abnormal behaviors are extracted to train models for the recognition and classification of such behaviors (Jordan and Mitchell, 2015). For instance, Wang and Liu (2021) combined the riskiness of ship collision with bridges and the abnormal behavior of ships, and proposed the criteria for determining the abnormal behavior in specific waters. Subsequently, a large amount of trajectory data is constructed by utilizing ship motion equations and random data generation methods. On this basis, a Back Propagation (BP) neural network classification method is used to establish a ship abnormal behavior recognition model. Rong et al. (2024) proposed a machine learning method to recognize abnormal behavior by generating ship parameter curves and applying a sliding window algorithm, this method employ a clustering algorithm to group similar abnormal behavior patterns, and then the clustered data are used to train a random forest model, which can effectively recognize multiple abnormal behaviors within historical AIS data. Wei et al. (2022) used density-based spatial clustering algorithm with noise to recognize the ship traffic patterns by combining spatio temporal and motion features, and designed an improved support model based on the weighted hybrid kernel function and the differential operator, differential operator-based improved support vector machine for abnormal behavior recognition. **Deep learning methods:** Deep learning based methods have been widely used in various fields such as computer vision (He et al., 2025) and natural language processing (Wang et al., 2025). due to their automatic feature extraction capability and powerful feature representation (LeCun et al., 2015). In recent years, in order to better promote the development of the marine field, deep learning methods have also been gradually introduced into the field of abnormal behavior recognition. In particular, the deep learning architecture represented by Transformer (Vaswani et al., 2017), as well as proved to be outstanding in natural language processing (Shahin and Ismail, 2024; Buscaldi et al., 2024; Li et al., 2024a; Oko et al., 2024; Kashi et al., 2025) and time series analysis (Feng et al., 2024; Piao et al., 2024; Kang and Kang, 2024; Lin et al., 2024). The core advantage of Transformer is its Self-Attention mechanism (Liu et al., 2024), which is able to effectively capture global information and long-range dependencies in sequential data without relying on traditional recursive structures. This feature makes it more promising than traditional models in dealing with long time series data, and enables a better understanding of the global context within the trajectories. For instance, Liu et al. (2025) proposed a GSA anomaly recognition model based on the Transformer architecture, which effectively captures the dependency relationship between trajectory features and time series data by introducing the GSA module, thus realizing the accurate recognition of anomalous ship behaviors. Sun et al. (2024) proposed a self-supervised representation learning model SSRLM, which can captures the dependencies between features through a fine-tuning mechanism, and achieves good metrics on the self-made dataset HN_MULMI as well as on the public dataset. Hou et al. (2025) proposed an unsupervised Transformer-VAE model that models temporal relationships through a multi-head self-attention mechanism and used reconstruction error as the criterion. It can effectively recognize various abnormal navigation behaviors of ships and improve maritime traffic safety. Zhang et al. (2024) constructed a TrajBERT-DSSM model that integrates Geohash encoding, TrajBERT trajectory representation, and DSSM semantic matching. By learning the spatiotemporal patterns of trajectories from AIS data

and comparing them with historical tracks in high dimension, the model can accurately predict ship destinations based on the most similar past track destination ports. Based on the research evolution in this field and the insights from existing research, this study constructs a dual path model architecture for the recognition of abnormal behaviors of single trawler vessels. This architecture consists of two functionally complementary representation paths: The first focuses on short-term mutation patterns at the local scale to capture the fine grained dynamic features of abnormal behaviors. The second emphasizes the long sequence trajectory structure at the global scale to model the macroscopic spatiotemporal patterns of fishing vessel operations. Two paths work together to enable the model to simultaneously possess sensitivity to local disturbances and robust perception of the overall situation, thereby achieving more reliable abnormal behavior recognition.

3. Problem formulation

Problem 1: Although the trajectory data of a single trawler can be relatively easily obtained through existing platforms, the annotation work remains costly and time consuming. To improve data utilization efficiency and accelerate model deployment, this study introduces a semi-supervised learning method based on dual modal random forests. This method uses a small amount of labeled samples to learn the feature patterns of each modality (motion features and geographic features) and automatically generates pseudo labels for large amounts of unlabeled data. Samples with consistent predicted labels from the two modalities will be directly added to the pseudo labeled data set for training, while samples with inconsistent predicted results will undergo rapid manual review to ensure the reliability of the labeling. The basic workflow is as follows.

$$\hat{y}_j^{motion} = RF_{motion}(\cdot) \tag{1}$$

$$\hat{y}_j^{geo} = RF_{geo}(\cdot) \tag{2}$$

$$\hat{y}_j = \begin{cases} \text{labeling}, & \text{if } \hat{y}_j^{\text{motion}} = \hat{y}_j^{\text{geo}}, \\ \text{manual review}, & \text{if } \hat{y}_j^{\text{motion}} \neq \hat{y}_j^{\text{geo}} \end{cases} \tag{3}$$

where $\hat{y}_j^{motion}$ and $\hat{y}_j^{geo}$ denote the predicted labels for the j-th trajectory sample obtained from the motion and geographical random forests, respectively. $RF_{motion}(\cdot)$ and $RF_{geo}(\cdot)$ represent the dual modal random forest models trained based on motion features and geographical features. Finally, $\hat{y}_j$ represents the resulting label used for pseudo labeling, if the two modal predictions are consistent ($\hat{y}_j^{motion} = \hat{y}_j^{geo}$), the sample is automatically labeled, otherwise, it is sent for manual review to ensure labeling reliability.

Problem 2: In recent years, with the increasing demand for marine supervision, fishing vessel behavior recognition based on AIS data has received extensive attention. However, most of the existing studies focus on the generalized trajectory analysis of multiple types of fishing vessels, and there is still a lack of specialized modeling for the precise recognition of single trawler during the closed season. Single trawler usually exhibit complex behavioral characteristics combining long-time offshore sailing and high-frequency cyclic operation in specific sea areas during illegal operation. Such trajectories contain both global dependencies at long time scales and embody perturbation variations at short time scales, posing significant challenges to traditional rule-based or single-scale modeling methods. Therefore, this section formalizes the problem as an anomalous behavior recognition task based on AIS trajectory sequences, as defined below:

$$\hat{y} = \text{softmax}\left(W_o \cdot \text{concat}\left(\mathbf{H}_\text{A} \cdot v_1, \ \mathbf{H}_\text{D} \cdot v_2 \right) + b_o \right) \tag{4}$$

the above formula represents the forward calculation process of SeaTraNet. The input AIS trajectory sequence is respectively processed by AIS-Encoder to extract the global sequence feature $\mathbf{H}_\text{A}$ and by DResG-Conv to extract the local convolution feature $\mathbf{H}_\text{D}$. Then, through the gating weights v_1, v_2, the contributions of the two types of features are dynamically adjusted and concatenated along the channel to form a fused representation. Finally, it undergoes a linear transformation W_o, b_o and is output through softmax to obtain the category probability $\hat{y}$, thereby achieving the classification or anomaly recognition of fishing vessel behaviors.

4. Methodology

The SeaTraNet architecture is deeply inspired by three types of classical neural network designs: the Transformer encoder structure, the Convolutional Neural Network (CNN), and the residual connection mechanism.

In the introduction, it is emphasized that the trajectories of illegal single trawlers exhibit cross scale features, in order to utilize these features, this paper designs a dual path architecture, as shown in Fig. 1.

The upper path builds the AIS-Encoder. Firstly, it maps the original AIS sequence to high dimensional features through linear projection. Then, it sequentially feeds these features into Multi-Head Flash Attention-2 and Gated Linear Unit (GLU), and after each sub-layer, it is supplemented with LayerNorm & Add to form a deep self-attention stacking structure, gradually capturing the long range spatiotemporal dependencies of the trajectory. The lower path is constructed by cascading DResGConv layers, focusing on extracting local abnormal features such as mutations and oscillations within short windows. Ultimately, after the dual path features are spliced through Output Splicing, they are dynamically fused through learnable gating weights to obtain a unified representation that balances both global and local aspects. Then, Softmax is applied for classification. Next, this study will conduct an analysis of each of the innovative modules.

4.1. TrawlTag dataset

4.1.1. Semi-supervised learning

Semi-supervised learning refers to a category of supervised learning tasks and corresponding technical methods, which mainly rely on a small volume of labeled data and a large quantity of unlabeled data to conduct model training (Van Engelen and Hoos, 2020). As a critical branch of machine learning, semi-supervised learning lies between unsupervised learning (completely unlabeled data) and supervised learning (completely labeled data).

Recent studies have shown that combining a large amount of unlabeled data with a small amount of labeled data can significantly improve the learning accuracy and generalization ability of the model. For instance, the ReFixMatch (Nguyen and Yang, 2023) fully exploits the potential of unlabeled data by utilizing a bridging mechanism between high and low confidence predictions, thus achieving better performance than existing methods on ImageNet. In addition, the FlexMatch (Zhang et al., 2021) introduces a Curriculum Pseudo Labeling strategy, which dynamically adjusts the thresholds of each category according to the learning state of the model, effectively improving the learning effect in the case of extremely scarce labeled data. These studies emphasize the importance of unlabeled data in semi-supervised learning, particularly in practical application scenarios where labeling is costly or data acquisition is difficult.

Semi-supervised learning assumes that there are l independently and identically distributed samples $x_1, x_2, \ldots, x_l \in X$ and corresponding labels $y_1, y_2, \ldots, y_l \in Y$, and u unlabeled samples $x_{l+1}, x_{l+2}, \ldots, x_{l+u} \in X$. By combining these small number of labeled samples with a large number of unlabeled samples, better classification performance can be obtained compared to dropping unlabeled samples for supervised learning or dropping labeled samples for unsupervised learning, and the process of semi-supervised learning is shown in Fig. 2.

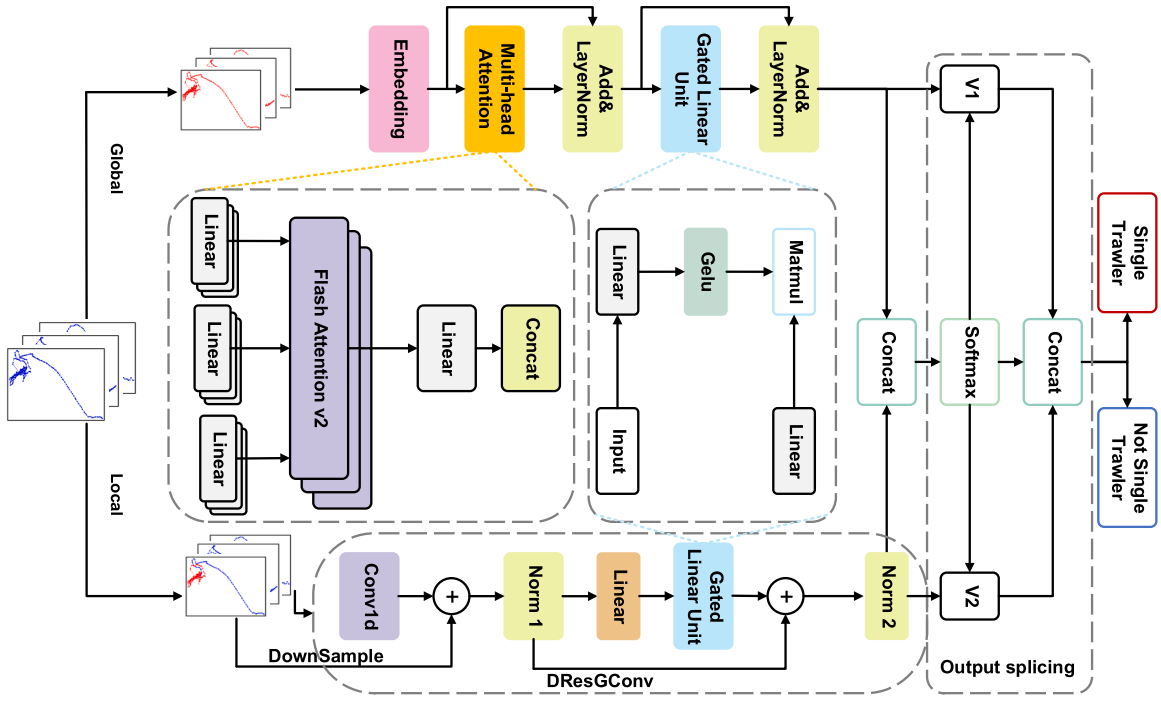


Fig. 1. The architecture of the model.

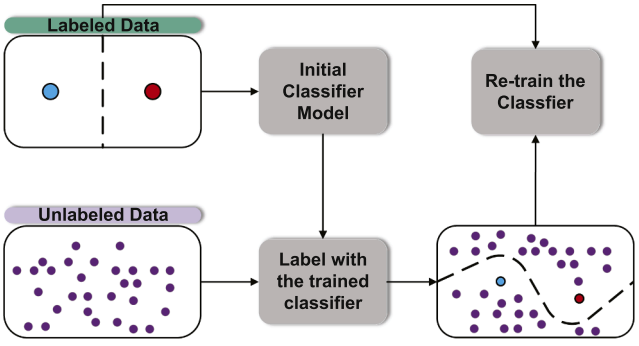


Fig. 2. Process for semi-supervised learning.

This figure illustrates the workflow of semi-supervised learning. The process can be viewed as clustering, where labeled data is used to mark clusters so that the cluster boundaries are away from high-density regions, or one-dimensional manifolds where the learning data is located. Initially, the labeled data (blue and red points) are sparsely distributed, while the unlabeled data (purple points) occupy most of the space, as the iterations proceed, high-confidence pseudo-labeled samples (red points) are progressively integrated, and the decision boundaries (dashed lines) are progressively adjusted to produce high-confidence pseudo-labeled labeled data.

4.1.2. Two modal labeling

To address the scarcity of labeled data for recognizing abnormal behaviors in single trawler vessels, and to reduce labor costs while maintaining the quality of labeling, this study proposed a semi-supervised pseudo-label generation method based on dual-modal random forest, as shown in the Fig. 3. This method fully utilizes the dynamic features (speed and heading) and spatial features (longitude and latitude) of the trajectory.

It conducts classification predictions through two independent random forests (Breiman, 2001), and selects high confidence samples based on probability and category consistency as pseudo labels to be added to the dataset, the process is as follows: First, the random forest model is trained. Let the training dataset be $D = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i represents the feature vector of the sample and $y_i \in \{0, 1\}$ represents the label. Each decision tree $h_t(x)$ is trained on a random subset $\tilde{D}_t \subset D$, and at each node m, the optimal feature partition j^* is selected to minimize the impurity index I_m:

$$I_m = 1 - \sum_{c=0}^{1} p_{m,c}^2 \tag{5}$$

where $p_{m,c}$ represents the proportion of samples belonging to category c in node m. Each tree outputs the category prediction:

$$\hat{y}_t(x) = h_t(x) \tag{6}$$

The overall prediction result of the random forest is obtained by integrating the outputs of all decision trees and is determined through the majority voting mechanism (Eq. 7) to finally decide the category label (Salman et al., 2024):

$$\hat{y}_{RF}(x) = \text{mode}\{\hat{y}_t(x) \mid t = 1, 2, \ldots, T\} \tag{7}$$

where mode represents the value that appears most frequently, and T represents the number of decision trees. Using Eq. (6), the prediction probabilities for each category can also be calculated:

$$P_{RF}(y = c \mid x) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}(\hat{y}_t(x) = c) \tag{8}$$

where $\mathbb{I}$ represents the indicator function. Second, this study divides the trajectory features into two modalities: (1) Dynamic mode: Speed and course. (2) Spatial mode: Longitude and latitude. Two random forests RF_{geo} and RF_{motion} are trained separately, the predicted categories $\hat{y}_{geo}, \hat{y}_{motion}$ and the corresponding probabilities P_{geo}, P_{motion} are obtained. Third, this study sets a criterion for generating pseudo labels (Eq. (9)). Only when the predicted categories of the two modalities are consistent and the predicted probabilities of both exceed the threshold

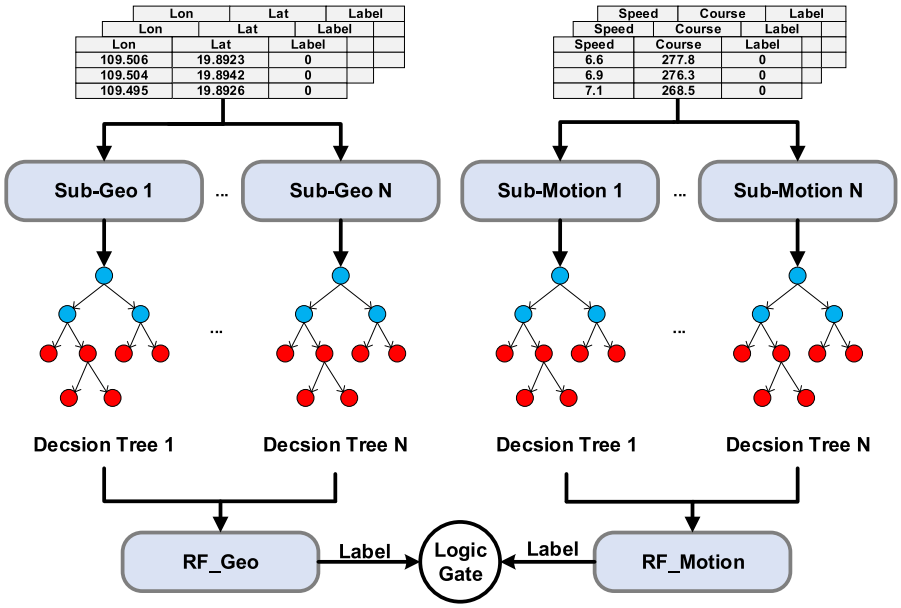


Fig. 3. Flow of two modal random forest.

τ, will the sample x_i be included in the pseudo label set D_{pseudo}:

$$x_i \in D_{\text{pseudo}} \iff \hat{y}_{\text{geo}}(x_i) = \hat{y}_{\text{motion}}(x_i) \wedge P_{\text{geo}}(x_i) > \tau \wedge P_{\text{motion}}(x_i) > \tau \tag{9}$$

Furthermore, for the high risk samples located near the decision boundary, the system automatically marks the inconsistent dual-modal output and triggers a rapid manual review channel: only when the two modal prediction results are different or the confidence levels simultaneously fall below the threshold, will the secondary annotation be initiated. In engineering practice, this strategy serves as an auxiliary module for trajectory analysis. On one hand, it mines weak behavioral features that are difficult for humans to detect in an almost automatic manner. On the other hand, through the continuous improvement through the collaboration between the model and humans, a high-quality and representative refined dataset is ultimately produced, providing a solid foundation for the subsequent model training.

4.2. Flash attention-2

The standard attention mechanism computes dot products between Query (Q), Key (K), and Value (V) followed by softmax. However, its quadratic complexity $O(L^2)$ causes prohibitive memory consumption for long AIS-based trajectories, limiting applicability in real world maritime scenarios.

To overcome this limitation and improve both memory and computational efficiency, Flash Attention-2 (Dao, 2023) is employed, an optimized attention mechanism that computes attention in a block-wise manner, thereby maximizing GPU shared memory utilization and minimizing redundant memory operations. It further integrates softmax normalization into the matrix computation pipeline, effectively reducing floating point overhead and eliminating intermediate memory bottlenecks.

Most importantly, Flash Attention-2 reduces the memory complexity from $O(L^2)$ to $O(L)$, enabling efficient processing of long range vessel trajectories with modest GPU memory usage, thereby improving scalability and accelerating both forward and backward passes. The detailed computational flow of Flash Attention-2 is illustrated in Fig. 4.

The computational process of Flash Attention-2 is as follows:

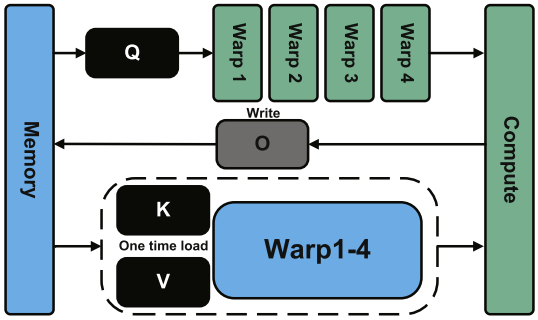


Fig. 4. Flow of flash attention-2.

Step 1: Input and Linear Mapping. Given the input sequence $X \in \mathbb{R}^{B\times L\times d_{model}}$, The Q, K and V matrices are obtained through linear mapping:

$$Q = XW_Q \tag{10}$$

$$K = XW_K \tag{11}$$

$$V = XW_V \tag{12}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{d_{model}\times(h\cdot d_k)}$, h represents the number of attention heads, d_k denotes the dimension of each head.

Step 2: Tiling. To avoid the $O(L^2)$ memory consumption caused by computing the complete $L \times L$ attention matrix, Flash Attention-2 uses warps as the processing units to concurrently process the calculations of Q, K, and V within the tiles, each warp is responsible for processing one row or a portion of rows of the tile. The threads within the warp collaborate to complete matrix multiplication, softmax, and weighted summation, using shared memory to reduce global memory access. In addition, Flash Attention-2 divides the sequence into small tiles based on their lengths, and each block is processed in the shared GPU memory:

$$K = \left[K_1, K_2, \ldots, K_{N_{\text{tile}}}\right] \quad K_j \in \mathbb{R}^{B_{\text{tile}}\times d_k} \tag{13}$$

$$V = \left[V_1, V_2, \ldots, V_{N_{\text{tile}}}\right] \quad V_j \in \mathbb{R}^{B_{\text{tile}}\times d_k} \tag{14}$$

each query tile $q_i \in \mathbb{R}^{d_k}$ streams over all key/value tiles. For a given Q-tile, the algorithm sequentially loads K and V into on-chip memory and computes partial attention results.

Step 3: Forward Pass. For the query q_i and the block K_j, V_j, Flash Attention-2 performs the calculation of local attention scores in registers and shared memory:

$$S_{ij} = Q_i \cdot K_j^{\top} \tag{15}$$

this operation is implemented by a fused warp-level kernel, and all intermediate results are not written to the global memory. Next, the algorithm performs streaming softmax normalization on the sub-blocks S_{ij} of the partition within the same fusion kernel, and immediately performs weighted summation on the corresponding value sub-blocks V_j. The normalized attention weights are:

$$A_{ij} = \mathrm{softmax}(S_{ij}) \tag{16}$$

based on this, calculate the local output:

$$O_{ij} = A_{ij} \cdot V_j \tag{17}$$

Since softmax, matrix multiplication and weighted operations are all integrated within a single kernel scheduling, the algorithm avoids constructing a complete $L \times L$ attention matrix, thereby significantly reducing the amount of memory access. Then, the calculated local output O_{ij} is cumulatively added to the output of the current query sub-block:

$$O_i \mathrel{+}= O_{ij} \tag{18}$$

as the query sub-blocks sequentially traverse all the key and value sub-blocks, the complete attention output $O \in \mathbb{R}^{B\times L\times(h\cdot d_k)}$ is finally obtained. Flash Attention-2 avoids constructing the complete $L \times L$ attention matrix by computing in blocks, which results in linear growth of memory consumption with the sequence length. At the same time, by relying on the fused kernel to integrate multiple operators in the forward and backward processes, it significantly reduces the amount of memory access and overall memory overhead, thereby achieving efficient and scalable attention computation.

4.3. Gated linear unit

In order to enhance the feature extraction capability of SeaTraNet in trajectory anomaly recognition, the GLU, which is a nonlinear transformation module with a gating mechanism, is used to enhance the model's ability to model complex patterns in AIS data by dynamically adjusting the weights of input features (Shazeer, 2020). GLU is chosen owing to its parallel processing mechanism. Compared to standard FFN, GLU introduces a gating mechanism, which provides stronger feature filtering capability and the ability to represent nonlinear perturbations.

The core idea of GLU is to divide the input features into two parallel paths: one path performs a linear transformation, while the other generates a gating signal through a nonlinear activation function. The final output is obtained by element-wise multiplication of the two paths. The computational process of GLU is defined as follows:

$$glu(x) = (xW_1 + b_1) \odot Gelu(xW_2 + b_2) \cdot W_3 \tag{19}$$

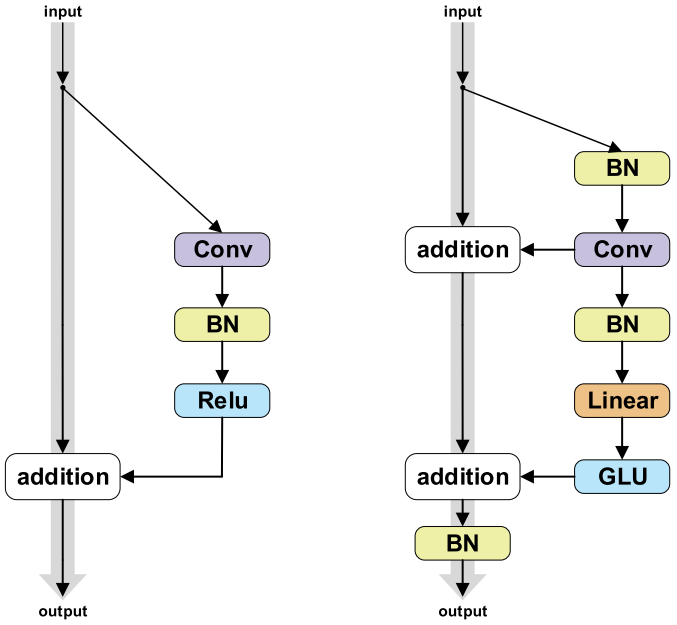
where $x \in \mathbb{R}^{B\times L\times d_{model}}$ (B is the batch size, L is the sequence length, d_{model} is the model dimension) is the model input, $W_1, W_2 \in \mathbb{R}^{d_{model}\times d_{ff}}$ and $b1, b2 \in \mathbb{R}^{d_{ff}}$ are the parameters about the linear transformation, d_{ff} is the forward dimension, $W_3 \in \mathbb{R}^{d_{ff}\times d_{model}}$ is used to map the result back to the original dimension, and $Gelu(\cdot)$ is the Gaussian Error Linear Unit function, which is used for generating gating the gated signals, and is calculated as follows:

$$Gelu(x) = x \cdot \phi(x) \tag{20}$$

where $\phi(x)$ is the cumulative distribution function of the standard normal distribution and denotes the integral of the probability density of $N(0,1)$ over $(-\infty, x]$:

$$\phi(x) = p(Z \le x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \tag{21}$$

where $Z \sim N(0,1)$ is a standard normal random variable.



(a) ResNet (b) DResGConv

Fig. 5. The architectures of ResNet and DResGConv.

4.4. Dual residual gated convolution

Inspired by ResNet (Fig. 5(a)), this paper proposes the DResGConv (Fig. 5(b)) module to enhance the model's ability to extract local spatiotemporal features from the AIS trajectories of a single trawler. This module effectively captures short-term variation patterns in the trajectory sequence through the integration of convolution, double residual connections, and GLU, and provides sharper local spatiotemporal features. The calculation steps of DResGConv are as follows:

Step 1: Pre-Convolution normalization. To ensure stable back-propagation in deep architectures, a Batch Normalization (BN) is inserted before the convolution layer, which forces the channel-wise statistics to zero-mean and unit-variance. This strategy keeps the gradients of convolutional kernels in an effective range, significantly accelerating convergence and improving generalization (Eq. 22):

$$\mathcal{H}_1 = \mathrm{Conv}(\mathrm{BN}(x)) \tag{22}$$

Step 2: Primary residual connection. The convolutional output $\mathcal{H}_1$ is added to the original input x, constituting the first residual path. This design encourages the block to learn a residual mapping $\mathcal{F}(x)$ while the identity branch ensures lossless gradient back-propagation and preserves low-level details. Consequently, the module can freely upscale or downscale channels and still stabilize the training of deep networks:

$$\mathrm{r}_1 = \mathcal{H}_1 + x \tag{23}$$

Step 3: Lightweight channel gating sub-network. After the primary residual, a BN–Linear–GLU path is introduced to capture cross-channel dependencies and perform adaptive scaling. BN stabilizes the feature distribution, Linear compresses the channel dimension into a compact representation, and GLU achieves nonlinear filtering through learnable Gelu gating, significantly enhancing the representation capability of the residual block:

$$\mathcal{H}_2 = \mathrm{Linear}\big(\mathrm{BN}(\mathcal{H}_1)\big) \tag{24}$$

$$\mathcal{H}_3 = \underbrace{(\mathcal{H}_2\mathbf{W}_1 + \mathbf{b}_1)}_{\text{value}} \odot Gelu(\underbrace{\mathcal{H}_2\mathbf{W}_2 + \mathbf{b}_2)}_{\text{gate}} \tag{25}$$

where $Gelu(\cdot)$ denotes the Gaussian Error Linear Unit function and $\odot$ denotes element-wise multiplication.

Step 4: Second residual fusion. The final output is obtained by a second residual addition, followed by BN to ensure distribution consistency:

$$\mathrm{Out} = \mathrm{BN}\left(\mathbf{r}_1 + \mathcal{H}_3\right) \tag{26}$$

In conclusion, DResGConv achieves precise modeling of fine-grained spatio-temporal features through the collaborative optimization of convolution, double residual, and gating, with low parameter overhead, and both discriminative power and efficiency.

4.5. Output splicing

To achieve the adaptive coupling of the AIS-Encoder features and the local features of DResGConv, SeaTraNet embeds a simple yet efficient gating unit at the end, which is used to dynamically balance the contribution ratios of the two paths. Specifically, the two sets of features are first concatenated along the channel dimension to obtain the initial fusion representation:

$$c = W_1 \cdot \mathrm{concat}(g, l) \tag{27}$$

where g represents the features extracted by the AIS-Encoder, l represents the local features generated by DResGConv, $\mathrm{concat}(\cdot)$ is the operation of concatenating along the channel dimension, and W_1 is the parameter of the linear transformation used to initially integrate and fuse the post-fusion representations. Subsequently, by applying the softmax operation to the fused feature c, the dynamic weights that control the contribution degree of the two paths can be obtained:

$$v_1, v_2 = \mathrm{softmax}(c) \tag{28}$$

where v_1 and v_2 correspond to the adaptive weights for the global features and local features respectively. The final output is obtained by rescaling and concatenating the two sets of features according to their respective weights:

$$\mathrm{out} = \mathrm{concat}(g \cdot v_1, l \cdot v_2) \tag{29}$$

This module can achieve fine-grained and dynamic selection of multiple source features while maintaining a lightweight structure, thereby significantly enhancing the effectiveness of feature fusion.

5. Experiments

- **Experimental environment settings:** All experiments are conducted on a workbench equipped with an NVIDIA Tesla V100 (16GB), Intel Xeon Gold 6132 CPU @ 2.60GHz. The models are implemented in Python 3.9 using PyTorch 2.5.1 with CUDA 12.1 for GPU acceleration.
- **Parameter Settings:** The encoder hidden layer dimension is 128, the number of convolutional channels is 64, the convolutional kernel size is 3, the number of multi-attention heads is 8, the number of encoder layers is 6.

5.1. Description of dataset

The dataset of this study comes from the Watchman platform, which provides AIS data of ships, covering the sea area around Hainan Island and part of North Bay. The dataset records the dynamic behavior of ships in different time periods, containing key information such as time, position, speed and heading, which lays the foundation for subsequent trajectory prediction and abnormal behavior analysis. The specific data coverage is shown in Fig. 6.

The raw data format adopted by the platform is shown in Table 1. where time denotes the timestamp of data acquisition, lon and lat denote the longitude and latitude of the ship respectively, which are used to describe its geographic location, speed represents the sailing speed of the ship at that moment in time, course reflects the ship's heading, and TargetId serves as a unique identifier of the target trajectory. These key features together constitute the dynamic trajectory information of the ship.

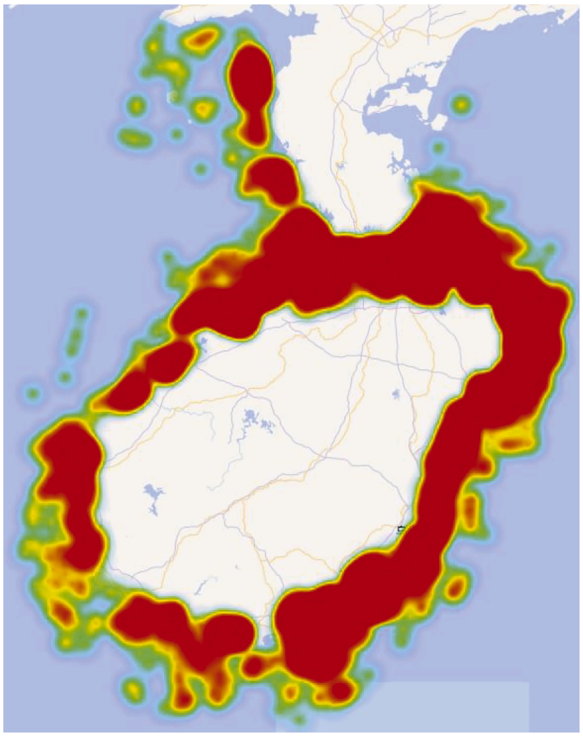


Fig. 6. The system interface of the Watchman platform.

Table 1
The Style of the Raw Data.

Time	Lon	Lat	Speed	Course	TargetId
2025-03-17 05:05:45	110.2	20.5	3.96	347.1	ID_001
2025-03-17 05:06:45	110.3	20.6	4.30	344.6	ID_001
2025-03-17 05:07:23	110.5	20.8	4.24	343.2	ID_001

Table 2
The style of the Preprocessed datat.

Time	Lon	Lat	Speed	Course
2025-03-17 05:05:45	110.2	20.5	3.96	347.1
2025-03-17 05:06:45	110.3	20.6	4.30	344.6
2025-03-17 05:07:45	110.5	20.8	4.24	343.2

Considering the problem of sequence length inconsistency, this study uses a fixed timestamp and removes the TargetId to eliminate the effect of unequal time series on the model. The format of the processed data is shown in Table 2:

5.2. Processed of dataset

This study collects 21,224 single trawler samples from the platform and constructs the initial training set based on 19.55 % of the manually labeled data. Before formal training, the raw data are preprocessed to eliminate the stationary segments (the trajectory segments where the speed of the vessel is 0). Subsequently, features are extracted from the labeled data for initial model training, the same feature extraction process is applied to the unlabeled data. To prevent data contamination, the unlabeled data are also cleaned and validated during processing, only valid trajectories are retained.

In the self-training framework, the number of trees is set to 100, the random seed is set to 42, and the confidence threshold is set to 0.9. The initial model is trained with labeled data, then iteratively generates pseudo-labels for the unlabeled data, and the high-confidence sam-

ples are included in the extended training set. The self-training is performed for a maximum of 50 rounds, and is terminated early if no new high-confidence samples are generated in a round. The number of new pseudo-labeled samples is recorded for each iteration round. Eventually, in subsequent experiments, the SeaTraNet will be trained on the extended training set, which combines the initial labeled data with all high-confidence pseudo-labeled samples. For model evaluation, the labeled dataset is split into training, validation, and test sets with a ratio of 0.7 : 0.15 : 0.15. The training set is used for initial model training and iterative self-training with pseudo-labels, the validation set is used to monitor model performance and tune hyper-parameters, and the test set is held out for final evaluation. The splitting is performed in a stratified manner to preserve the proportion of normal and abnormal behavior samples across all sets.

5.3. Evaluation criteria

In order to better reflect the experimental results, this study chooses accuracy, precision, recall, and F1-Score as the evaluation criteria, which are very classical indicators, and the formulas of the four indicators are placed below for easy understanding.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{30}$$

$$Precision = \frac{TP}{TP + FP} \tag{31}$$

$$Recall = \frac{TP}{TP + FN} \tag{32}$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{33}$$

where (1) TP (True Positive): the number of samples that are predicted to be positive and are truly positive. (2) TN (True Negative): the number of samples that are predicted to be negative and are really negative. (3) FP (False Positive): the number of samples that are predicted to be positive but are really negative. (4) FN (False Negative): the number of samples that are predicted to be negative but are really positive.

5.4. Explainability criterion

It is well known that deep learning models are difficult to understand due to their complex network structure, and work on the explainability of deep learning can lead to a better understanding of the internal logic (Samek et al., 2017). In this section, a widely adopted technique is first introduced to explain the recognition process of SeaTraNet. Layer Relevance Propagation (LRP) explains the classifier's decisions separately (Bach et al., 2015), which uses a rule of local redistribution to redistribute the predictions until a score is assigned to each input variable, The key property of the redistribution process refers to relevance preservation, which can be described as:

$$\sum_i R_i = \sum_j R_j = \sum_k R_k = \cdots = f(x) \tag{34}$$

The reallocation process preserves the relevance scores and no relevance is artificially added or removed during this period, and the relevance score R_i for each input feature determines the contribution of this feature to the results in a way that decomposes the function $f(x)$, unlike the sensitivity analysis absolute results.

5.5. Ablation experiments

In order to investigate the effect of DResGConv module, GLU module and Flash Attention-2 module on SeaTraNet, several sets of structural ablation experiments are designed in this section, the results are shown in Table 3.

As shown in Table 3, a series of ablation experiments are conducted to validate the effectiveness of each enhancement module. The base model (Base) serves as the performance benchmark when no enhancement module is integrated, achieving an Accuracy of 0.8073 and an F1-Score of 0.7738. Each module is introduced into the base model in sequence to evaluate its independent contribution:

- With the DResGConv module, the F1-Score improves markedly to 0.8256, highlighting its effectiveness in extracting local structural features.
- With the GLU module, the F1-Score reaches 0.7996, showing benefits in feature selection despite modest overall gains.
- With the Flash Attention-2 module, the F1-Score increases to 0.8082. Notably, this configuration also reduces memory consumption (3698 MiB → 2718 MiB) and achieves the fastest inference time (0.0050 s/sample), demonstrating both accuracy and efficiency advantages.

The results indicate that each module individually contributes positively to model performance. Next, different module combinations are further examined:

- With the GLU and Flash Attention-2 modules: The F1-Score improves to 0.8092, while maintaining low memory usage (2968 MiB) and fast inference (0.0056 s/sample), suggesting complementary benefits between gated feature selection and efficient sequence modeling.
- With the DResGConv and GLU modules: This combination achieves an F1-Score of 0.7939, with Recall decreasing slightly to 0.7118 but Precision reaching 0.8974, the highest across all settings. This indicates enhanced discrimination confidence, albeit at a slower inference speed (0.0088 s/sample).
- With the DResGConv and Flash Attention-2 modules: The F1-Score rises to 0.8215, balancing local structural modeling and efficient sequence representation. Moreover, the inference time is reduced to 0.0054 s/sample, further validating the efficiency of Flash Attention-2.

Through analysis, it is found that integrating all three modules (DResGConv, GLU, and Flash Attention-2) can achieve the best overall performance, with an Accuracy of 0.8559 and an F1-Score of 0.8394. At the same time, the memory usage is kept moderate (3032 MiB), and inference remains efficient (0.0062 s/sample). This demonstrates that SeaTraNet achieves superior accuracy while ensuring practical deployability under resource-constrained conditions. Furthermore, to systematically evaluate the contributions of DResGConv and AIS-Encoder to the performance of SeaTraNet, this study designed a set of ablation experiments: sequentially removing or replacing the single/double path DResGConv and the single/double path AIS-Encoder, and comparing them with SeaTraNet. The results are shown in the Table 4.

As can be seen from the table, SeaTraNet significantly outperforms the single structures of pure dual-path AIS-Encoder or pure dual-path DResGConv in terms of accuracy, recall and efficiency: its Accuracy, Precision, Recall and F1-Score all reach the highest values simultaneously, while the memory usage (3032 MiB) and inference latency (0.0062) are at a relatively reasonable level. It can be seen that the fusion strategy effectively integrates the complementary characteristics of the two paths, AIS-Encoder ensures positive class coverage, and DResGConv provides high discrimination accuracy, thereby improving the recall rate by 4.7 % and the F1-Score by 3.8 % while maintaining high accuracy without introducing additional computational burden. In contrast, the single structure: dual-path AIS-Encoder has a high recall rate but slightly lower accuracy and efficiency, dual-path DResGConv has outstanding accuracy but insufficient recall, resulting in the lowest F1-Score. In summary, SeaTraNet's dual-path fusion achieves the optimal trade-off between classification performance and resource consumption, verifying the necessity and effectiveness of collaborative design. Overall, through the combination of well-designed modules, SeaTraNet has been comprehensively enhanced in feature modeling, sequence representation and classification discrimination, providing a more efficient and robust technical path for ship trajectory anomaly recognition task.

Table 3
The results of ablation experiments.

Base	DResGConv	GLU	Flash Attention-2	Accuracy	Precision	Recall	F1-Score	Memory	Time(s/sample)
✓				0.8073	0.8112	0.7394	0.7738	3698 MiB	0.0091
✓	✓			0.8438	0.8520	0.8008	0.8256	3744 MiB	0.0087
✓		✓		0.8299	0.8784	0.7331	0.7996	3948 MiB	0.0089
✓			✓	0.8385	0.8950	0.7368	0.8082	2718 MiB	0.0050
✓		✓	✓	0.8365	0.8616	0.7630	0.8092	2968 MiB	0.0056
✓	✓		✓	0.8448	0.8585	0.7871	0.8215	2764 MiB	0.0054
✓	✓	✓		0.8305	0.8974	0.7118	0.7939	3994 MiB	0.0088
✓	✓	✓	✓	0.8559	0.8645	0.8158	0.8394	3032 MiB	0.0062

Table 4
Benchmarking single path models and their parallel fusion architecture.

Metrics	DResGConv		AIS-Encoder		SeaTraNet
	Single	Dual	Single	Dual	
Accuracy	0.7785	0.8350	0.8073	0.8180	**0.8559**
Precision	0.8230	0.8510	0.8112	0.8270	**0.8645**
Recall	0.7120	0.7250	0.7394	0.8020	**0.8158**
F1-Score	0.7430	0.7800	0.7738	0.8120	**0.8394**
Memory	982 MiB	1580 MiB	3698 MiB	4200 MiB	**3032 MiB**
Time (s/sample)	0.0045	0.0048	0.0091	0.0096	**0.0062**

Table 5
Performance of different models on the TrawlTag dataset.

Model	Accuracy	Precision	Recall	F1-Score	Memory	Time(s/sample)
TrAISformer	0.8090	0.7868	0.8045	0.7955	4606 MiB	0.0024
CNN-LSTM	0.8194	0.7852	0.8383	0.8109	2879 MiB	0.0021
Dlinear	0.7101	0.6800	0.7030	0.6913	2004 MiB	0.0014
GhostNet	0.8142	0.8279	0.7946	0.8112	2398 MiB	0.0059
Resnet	0.7916	0.7943	0.7406	0.7665	2649 MiB	0.0011
SegRNN	0.7726	0.7547	0.7519	0.7533	2418 MiB	0.0010
TimesNet	0.8403	0.8246	0.8308	0.8277	3397 MiB	0.0069
FiLM	0.7743	0.7595	0.7481	0.7538	2623 MiB	0.0013
Our	**0.8559**	**0.8645**	**0.8158**	**0.8394**	**3032 MiB**	**0.0062**

5.6. Validation on a TrawlTag dataset

To further validate the stability and robustness of SeaTraNet in semi-supervised scenarios, this subsection conducts extended experiments on the TrawlTag dataset and selects eight mainstream baseline models for comparison, including TrAISformer (Li et al., 2024b), CNN-LSTM, Dlinear (Zeng et al., 2023), GhostNet (Han et al., 2020), ResNet (He et al., 2016), SegRNN (Lin et al., 2023), TimesNet (Wu et al., 2022) and FiLM (Perez et al., 2018).This dataset is constructed by automated label generation method, which has certain label noise and behavioral ambiguity, and is closer to real maritime application scenarios.

As shown in Table 5, SeaTraNet performs optimally in all key performance indicators, with an accuracy of 0.8559, precision of 0.8645, recall of 0.8158, and F1-Score of 0.8394, which is significantly better than the other comparison methods. Compared with the second best model, TimesNet (F1-Score 0.8277), SeaTraNet improves about 1.41 %, and has a more obvious precision advantage compared with other models such as CNN-LSTM. This result verifies the stronger discriminative ability and robustness of SeaTraNet in noisy data environments. Especially in terms of precision, SeaTraNet performs especially well, reaching 0.8645, surpassing GhostNet (0.8279) and TimesNet (0.8246), indicating that the model has a significant advantage in reducing false alarms. Although the inference time is 0.0062 sec/sample, which is slightly higher than lightweight models such as CNN-LSTM and FiLM, its memory usage (3032 MiB) is still kept at a moderate level, maintaining a good efficiency-performance balance. Overall, SeaTraNet maintains its leading position on the TrawlTag dataset, proving that it is able to effectively cope with the challenges of label uncertainty and behavioral diversity, and has the potential to be generalized to complex, low-quality data scenarios. The experimental results further support the feasibility of deploying SeaTraNet in real-world applications.

In addition, in order to evaluate more intuitively the performance of SeaTraNet in the task of recognizing the abnormal behaviors of single-trawler fishing vessels, Fig. 7 in this subsection shows the Receiver Operating Characteristic (ROC) curves and their corresponding Area Under Curve (AUC) values of each model on the test set, which are used to measure the classification ability of the model in the recognition of abnormal trajectories.

As can be seen in Fig. 7, SeaTraNet (Fig. 7i) achieved the highest score of 0.9304 on the AUC metric, significantly outperforming all baseline models. Specifically, CNN-LSTM (Fig. 7a) is 0.9197, ResNet (Fig. 7b) is 0.9002, GhostNet (Fig. 7c) is 0.8583, Dlinear (Fig. 7d) is 0.8300, TrAISformer (Fig. 7e) is 0.9013, TimesNet (Fig. 7f) is 0.9075, SegRNN (Fig. 7g) is 0.8878, and FiLM (Fig. 7h) is 0.8596. Compared with the second best CNN-LSTM, the AUC of SeaTraNet is improved by 1.16 %, and compared with the worst-performing Dlinear, the improvement is up to 12.1 %, which indicates that this model has a stronger classification ability and significant advantages in the abnormal behavior recognition task. This indicates that the model has stronger classification ability and significant advantages in the task of abnormal trajectory recognition. Further observation of the ROC curve pattern shows that SeaTraNet still maintains a high true positive rate (TPR) in the region of low false positive rate (FPR), which indicates that it is able to control the false alarm rate while effectively recognizing illegal operation behaviors. This performance is especially critical for high-risk application scenarios such as maritime law enforcement. In contrast, although CNN-LSTM and ResNet perform well in the low FPR region, their overall curves are still slightly inferior to SeaTraNet, while GhostNet, Dlinear and FiLM show obvious performance degradation in this region, making it difficult to meet the high requirements for the accuracy of recognizing abnormal behaviors.

5.7. Comparative experiment

In order to fully evaluate the performance of SeaTraNet, this section conducts comparative experiments with several representative benchmark models on the HN_STV (Gu et al., 2024) dataset, and the results are shown in Table 6. It is worth pointing out that although MFGTN (Gu et al., 2024) achieves better performance on this dataset (accuracy 84.92 %, precision 84.82 %, recall 85.53 %, f1-score 84.83 %), which further validates the evaluation value of this dataset, it is not included in this comparison of benchmark models due to the significant difference between its experimental environment and the settings. Overall, SeaTraNet achieves optimal performance in several key evaluation indexes, showing excellent comprehensive performance and robustness.

Specifically, SeaTraNet achieves an accuracy of 0.8571, which exceeds TrAISformer (0.8472) and CNN-LSTM (0.8469), and is the best performance among all the compared methods. Meanwhile, its F1 score is 0.8554, which is 1.29 percentage points higher than TrAISformer, fully reflecting SeaTraNet's ability to balance between classification precision and recall. From the perspective of Precision, SeaTraNet also achieves the highest value of 0.8802, which is significantly better than CNN-LSTM (0.8613) and DLinear (0.8565), indicating that SeaTraNet has a stronger ability to recognize anomalous trajectories

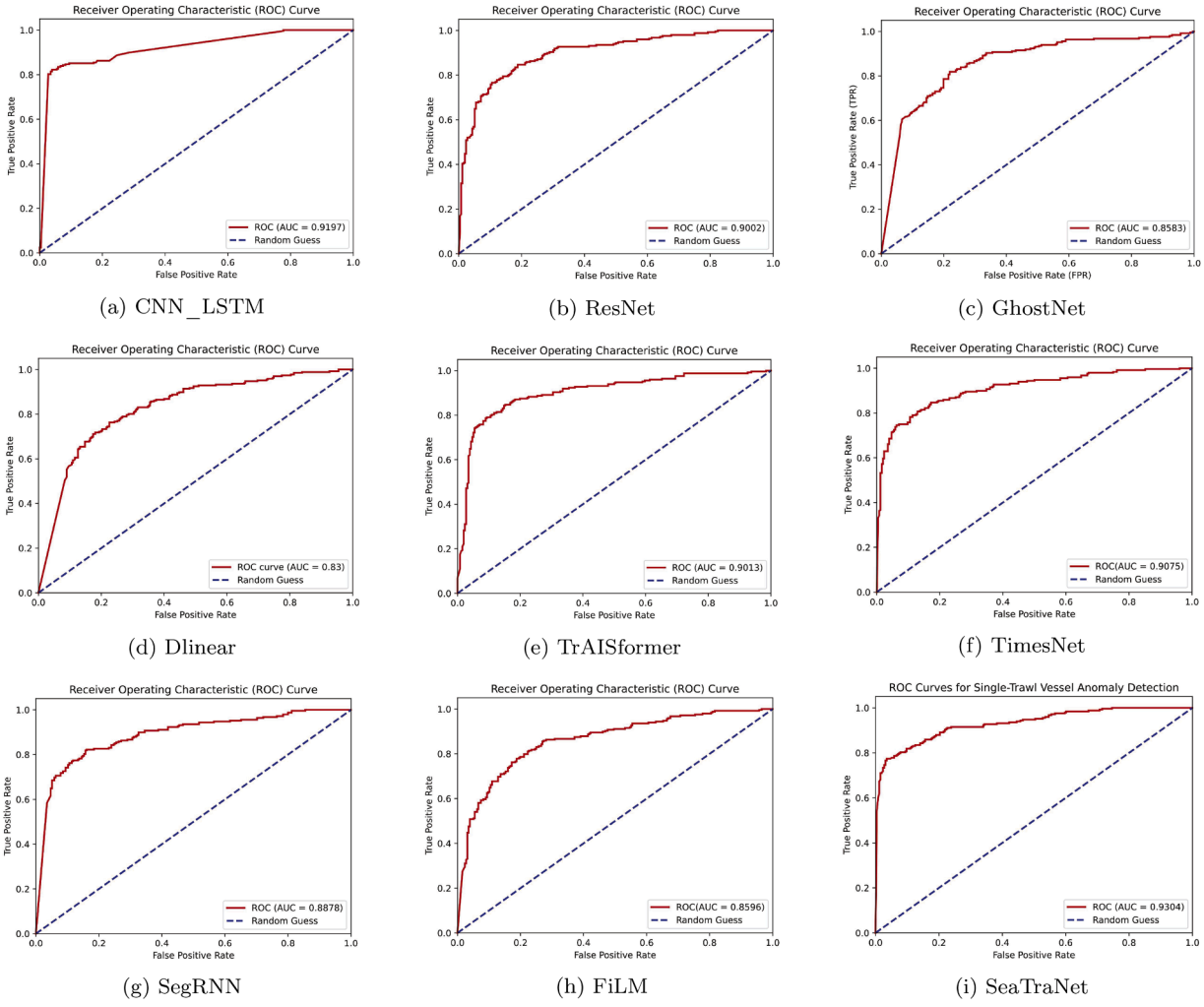


Fig. 7. ROC curves for different models.

Table 6
Performance of different models on the HN_STV dataset.

Model	Accuracy	Precision	Recall	F1-Score	Memory	Time(s/sample)
TrAISformer	0.8472	0.8548	0.8306	0.8425	15,344 MiB	0.0024
CNN-LSTM	0.8469	0.8613	0.8024	0.8306	9590 MiB	0.0024
Dlinear	0.8333	0.8565	0.7944	0.8243	6590 MiB	0.0018
GhostNet	0.7837	0.7421	0.8588	0.7962	4987 MiB	0.0071
ResNet	0.8035	0.7508	0.8991	0.8183	8823 MiB	0.0050
SegRNN	0.8290	0.8320	0.8226	0.8185	8053 MiB	0.0016
TimesNet	0.8333	0.8417	0.8145	0.8279	10,728 MiB	0.0016
FiLM	0.7917	0.8071	0.7661	0.7835	6514 MiB	0.0035
Our	**0.8571**	**0.8802**	**0.8320**	**0.8554**	**10094 MiB**	**0.0051**

and a lower false alarm rate. Although it is slightly lower than ResNet (0.8991) and GhostNet (0.8588) in terms of Recall, it achieves the overall performance lead by virtue of a higher precision rate while maintaining the recall capability. In terms of resource consumption, SeaTraNet consumes 10,094 MiB of memory, which is significantly lower than TrAISformer (15344 MiB) and TimesNet (10728 MiB), and the inference time is 0.0051 s/sample, which realizes a good compromise between model performance and inference efficiency. Compared with GhostNet, although the running time is slightly higher, its classification performance is significantly improved with obvious advantages. SeaTraNet outperforms the existing mainstream methods in key indexes such as accuracy, precision and F1-Score, and maintains the competitiveness in terms of resource consumption and running efficiency, which verifies its validity and practicability in modeling the behavior of complex ship trajectories. The results further indicate that SeaTraNet has strong promotion potential and application value in AIS anomaly recognition tasks.

5.8. Structural collaborative exploration

In order to conduct a detailed analysis of the feature contributions of DResGConv and AIS-Encoder during the model process, six typical trajectories are selected, and their attention weights in the test are calculated, as shown in Figs. 8–10. To facilitate comparison, the attention weights for the first 300 steps are fixed, and a detailed analysis is conducted on their contribution to the features. Figs. 8a and Fig. 8b, Fig. 8c and Fig. 8d, Fig. 10a and Fig. 10b, Fig. 9(c) and Fig. 9(d) show the performance of DResGConv and AIS-Encoder under different strides. The DResGConv model ignores the features on the left and right of the first 30 steps, the first 35 steps, the first 25 steps, and the first 55 steps, while the AIS-Encoder can better capture these features. In Fig. 9(a) and Fig. 9(b), as well as Fig. 10c and Fig. 10d , DResGConv respectively ignores the features from the last 50 steps and the last 55 steps or so. However, AIS-Encoder is capable of capturing these details with great precision. Particularly, in steps 260 to 300 of Fig. 8a and Fig. 8b, steps 250 to 300 of Fig. 8c and Fig. 8d, the first 30 steps of Fig. 9(a) and Fig. 9(b), steps 250 to 300 of Fig. 9(c) and Fig. 9(d), steps 50 to 110 of Fig. 10a and Fig. 10b, the first 30 steps of Fig. 10c and Fig. 10d, and steps 60 to 100 of Fig. 10c and Fig. 10d, the features captured by DRes-

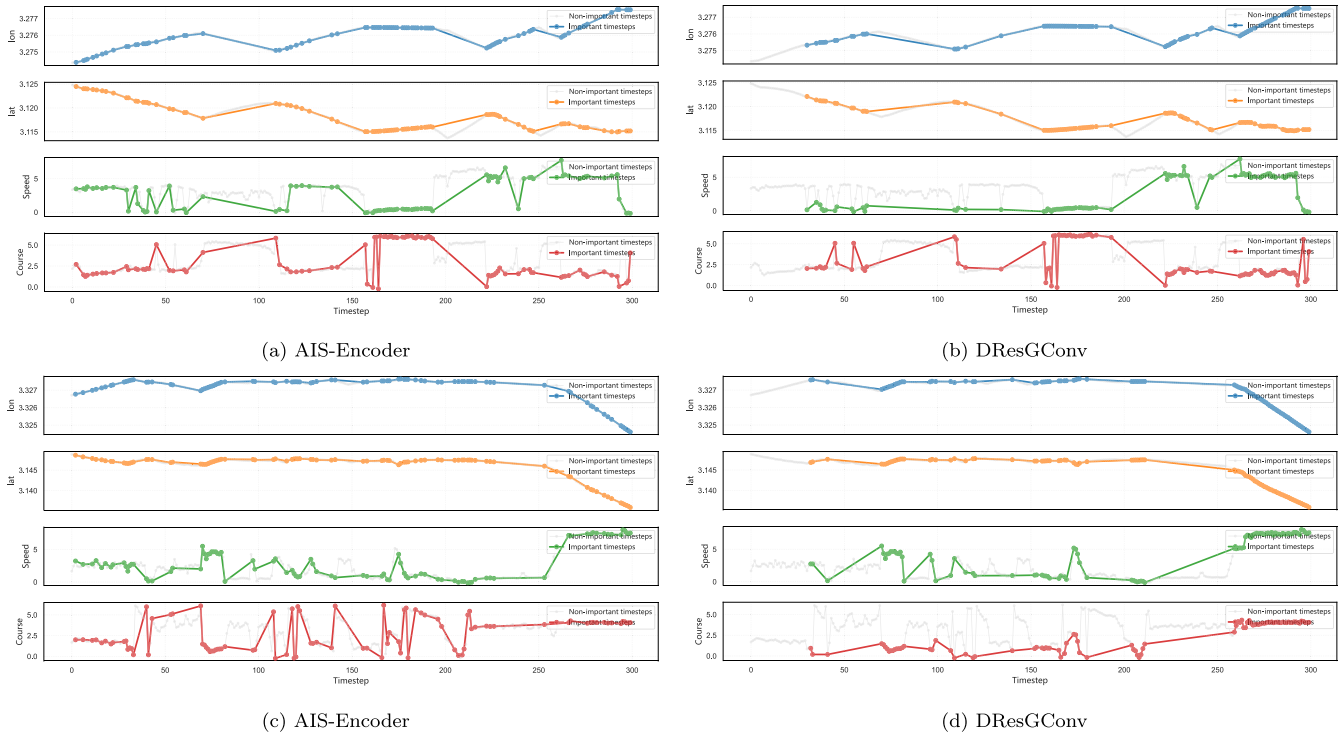


Fig. 8. Path attention: Trajectory 1-2.

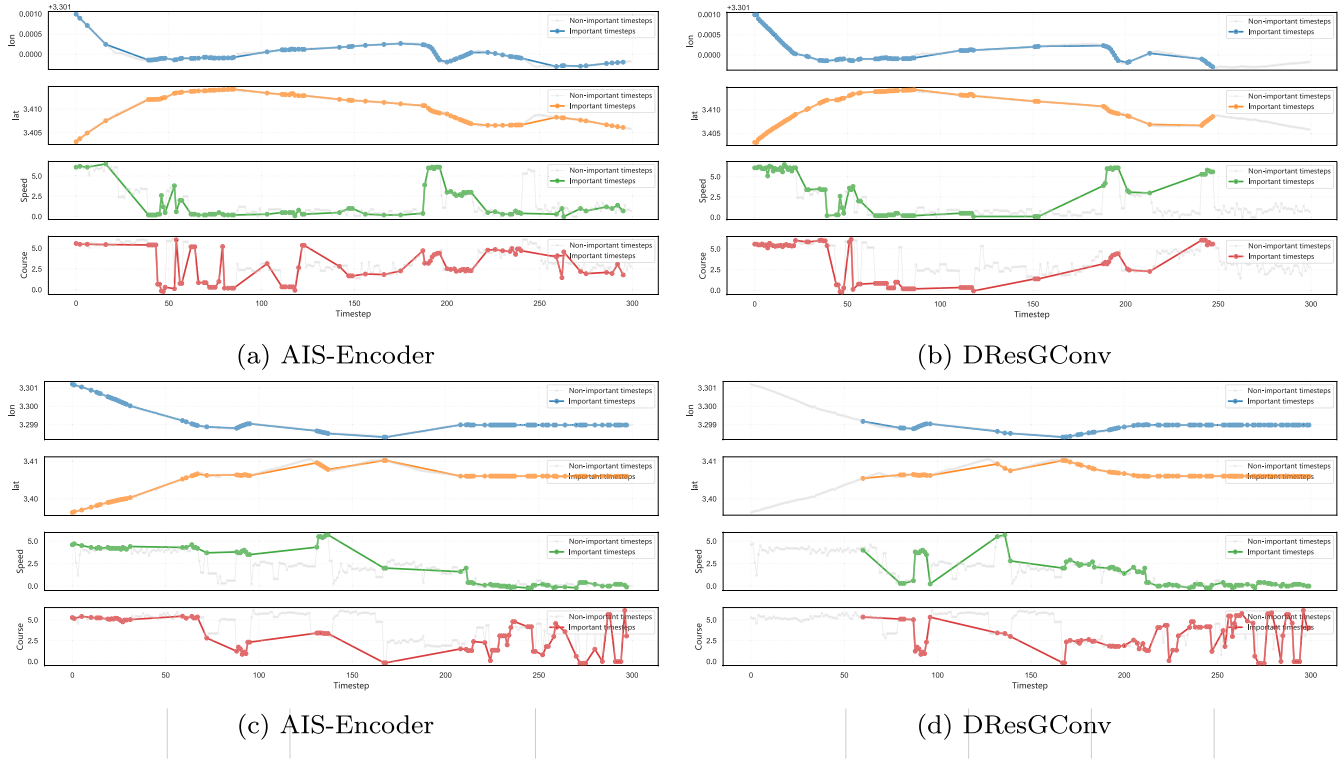


Fig. 9. Path attention: Trajectory 3-4.

GConv are significantly more detailed, while AIS-Encoder demonstrates its advantage in capturing global information.

By conducting an attention visualization analysis of the two structures, namely DResGConv and AIS-Encoder, the following conclusions can be drawn: The DResGConv structure demonstrates sensitivity to local features, with its attention distribution concentrated on specific time steps and local areas. This indicates that it can effectively capture subtle changes in ship behavior, such as sudden speed changes or rapid adjustments in course. This local focus ability enables DResGConv to perform well in recognizing the immediate operations and short-term behavior patterns of ships. In contrast, the attention distribution of the AIS-Encoder structure shows a focus on global features. Its attention map

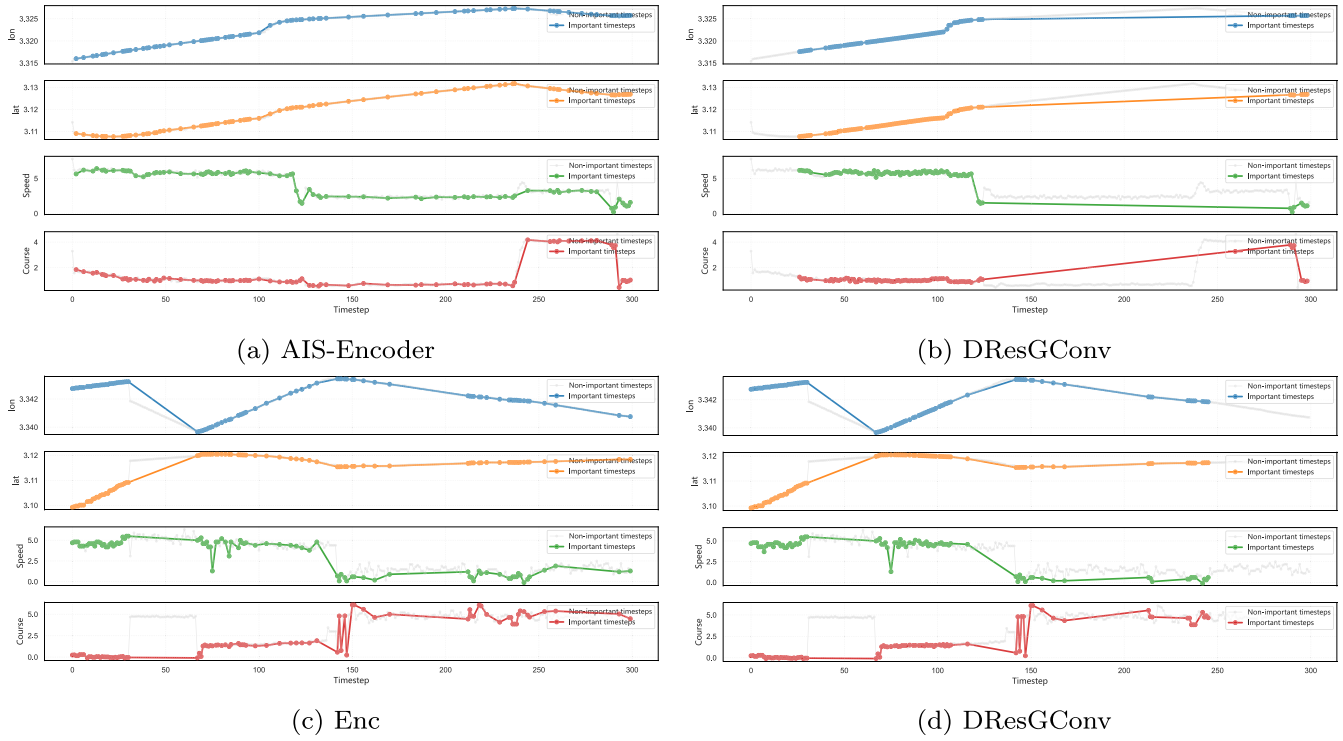


Fig. 10. Path attention: Trajectory 5-6.

Table 7
Performance of different models on open source datasets.

Model	Accuracy	Precision	Recall	F1-Score	Memory	Time(s/sample)
TrAISformer	0.8429	0.8421	0.8430	0.8421	4290 MiB	0.0073
CNN-LSTM	0.8645	0.8662	0.8646	0.8628	3890 MiB	0.0032
Dlinear	0.7193	0.7178	0.7197	0.7185	2560 MiB	0.0082
GhostNet	0.8082	0.8061	0.8086	0.8065	3170 MiB	0.0051
ResNet	0.8369	0.8374	0.8370	0.8372	4610 MiB	0.0049
SegRNN	0.7646	0.7622	0.7650	0.7623	3380 MiB	0.0036
TimesNet	0.8061	0.8070	0.8061	0.8026	5220 MiB	0.0020
FiLM	0.5862	0.5855	0.5866	0.5860	2860 MiB	0.0028
Our	**0.8712**	**0.8706**	**0.8715**	**0.8715**	**4920 MiB**	**0.0052**

covers a longer time span and a wider range of features. AIS-Encoder is more inclined to recognize and track the long-term patterns and overall trends of ship behavior, such as the stability of course and the continuous changes in latitude and longitude. This global perspective gives AIS-Encoder an advantage in understanding the overall movement trajectory and operation mode of ships. In summary, the differences in feature attention between DResGConv and AIS-Encoder indicate that they play different roles in the task of recognizing abnormal ship behavior. DResGConv focuses on capturing local details, while AIS-Encoder is dedicated to recognizing global patterns. The complementary nature of these two structures helps to enhance the model's ability to understand and recognize complex ship behavior.

5.9. Evaluation of generalization capabilities

In order to further validate the generalization ability of the proposed model SeaTraNet in different navigation tasks and scenarios, this study designs an extended comparative experiment by introducing a publicly available dataset (https://www.heywhale.com/mw/dataset/623b00c9ae5cf10017b18cc6/content) of multi-type fishing vessel operations as a new test environment. This dataset covers three typical fishing operations, namely trawling, gillnetting and purse seining, with significant operational heterogeneity and trajectory diversity, and can be used as an important benchmark for evaluating the generalization of the model in cross-type tasks.

In order to maintain the comparability with the previous experiments, the four basic features of longitude, latitude, speed and heading are still selected as model inputs in this study. In addition, considering that there are a large number of time periods in this dataset in which the ships are docked or stationary (with speed 0), in order to reduce the interference of invalid information, a masking mechanism is introduced in the preprocessing stage to shield the stationary segments, focusing on the modeling and recognition of the effective motion behaviors.

In this scenario, all model structures and training settings are kept consistent with the previous section, and the comparison results are shown in Table 7. Overall, SeaTraNet achieves the optimal performance in terms of Accuracy, Precision, Recall and F1-Score. Among them, the Accuracy reaches 0.8712 and the F1-Score is 0.8708, which is significantly better than other methods. Compared with the second best performance of CNN-LSTM (F1 = 0.8628), SeaTraNet improves 0.80 %, compared with TrAISformer (F1 = 0.8421), the improvement reaches 2.87 %.

In addition to performance, two key indicators for practical deployment-inference efficiency and memory usage are also highlighted. In terms of inference efficiency, the average inference time of SeaTraNet is 0.0052 s/sample. Although this is slightly higher than that of CNN-LSTM (0.0032 s/sample) and TimesNet (0.0020 s/sample), it remains within an acceptable range for real-world applications. Regarding memory consumption, SeaTraNet occupies 4920 MB of GPU memory during operation, which is lower than complex temporal models such as TimesNet (5220 MB) and ResNet (4610 MB), and only moderately higher than lightweight models like GhostNet (3170 MB) and FiLM (2860 MB), showing a good balance between performance and efficiency, and feasible for practical deployment.

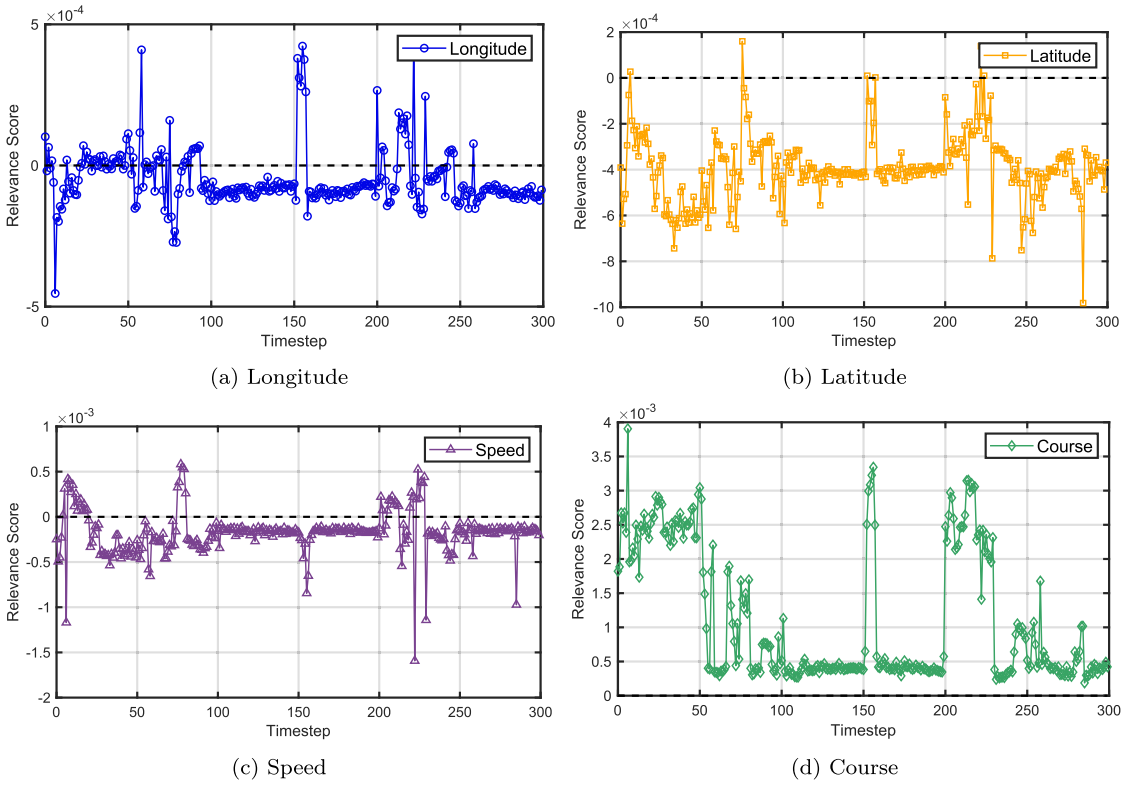


Fig. 11. LRP scores of the four features.

The extended experimental results fully show that SeaTraNet can still maintain stable performance in the face of diverse operational behaviors and complex trajectory characteristics, has good cross-scene generalization ability and practical value, and provides a solid foundation for the construction of generalized trajectory anomaly recognition system.

5.10. The explainability of SeaTraNet

To reveal the inherent decision making logic of SeaTraNet in the task of recognizing abnormal behaviors of single trawler vessels, this section employs the LRP method mentioned earlier to conduct reverse correlation decomposition on the trained SeaTraNet. Fig. 11 presents the time series curves of the relevance intensity of the four input variables to the model output within 0–300 time steps. Positive values indicate that the variable promotes an increase in the model output, while negative values indicate a suppression. By combining the relevance results with the geometric features of the trajectory, a detailed assessment of a single trajectory can be conducted.

The time series curve in Fig. 11 and the heatmap in Fig. 12 jointly show that the heading consistently provides a significant positive contribution throughout the entire trajectory.While the longitude, latitude, and speed have a weak positive correlation for a few time steps, they are all negatively correlated in the remaining periods.

The trajectory graph in Fig. 13 can provide spatial evidence for the above relevance results. Within the 0–50 time steps, the ship sailed at a nearly constant speed in a fixed direction at a uniform speed. From 50 to 110 steps, the course fluctuated sharply, and the speed also fluctuated synchronously. It is speculated that the ship has reached the first operation area and started to drop the nets. From 110 to 150 steps, the latitude, longitude, speed and course showed no significant changes, and the signal information is low, which can be regarded as a normal operation period. From 150 to 155 steps, the course changed again, and the latitude and longitude showed a turning point, indicating that the ship is preparing to turn to the next operation block. From 155 to 200 steps, the course changed again suddenly, and the latitude and longitude showed a turning point, suggesting that the ship is turning to the next operation area. From 200 to 230 steps, the latitude and longitude decreased slowly, and the speed remained low, corresponding to the process of net collection or ship relocation. During this period, the amplitude of each variable's relevance is small, consistent with the reduction in control intensity. From 230 to 300 steps, the course fluctuated violently again, the speed gradually decreased, and combined with the changes in latitude and longitude, it can be determined that the ship entered a new operation cycle.

Based on the correlation analysis, the system disassembled the single trawling operation chain. This trajectory completely replicated the typical single trawling net mode of high speed offshore operation, fixed-point cyclic operation, short-distance relocation, confirming the explainable mechanism of precise recognition achieved by SeaTraNet through its heading sensitivity and speed-position coupling characteristics.

6. Discussion

Through experimental validation, the SeaTraNet shows better performance in recognizing abnormal behaviors of single-trawler, and is able to recognize most of the abnormal behaviors effectively. However, in practical application scenarios, due to the complexity and diversity of the maritime environment, the frequent occurrence of some special situations can lead to misjudgments. These misjudgments may not only hide the real illegal behaviors, but also interfere with the subsequent regulation and decision-making. Therefore, in order to deeply understand the limitations of the model and enhance its robustness, it is necessary to specifically analyze the misjudgments results. To this end, some false recognition results are collected, and the trajectories corresponding to these results are visualized, where the red trajectories indicate that the model incorrectly recognizes a non-single trawler as a single trawler, and the blue trajectories indicate that the model incorrectly recognizes a single trawler as a non-single trawler. The results of plotting these trajectories are shown in Fig. 14.

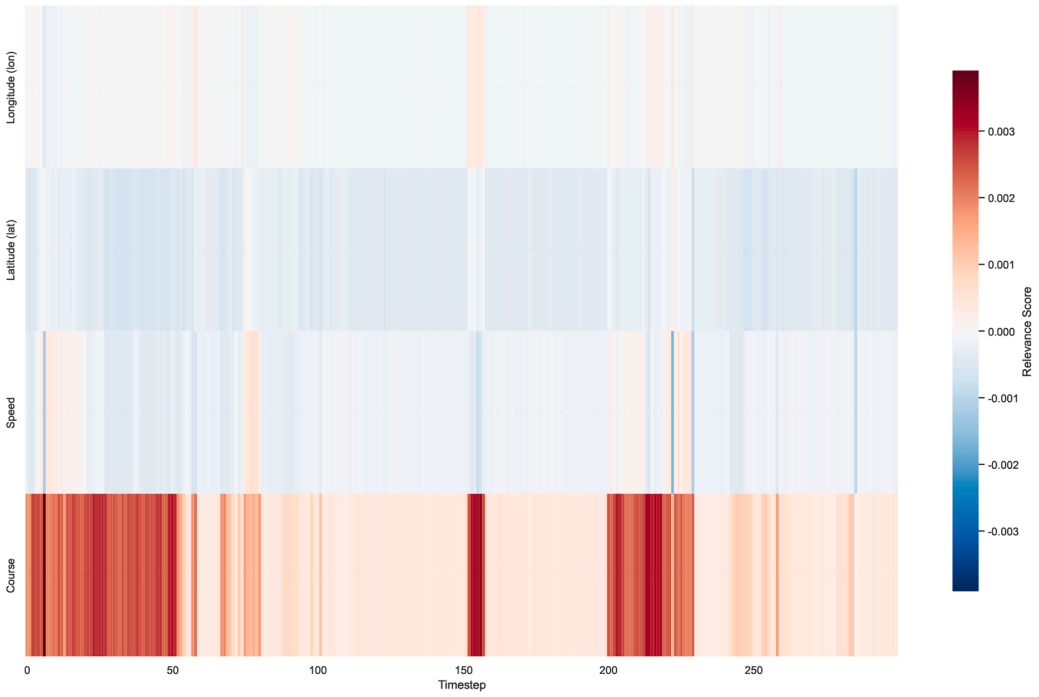
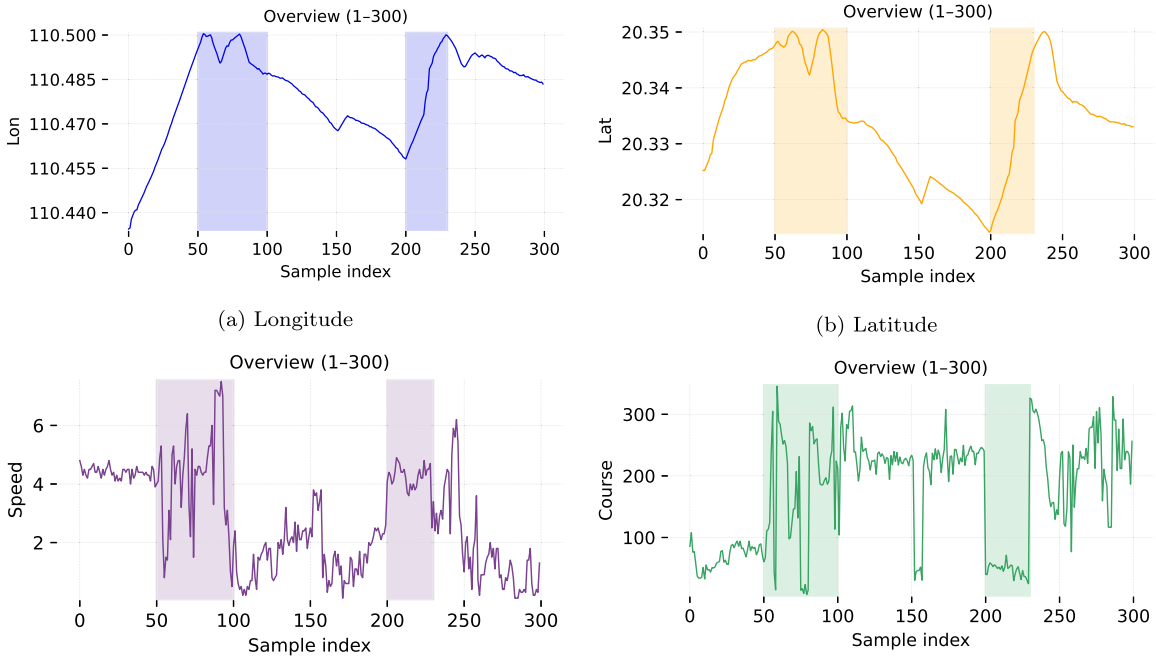


Fig. 12. The heat map of LRP.



(a) Longitude (b) Latitude

(c) Speed (d) Course

Fig. 13. The values corresponding to the four features.

From the figure, the scenarios that cause model recognition errors can be mainly divided into the following two types.

First, most misclassifications are closely related to insufficient length of ship data sequences. For instance, the trajectories in the second column of the first row and the fourth column of the third row show that when there are few data points, the behavioral patterns of the vessels are difficult to be captured completely. In this case, even a non-single trawler may exhibit special behavioral characteristics similar to those of a single trawler for a short period of time, such as sudden deceleration and frequent steering, which causes the model to mistake it for a single trawler. This phenomenon indicates that data sparsity is one of the key factors affecting model performance, especially when the time series is short or the sampling frequency is low, the model may not be able to accurately distinguish the essential differences in behavioral patterns.

Second, another type of misclassification stems from the fact that ship trajectories are affected by complex maritime factors, leading to abnormal changes in heading. For instance, the trajectories in the first column of the second row and the first column of the fourth row

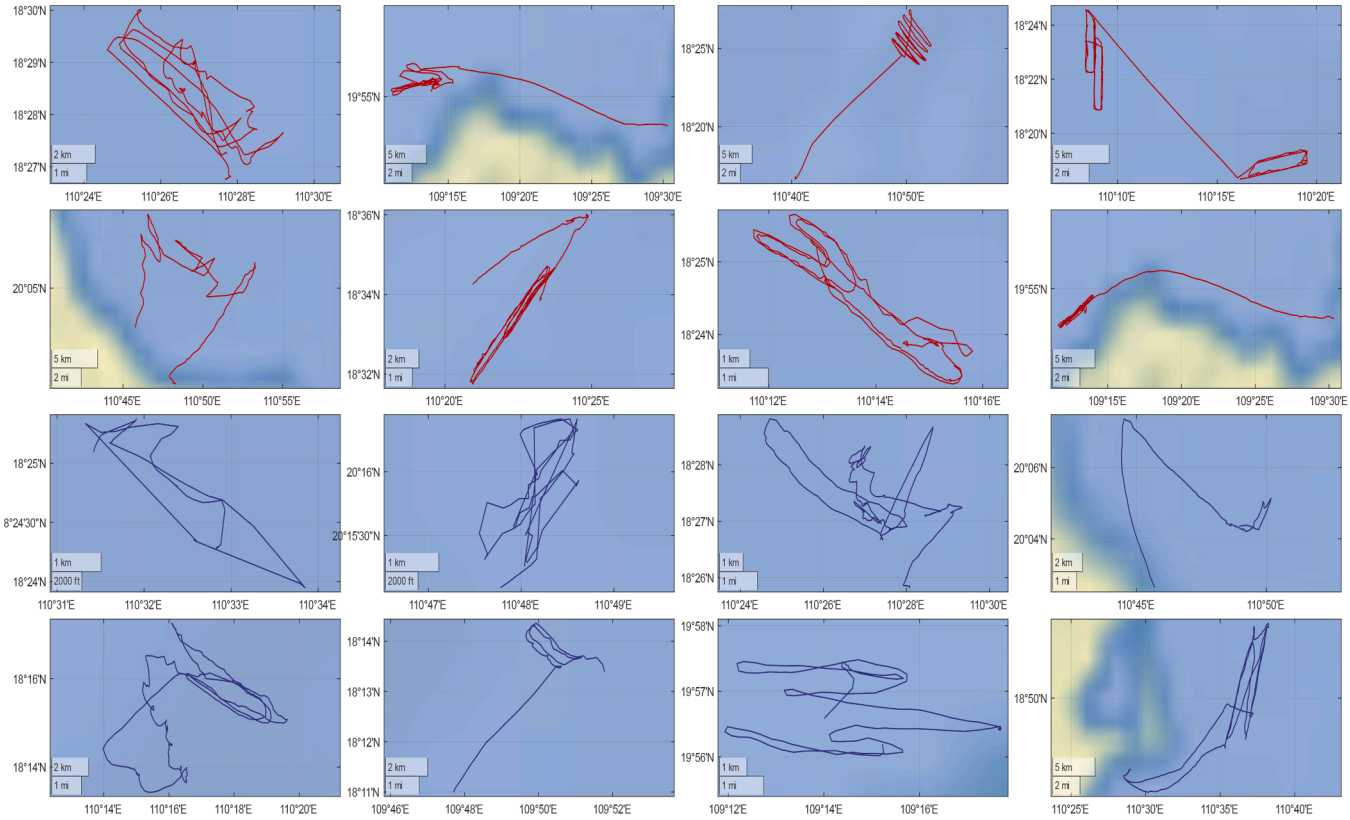


Fig. 14. Visualization of error trajectories.

show that the vessel may have been affected by ocean currents, winds, tides, or other external disturbances in the course of its voyage, which deviated from the conventional navigation logic. In this case, single-trawler may exhibit atypical movement patterns, while non-single-trawler fishing vessels may also be forced to adjust their course due to environmental factors and exhibit trajectory characteristics similar to those of single-trawler. These abnormal routes usually do not conform to conventional logical expectations, making it difficult for the model to classify them correctly without contextual information. It is worth noting that such trajectory changes caused by the external environment are not uncommon in real marine environments, thus placing higher requirements on the generalization ability of the model.

7. Conclusion

In order to explore the practical challenges faced by intelligent supervision of illegal activities of single trawler fishing vessels, this study proposes the SeaTraNet deep learning model, which achieves the collaborative modeling of local and global trajectory features through a dual-path fusion architecture. This model combines with the constructed TrawlTag dataset, forms an end-to-end solution for anomaly recognition of specific vessel types. A large number of experiments have confirmed that this solution not only leads in recognition accuracy, but also performs well in terms of inference efficiency and model interpretability, laying a solid foundation for practical applications. However, despite the encouraging results achieved, there are still some practical issues that need to be addressed before the application can be carried out. Firstly, in short time series or situations with sparse data, the behavior patterns of ships are difficult to be fully presented, which makes the model prone to being disturbed by local anomalies. Secondly, in the real sea area, affected by environmental factors such as ocean currents and wind waves, the trajectories deviate from the regular logic, causing different types of ships to exhibit similar movement characteristics, thereby weakening the discriminative and generalization capabilities of the model. Thirdly, the current datasets used in this study are primarily collected from the waters around the Beibu Gulf and Hainan Island and mainly focus on single-trawler fishing vessels. The geographical and operational homogeneity of the data may limit the generalization of SeaTraNet to other regions and vessel types. The future work can focus on addressing three key issues: Firstly, we will conduct research on the issue of intermittent trajectories. By leveraging advanced feature matching and identity association technologies, we will recognize the interrupted signal segments belonging to the same vessel, and on this basis, reconstruct and complete the missing trajectory sequences. Secondly, the AIS trajectories will be jointly modeled with multi-modal data such as satellite remote sensing, radar, meteorology, and ocean currents, in order to integrate more environmental and observational information, thereby enhancing the model's ability to adapt to complex sea conditions and cross-regional scenarios. Thirdly, we will explore more advanced semi-supervised data generation and expansion strategies to further enrich the dataset, and enhance its robustness across various scenarios. In addtion, we are committed to human-machine collaborative review and feedback loops, enabling the model to continuously iterate and optimize in actual regulatory scenarios, ultimately forming a more reliable, interpretable and applicable large-scale maritime law enforcement intelligent monitoring system.

CRediT authorship contribution statement

Lingkai Kong: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Zhuhua Hu:** Writing – review & editing, Software, Investigation, Funding acquisition, Formal analysis, Data curation; **Yaochi Zhao:** Supervision, Resources, Investigation, Formal analysis, Conceptualization; **Wei Wu:** Visualiza-

tion, Supervision, Investigation, Data curation; **Yanming Gu:** Supervision, Resources, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

Acknowledgement

This research was supported by the National Natural Science Foundation of China 62361024 and 62161010, the Key Research and Development Project of Hainan Province (ZDYF2022GXJS348 and ZDYF2024GXJS021), the National Key Research and Development Program of China (2022YFD2400504), and the Hainan University Postgraduate Innovative Talent Cultivation Project (HDJXAL2024Y0007, HDZBJC2024Y0008). The authors would like to thank the referees for their constructive suggestions.

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.oceaneng.2025.123947.

References

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., Samek, W., 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS ONE 10 (7), e0130140.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Buscaldi, D., Dessí, D., Motta, E., Murgia, M., Osborne, F., Recupero, D.R., 2024. Citation prediction by leveraging transformers and natural language processing heuristics. Inf. Process. Manage. 61 (1), 103583.

Dao, T., 2023. Flashattention-2: faster attention with better parallelism and work partitioning. arXiv preprint arXiv:2307.08691

Feng, S., Miao, C., Zhang, Z., Zhao, P., 2024. Latent diffusion transformer for probabilistic time series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38, pp. 11979–11987.

Gu, Y., Hu, Z., Zhao, Y., Liao, J., Zhang, W., 2024. Mfgtn: a multi-modal fast gated transformer for identifying single trawl marine fishing vessel. Ocean Eng. 303, 117711.

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C., 2020. GhostNet: more features from cheap operations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1580–1589.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

He, P., Chopin, F., Suuronen, P., Ferro, R. S.T., Lansley, J., 2021. Classification and illustrated definition of fishing gears. FAO Fisheries and Aquaculture Technical Paper (672), I–94.

He, Y., Yu, H., Liu, X., Yang, Z., Sun, W., Anwar, S., Mian, A., 2025. Deep learning based 3D segmentation in computer vision: a survey. Inf.n Fusion 115, 102722.

Hou, J., Zhou, H., Grifoll, M., Zhou, Y., Liu, J., Ye, Y., Zheng, P., 2025. A transformer-VAE approach for detecting ship trajectory anomalies in cross-sea bridge areas. J. Mar. Sci. Eng. 13 (5), 849.

Jieyang, P., Kimmig, A., Dongkun, W., Niu, Z., Zhi, F., Jiahai, W., Liu, X., Ovtcharova, J., 2023. A systematic review of data-driven approaches to fault diagnosis and early warning. J. Intell. Manuf. 34 (8), 3277–3304.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science 349 (6245), 255–260.

Kang, H., Kang, P., 2024. Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism. Knowl. Based Syst. 290, 111507.

Kashi, M., Lahmiri, S., Ait Mohamed, O., 2025. Comprehensive analysis of transformer networks in identifying informative sentences containing customer needs. Expert Syst. Appl. 273, 126785.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Li, X., Li, S., Song, S., Yang, J., Ma, J., Yu, J., 2024a. Pmet: precise model editing in a transformer. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38, pp. 18564–18572.

Li, Y., Wang, J., Li, T., Fu, Z., 2024b. TrAISformer: spatio-temporal ship trajectory prediction based on transformer. In: 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT). IEEE, pp. 1099–1104.

Lin, S., Lin, W., Wu, W., Wang, S., Wang, Y., 2024. Petformer: long-term time series forecasting via placeholder-enhanced transformer. IEEE Trans. Emerg. Top. Comput. Intell. 9 (2), 1189–1201

Lin, S., Lin, W., Wu, W., Zhao, F., Mo, R., Zhang, H., 2023. SegRNN: Segment recurrent neural network for long-term time series forecasting. arXiv preprint arXiv:2308.11200

Liu, H., Wu, C., Li, B., Zong, Z., Shu, Y., 2025. Research on ship anomaly detection algorithm based on transformer-GSA encoder. IEEE Trans. Intell. Transp. Syst.. 26 (6), 8752–8763

Liu, T., Wang, Y., Sun, J., Tian, Y., Huang, Y., Xue, T., Li, P., Liu, Y., 2024. The role of transformer models in advancing blockchain technology: a systematic survey. arXiv preprint arXiv:2409.02139

Mazzarella, F., Vespe, M., Santamaria, C., 2015. Sar ship detection and self-reporting data fusion based on traffic knowledge. IEEE Geosci. Remote Sens. Lett. 12 (8), 1685–1689.

Nguyen, K.-B., Yang, J.-S., 2023. Boosting semi-supervised learning by bridging high and low-confidence predictions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1028–1038.

Nie, J., Jiang, J., Li, Y., Wang, H., Ercisli, S., Lv, L., 2025. Data and domain knowledge dual-driven artificial intelligence: survey, applications, and challenges. Exp. Syst. 42 (1), e13425.

Oko, K., Song, Y., Suzuki, T., Wu, D., 2024. Pretrained transformer efficiently learns low-dimensional target functions in-context. Adv. Neural Inf. Process. Syst. 37, 77316–77365.

Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A., 2018. Film: visual reasoning with a general conditioning layer. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32.

Piao, X., Chen, Z., Murayama, T., Matsubara, Y., Sakurai, Y., 2024. Fredformer: frequency debiased transformer for time series forecasting. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 2400–2410.

Ribeiro, C.V., Paes, A., de Oliveira, D., 2023. Ais-based maritime anomaly traffic detection: a review. Expert Syst. Appl. 231, 120561.

Rong, H., Teixeira, A.P., Soares, C.G., 2024. A framework for ship abnormal behaviour detection and classification using AIS data. Reliab. Eng. Syst. Saf. 247, 110105.

Rybicki, T., Masek, M., Lam, C.P., 2024. Maritime behaviour anomaly detection with seasonal context. International Society of Photogrammetry and Remote Sensing.

Salman, H.A., Kalakech, A., Steiti, A., 2024. Random forest algorithm overview. Babylonian J. Mach. Learn. 2024, 69–79.

Samek, W., Wiegand, T., Müller, K.-R., 2017. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296

Shaheen, F., Verma, B., Asafuddoula, M., 2016. Impact of automatic feature extraction in deep learning architecture. In: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 1–8, year = 2016, organization = IEEE.

Shahin, N., Ismail, L., 2024. From rule-based models to deep learning transformers architectures for natural language processing and sign language translation systems: survey, taxonomy and performance evaluation. Artif. Intell. Rev. 57 (10), 271.

Shahir, A.Y., Tayebi, M.A., Glässer, U., Charalampous, T., Zohrevand, Z., Wehn, H., 2019. Mining vessel trajectories for illegal fishing detection. In: 2019 IEEE International Conference on Big Data (Big Data). IEEE, pp. 1917–1927.

Shazeer, N., 2020. Glu variants improve transformer. arXiv preprint arXiv:2002.05202

Sun, Y., Zhao, Y., Hu, Z., Wu, W., Xia, J., Wang, Y., 2024. Ssrlm: a self-supervised representation learning method for identifying one ship with multi-mmsi codes. Ocean Eng. 312, 119186.

Van Engelen, J.E., Hoos, H.H., 2020. A survey on semi-supervised learning. Mach. Learn. 109 (2), 373–440.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30 (6000–6010).

Wan, H., Xiao, Y., Xu, W., Fu, S., Gao, P., 2023. A cognitive approach for identification of ship illegal behaviors by using knowledge graph. In: 2023 7th International Conference on Transportation Information and Safety (ICTIS). IEEE, pp. 732–737.

Wan, Y., Liu, Y., Chen, Z., Chen, C., Li, X., Hu, F., Packianather, M., 2024. Making knowledge graphs work for smart manufacturing: research topics, applications and prospects. J. Manuf. Syst. 76, 103–132.

Wang, H., Liu, Y., 2021. Abnormality behavior recognition method for ship in bridge waters considering small-sample problem. In: 2021 7th International Conference on Hydraulic and Civil Engineering & Smart Water Conservancy and Intelligent Disaster Reduction Forum (ICHCE & SWIDR). IEEE, pp. 1275–1280.

Wang, Y., Gong, C., Ji, X., Yuan, Q., 2025. Text classification for evaluating digital technology adoption maturity based on BERT: an evidence of industrial AI from china. Technol. Forecast. Soc. Change 211, 123903.

Wei, Z., Xie, X., Zhang, X., 2022. Maritime anomaly detection based on a support vector machine. Soft Comput. 26 (21), 11553–11566.

Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M., 2022. TimesNet: Temporal 2D-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186

Zeng, A., Chen, M., Zhang, L., Xu, Q., 2023. Are transformers effective for time series forecasting? In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 37, pp. 11121–11128.

Zhang, B., Wang, Y., Hou, W., Wu, H., Wang, J., Okumura, M., Shinozaki, T., 2021. Flexmatch: boosting semi-supervised learning with curriculum pseudo labeling. Adv. Neural Inf. Process. Syst. 34, 18408–18419.

Zhang, C., Bin, J., Liu, Z., 2024. TrajBERT - DSSM: deep bidirectional transformers for vessel trajectory understanding and destination prediction. Ocean Eng. 297, 117147.