Article

A Transformer–VAE Approach for Detecting Ship Trajectory Anomalies in Cross-Sea Bridge Areas

Jiawei Hou [1,2,3], Hongzhu Zhou [1,2,3], Manel Grifoll [4], Yusheng Zhou [5], Jiao Liu [1,2,3], Yun Ye [1,2,3,*] and Pengjun Zheng [1,2,3,*]

1   Faculty of Maritime and Transportation, Ningbo University, Ningbo 315211, China;
    2211120006@nbu.edu.cn (J.H.); zhouhongzhu@nbu.edu.cn (H.Z.); liujiao@nbu.edu.cn (J.L.)
2   Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University,
    Nanjing 210096, China
3   National Traffic Management Engineering & Technology Research Center Ningbo University Sub-Center,
    Ningbo 315211, China
4   Barcelona School of Nautical Studies, Universitat Politècnica de Catalunya (UPC—BarcelonaTech),
    08003 Barcelona, Spain; manel.grifoll@upc.edu
5   Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong;
    yusheng1.zhou@polyu.edu.hk
*   Correspondence: yeyun1@nbu.edu.cn (Y.Y.); zhengpengjun@nbu.edu.cn (P.Z.)

Abstract: Abnormal ship navigation behaviors in cross-sea bridge waters pose significant threats to maritime safety, creating a critical need for accurate anomaly detection methods. Ship AIS trajectory data contain complex temporal features but often lack explicit labels. Most existing anomaly detection methods heavily rely on labeled or semi-supervised data, thus limiting their applicability in scenarios involving completely unlabeled ship trajectory data. Furthermore, these methods struggle to capture long-term temporal dependencies inherent in trajectory data. To address these limitations, this study proposes an unsupervised trajectory anomaly detection model combining a transformer architecture with a variational autoencoder (transformer–VAE). By training on large volumes of unlabeled normal trajectory data, the transformer–VAE employs a multi-head self-attention mechanism to model both local and global temporal relationships within the latent feature space. This approach significantly enhances the model's ability to learn and reconstruct normal trajectory patterns, with reconstruction errors serving as the criterion for anomaly detection. Experimental results show that the transformer–VAE outperforms conventional VAE and LSTM–VAE in reconstruction accuracy and achieves better detection balance and robustness compared to LSTM—VAE and transformer–GAN in anomaly detection. The model effectively identifies abnormal behaviors such as sudden changes in speed, heading, and trajectory deviation under fully unsupervised conditions. Preliminary experiments using the POT method validate the feasibility of dynamic thresholding, enhancing the model's adaptability in complex maritime environments. Overall, the proposed approach enables early identification and proactive warning of potential risks, contributing to improved maritime traffic safety.

Keywords: AIS data; trajectory reconstruction; anomaly detection; transformer; variational autoencoder

1. Introduction

With the rapid development of global shipping, ships have progressively evolved toward larger sizes, greater specialization, and higher speeds, leading to a continuous

increase in maritime traffic and a more complex navigation environment. Concurrently, the accelerated construction of bridges in coastal waters has significantly transformed the spatial characteristics of navigational channels around bridge areas, including changes in current velocity, flow direction, and other hydrological conditions. These changes have introduced greater uncertainty into ship navigation, particularly when passing through bridge regions, where precise control of speed and heading is critically required. However, due to external interference or operational errors, ships frequently deviate from established routes, exhibit abnormal speed variations, or engage in reverse-direction navigation near bridge areas, thereby elevating the risk of ship-bridge collisions. Such incidents not only threaten infrastructure and vessel safety but also cause substantial economic losses, environmental damage, and human casualties.

Typically, ship anomaly detection involves analyzing historical ship trajectory data, including speed, heading, and spatial trajectory patterns, to determine whether ship movements conform to normal navigational patterns. AIS trajectory data contain complex temporal dependencies and strong correlations across multiple features. Yet, most existing anomaly detection methods for multivariate sequences rely heavily on labeled or partially labeled data, significantly limiting their practical applicability in real-world maritime scenarios where explicit anomaly labels are unavailable.

Additionally, current anomaly detection models, primarily based on recurrent neural networks (RNN) and their variants such as LSTM and GRU, exhibit limitations in capturing long-term temporal dependencies and adequately learning complex trajectory features, resulting in suboptimal detection accuracy and limited generalization capabilities.

To address these challenges, this study proposes a novel unsupervised anomaly detection method based on integrating a transformer with a variational autoencoder (transformer–VAE), specifically designed for ship trajectory anomaly detection in cross-sea bridge navigation areas. By learning long-term dependency features from extensive historical normal trajectory data, the proposed transformer–VAE leverages the multi-head self-attention mechanism inherent in the transformer [1] architectures to accurately capture complex temporal dynamics in trajectory latent feature spaces, significantly enhancing the model's capability for trajectory reconstruction. Reconstruction errors are subsequently utilized as an anomaly detection metric, and anomaly thresholds are determined using a quantile-based strategy (e.g., 90th, 95th, or 98th percentiles), thus enabling fully unsupervised anomaly detection without relying on explicit anomaly labels.

The structure of the paper is organized as follows: the introduction is followed by Section 2, which reviews studies on the ship trajectory anomaly detection. Section 3 describes the research methodology employed in this study. Section 4 details the practical application results of the model. Section 5 discusses the strengths and weaknesses of the approach taken in this paper. Finally, Section 6 concludes with the main findings and suggests directions for future research.

2. Literature Review

Anomaly detection aims to identify patterns in data that deviate from normal behaviors. This task has important applications across various domains, including medical insurance, cybersecurity intrusion detection, trajectory anomaly detection, and military surveillance for monitoring adversarial activities. In the context of vessel anomaly detection, analyzing AIS data enables the learning of vessel movement patterns and the application of relevant algorithms to extract anomalous vessel states.

In recent years, several scholars have reviewed and summarized the detection methods for vessel anomalous behaviors, which can be mainly divided into the following two categories: rule-based methods and data-driven methods. Rule-based vessel trajectory anomaly

detection methods identify abnormal trajectories by applying predefined navigational rules. For instance, He et al. [2] proposed a rule-based collision-avoidance path planning method for multi-ship encounters that incorporates COLREGs and vessel maneuverability, which can serve as a foundation for anomaly detection in structured navigational scenarios. These approaches typically rely on fixed parameters and thresholds, such as deviations in speed or course. However, due to varying navigational conditions across different maritime regions, such methods often lack robustness and generalization capability in dynamic environments.

Data-driven approaches can be broadly categorized into statistical analysis methods and machine learning methods. Statistical approaches generally assume that normal vessel behaviors follow specific probabilistic distributions, and anomalies are identified through statistical inference or density-based modeling. Representative studies include Sidibé and Gao [3], who constructed statistical models from historical data and applied hypothesis testing for anomaly identification. Gao et al. [4] applied an ordered probit model to analyze collision accident reports from the Yangtze River estuary, identifying key environmental and vessel-related factors that influence incident severity. Mascaro et al. [5] modeled trajectories as dynamic Bayesian networks using time-series data; Xie et al. [6] combined Gaussian mixture models (GMM) with an improved autoencoder and used dynamic time warping (DTW) to measure reconstruction deviations; Smith et al. [7] employed Kalman filters to implement a gated anomaly detection mechanism; and Guillarme [8] proposed a learning framework based on satellite AIS data, integrating point-of-interest extraction, DTW similarity, and Markov modeling for anomaly detection. Rong et al. [9,10] utilized Gaussian process-based trajectory prediction models and maritime traffic probability models to detect abnormal courses based on route and waypoint features. Anneken et al. [11] employed KDE and GMM to detect anomalous trajectory points, though their method did not account for entire trajectory structures. Nguyen et al. [12] developed the GeoTrackNet model, which identifies low-probability trajectories as anomalies based on probabilistic modeling. In general, statistical methods offer efficient computation and interpretable models but are often limited by their reliance on strong distributional assumptions, especially in complex and dynamic maritime environments. This limits their applicability in real-world navigation scenarios with high uncertainty and pattern diversity.

Machine learning methods aim to construct normal behavior models using historical motion data and compare incoming data against these models to identify significant deviations. These methods are typically divided into supervised and unsupervised learning approaches. In the category of unsupervised learning, Liu et al. [13] optimized K-means clustering using an immune genetic algorithm to extract core trajectories and identify anomalies based on spatial deviation. Fu et al. [14] proposed a similarity-based approach, while Djenouri et al. [15] trained multiple CNN models on clustered data and aggregated their outputs for anomaly detection. Lei et al. [16] introduced the MT-MAD framework, combining DBSCAN clustering and anomaly scoring to detect abnormal trajectories. Liu et al. [17] applied density-based clustering and integrated speed and course features for behavioral analysis. To handle large-scale datasets, Nooshin and Li [18,19] proposed improvements to DBSCAN. Kontopoulos et al. [20] refined trajectory continuity modeling through heading change point extraction and trajectory line clustering. Yang et al. [21] introduced DBTCAN, which uses the Hausdorff distance to measure similarity between trajectories of varying lengths. Jin [22] applied the DP algorithm to identify turning points and constructed anomaly factors using DBSCAN and KDE. Bai et al. [23] developed an adaptive threshold DBSCAN method incorporating DTW and an improved K-adaptive nearest neighbor algorithm for improved clustering accuracy.
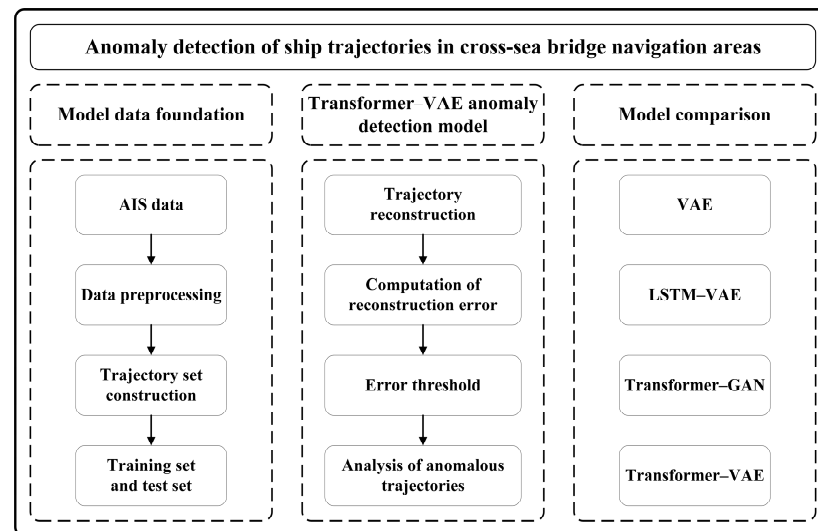
In supervised learning, deep neural networks have demonstrated strong capabilities in modeling temporal dependencies and identifying anomalies. Singh et al. [24] utilized a regression-based RNN model to predict future trajectories and assess uncertainty. Zhao and Shi [25] combined DBSCAN with LSTM networks for trajectory clustering and prediction. Mantecón et al. [26] designed a CNN-based framework to classify vessel movement states. Nguyen et al. [27] proposed a VRNN-based model with a backward detection strategy for adaptive anomaly modeling. Karataş [28] employed LSTM to predict arrival time and future positions for anomaly detection. Rhodes et al. [29] developed a fuzzy neural network for behavioral prediction. Xie et al. [30] integrated graph neural networks (GNNs) with a dynamic threshold mechanism to enhance detection accuracy. Liang and Tang et al. [31,32] used LSTM models to predict trajectories and identify anomalies based on deviation thresholds. Maganaris et al. [33] applied deep RNNs to reconstruct motion patterns and detect anomalies via reconstruction error. In the domain of image-based trajectory representation, Sadeghi and Matwin [34] transformed trajectories into Hankel matrices and applied image segmentation to identify anomaly peaks, while Liang et al. [35] employed a WGAN to train on normal trajectory images and detect anomalies through reconstruction differences. Chen et al. [36] developed a YOLO-based rotation-aware ship detection model that integrates a feature decoupling head and attention mechanisms to accurately detect inclined ships in maritime surveillance videos. In addition, reinforcement learning-based approaches have also been explored for trajectory behavior modeling. Chen et al. [37] proposed an A-guided double deep Q-network (A-DDQN) model for intelligent route generation under dynamic maritime conditions. Although originally designed for navigation optimization, such models provide a behavior-driven formulation of vessel motion and may offer complementary insights for identifying trajectory-level anomalies.

Despite the notable progress made by existing trajectory anomaly detection models, several challenges remain. RNNs and their variants, such as LSTM and GRU, have been widely used to capture the temporal dependencies in sequential AIS data. However, their ability to model long-range dependencies is hindered by issues such as gradient vanishing and explosion, limiting their effectiveness in capturing global navigational patterns. Furthermore, although unsupervised learning methods do not require labeled anomalies for training, many studies still rely on labeled anomaly data for model evaluation and threshold selection, which introduces a disconnect between academic research and real-world applications where labeled anomalies are often unavailable. Consequently, the robustness and reliability of these models under fully unlabeled, real-world conditions remain insufficiently validated.

To address these limitations, this study proposes a novel unsupervised trajectory anomaly detection framework that integrates a transformer-based encoder with a variational autoencoder. The core idea is to learn the spatiotemporal navigation patterns of normal vessel trajectories and detect anomalies based on reconstruction errors, enabling effective anomaly detection in label-scarce maritime scenarios.

3. Methods

As illustrated in Figure 1, this study proposes an unsupervised ship trajectory anomaly detection framework based on a transformer–VAE, aiming to leverage the reconstruction capability of deep learning models to accurately identify abnormal trajectories of vessels navigating under sea-crossing bridges.

```
┌─────────────────────────────────────────────────────────────────┐
│   ╔═══════════════════════════════════════════════════════════╗   │
│   ║  Anomaly detection of ship trajectories in cross-sea       ║   │
│   ║                 bridge navigation areas                    ║   │
│   ╚═══════════════════════════════════════════════════════════╝   │
│                                                                   │
│   ┌ Model data foundation ┐  ┌Transformer–VAE anomaly┐  ┌ Model comparison ┐ │
│                              │  detection model      │                       │
│   ┌──────────────────┐       ┌──────────────────┐       ┌──────────────────┐ │
│   │     AIS data     │       │    Trajectory    │       │       VAE        │ │
│   └────────┬─────────┘       │  reconstruction  │       └──────────────────┘ │
│            ↓                 └────────┬─────────┘                             │
│   ┌──────────────────┐                ↓                 ┌──────────────────┐ │
│   │ Data preprocessing│      ┌──────────────────┐       │     LSTM–VAE     │ │
│   └────────┬─────────┘       │  Computation of  │       └──────────────────┘ │
│            ↓                 │ reconstruction error│                          │
│   ┌──────────────────┐       └────────┬─────────┘       ┌──────────────────┐ │
│   │  Trajectory set  │                ↓                 │  Transformer–GAN │ │
│   │   construction   │      ┌──────────────────┐       └──────────────────┘ │
│   └────────┬─────────┘       │  Error threshold │                             │
│            ↓                 └────────┬─────────┘       ┌──────────────────┐ │
│   ┌──────────────────┐                ↓                 │  Transformer–VAE │ │
│   │   Training set   │      ┌──────────────────┐       └──────────────────┘ │
│   │   and test set   │       │ Analysis of anomalous │                        │
│   └──────────────────┘       │    trajectories  │                            │
│                              └──────────────────┘                            │
└─────────────────────────────────────────────────────────────────┘
```

Figure 1. The flowchart of ship trajectory anomaly detection in cross-sea bridge navigation areas.

The proposed methodology consists of three main stages: trajectory data preprocessing and trajectory set construction, model development and training, and anomaly detection based on reconstruction errors. Firstly, AIS trajectory data undergoes strict preprocessing procedures, including denoising, interpolation, and trajectory segmentation, followed by latitude resampling to ensure a consistent number of time steps for all input trajectories. The preprocessed data constitute the ship trajectory dataset, with the training set carefully selected through manual inspection and expert evaluation to ensure that the model learns only normal navigation patterns compliant with maritime regulations. Subsequently, the transformer–VAE model is constructed, where the transformer encoder captures long-range dependencies within trajectory sequences using a self-attention mechanism, while the VAE structure models the latent distribution of normal trajectories, enhancing the model's ability to distinguish anomalies. The model is trained exclusively on normal trajectories, with the loss function comprising mean squared error (MSE) and Kullback–Leibler divergence (KLD). To evaluate the effectiveness of the proposed method in learning normal trajectory patterns and achieving high reconstruction accuracy, LSTM–VAE and standard VAE models are used as baseline models. The focus is on comparing the performance of different encoder architectures in trajectory modeling and reconstruction tasks. In the anomaly detection phase, the transformer–VAE model identifies trajectory anomalies in the test set based on reconstruction errors. Due to the absence of ground truth labels, a quantile-based thresholding strategy is initially applied, where trajectories exceeding the threshold are classified as anomalies and subsequently visualized. To further evaluate detection performance, comparative experiments are conducted using two representative generative models—LSTM–VAE and transformer–GAN—for benchmark analysis. Additionally, to enhance adaptability in dynamic maritime environments, a preliminary experiment is conducted using the POT method from extreme value theory to construct a dynamic threshold, validating its feasibility and potential for broader application.

Most research on vessel anomaly detection assumes that abnormal trajectories result from abnormal vessel behaviors, and the detected anomalies in trajectories indicate abnormal behaviors. Therefore, this study does not distinguish between abnormal trajectories and abnormal behaviors. Vessel abnormal trajectories typically refer to deviations from expected or typical routes during navigation, with these anomalies manifesting as significant deviations in position, direction, or speed. This research primarily focuses on three representative types of vessel trajectory anomalies: position deviation, heading anomaly, and speed anomaly.

3.1. Data Preprocessing

In the data preprocessing stage, it is first necessary to eliminate evidently erroneous records to ensure the accuracy and reliability of trajectory anomaly detection. This includes removing data points with invalid latitude or longitude values (e.g., longitude exceeding $\pm 180°$, latitude exceeding $\pm 90°$), as well as those located outside the defined study area. For abnormal speed data, a maximum speed threshold of 30 knots is set as the upper limit for vessel speed. Values exceeding this threshold are typically caused by AIS transmission errors or noise interference, rather than actual vessel behavior, and should therefore be removed. In addition, it is important to address positional drift in AIS data, where abrupt changes in latitude and longitude cause trajectory points to deviate significantly from the true sailing path. Such errors are often the result of AIS device malfunctions or signal interference, rather than genuine navigational anomalies. Accordingly, these abruptly shifted or drifting points should be excluded to avoid misleading subsequent trajectory modeling and anomaly detection processes.

To maintain the temporal continuity of trajectory data, missing trajectory points were interpolated after the removal of outliers. Linear interpolation was applied to regular numerical features, while a circular interpolation strategy was used for COG due to its periodic nature. Specifically, COG values were first mapped onto the unit circle, and interpolation was performed separately on the sine and cosine components before converting them back to angular form. This approach ensures smooth directional transitions and avoids angular discontinuities. In addition, to meet the model's requirements for consistent input structure and fixed sequence length, all trajectories were resampled at equal intervals along the latitude dimension, ensuring that each sequence contains the same number of trajectory points.

To address data privacy concerns, all personally identifiable information (PII) was removed from the AIS dataset prior to analysis. This includes vessel identifiers such as MMSI numbers, IMO codes, and ship names. Only non-identifiable behavioral features—including geographic position, SOG, and COG—were retained for trajectory modeling. The data processing procedure complies with data minimization and anonymization principles in line with ethical research standards and the relevant provisions of the General Data Protection Regulation (GDPR).

This study selected four key features as model input, as shown in Equation (1):

$$Tra_x = (LON_x, LAT_x, SOG_x, COG_x) \tag{1}$$

3.2. Transformer-Based Variational Autoencoder Model

A variational autoencoder (VAE) is a type of deep Bayesian network that extends the conventional autoencoder through variational inference. It maps input data into a probabilistic distribution to learn the variability of the input data, making it a probabilistic counterpart of the autoencoder (AE). Unlike a standard AE, a VAE leverages deep neural networks to model two complex conditional probability densities separately, allowing it to learn meaningful latent representations while maintaining probabilistic consistency.

Existing VAE-based anomaly detection models struggle to propagate long-term dependencies among latent variables in the latent space. To address this limitation, this study proposes a ship trajectory anomaly detection model that integrates a transformer encoder with a VAE. The transformer model is entirely built on the attention mechanism, leveraging self-attention and feedforward neural networks (FNNs) to enable self-learning and self-adjustment without relying on prior historical experience. Additionally, its outstanding parallelization capability enhances computational efficiency. In the proposed model, the trajectory time-series feature hidden states output by the transformer encoder and the

sampled latent variables from the previous time step are used to generate the approximate posterior distribution of the latent variables at the current time step. The inference network takes the hidden states of the temporal features and the sampled approximate posterior latent variables as inputs, reconstructing trajectory data through the generative network. The key advantage of this approach is that it leverages the multi-head self-attention mechanism of the transformer encoder to establish long-term dependencies among latent variables in the VAE latent space, overcoming the limitations of traditional VAE-based methods in capturing temporal dependencies.

In the inference network, the transformer model first encodes the input trajectory sequence $x_{1:T} \in \mathbb{R}^{P \times T}$, where each time step is represented by a trajectory feature vector $x_t = [LON_t, LAT_t, SOG_t, COG_t]$. Here, T denotes the trajectory length and P represents the dimensionality of trajectory features. The encoder outputs a sequence of temporal hidden representations $e_{1:T} = [e_1, \ldots, e_t, \ldots, e_T] \in \mathbb{R}^{d \times T}$, which are used to capture global dependencies within the trajectory sequence, where d represents the output dimension of the transformer encoder. To estimate the approximate posterior distribution of the latent variable at each time step t, the hidden feature vector e_t is concatenated with the sampled latent variable from the previous time step z_{t-1}, and the combined vector is passed through an MLP to compute the parameters of the approximate posterior distribution of z_t. In the generative network, the sampled latent variable z_t at each time step, together with the corresponding temporal feature e_t, is the input to the decoder. A final linear mapping layer is then applied to project the decoded features back to the original data space, producing the reconstructed value for each time step. MSE and KLD are used as loss functions to optimize the parameters of both the inference and generative networks. The reconstruction error is then used as an anomaly score to detect abnormal ship trajectory patterns.

To enable nonlinear transformation and temporal dependency modeling between the latent variables z_{t-1} and z_t, the sampled latent variable from the previous time step z_{t-1} is concatenated with the current hidden representation e_t, which is generated by the transformer encoder. This concatenated vector is then used to estimate the latent variable z_t at the current time step, thereby capturing temporal dependencies within the latent space. Here, e_t is a hidden feature vector obtained from the temporal representation sequence $e_{1:T}$ output by the transformer encoder. The computation of $e_{1:T}$ proceeds as follows.

To incorporate the positional information of each time step in the trajectory sequence and enable the model to have sequence-awareness, absolute positional encoding is applied to the trajectory sequence $x_{1:T}$. The encoded result is denoted as $x_{1:T}^p$, as shown in Equation (2):

$$x_{1:T}^p = (W_x x_{1:T} + b) + w_p, \tag{2}$$

where $W_x \in \mathbb{R}^{d \times T}$ represents the network parameters, $b \in \mathbb{R}^{d \times T}$ is the bias term, and $w_p \in \mathbb{R}^{d \times T}$ denotes the positional encoding matrix. The positional encoding is constructed using trigonometric functions, where sine functions are used for even-indexed positions and cosine functions for odd-indexed positions. The specific formulation is as follows:

$$w_p = \begin{bmatrix} sin(w_1 \cdot 1) & \cdots & sin(w_1 \cdot t) & \cdots & sin(w_1 \cdot T) \\ cos(w_1 \cdot 1) & \cdots & cos(w_1 \cdot t) & \cdots & cos(w_1 \cdot T) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ sin(w_k \cdot 1) & \cdots & sin(w_k \cdot t) & \cdots & sin(w_k \cdot T) \\ cos(w_k \cdot 1) & \cdots & cos(w_k \cdot t) & \cdots & cos(w_k \cdot T) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ sin(w_{d/2} \cdot 1) & \cdots & sin(w_{d/2} \cdot t) & \cdots & sin(w_{d/2} \cdot T) \\ cos(w_{d/2} \cdot 1) & \cdots & cos(w_{d/2} \cdot t) & \cdots & cos(w_{d/2} \cdot T) \end{bmatrix}, \tag{3}$$

where $w_k = 1/10,000^{\frac{2k}{d}}, 1 \leq k \leq d/2, 1 \leq t \leq T$. Next, the multi-head self-attention mechanism is applied to the position-encoded trajectory sequence to model and extract the global dependencies between different time steps of the trajectory. The transformer encoder maps each trajectory input sequence $x^p_{1:T}$ to a set of temporal hidden feature vectors $e_{1:T} = [e_1, \ldots, e_t, \ldots, e_T] \in \mathbb{R}^{d \times T}$, where the computation process is outlined in Equations (4)–(7).

$$Q_h = x^p_{1:T} W^Q_h, K_h = x^p_{1:T} W^K_h, V_h = x^p_{1:T} W^V_h, \tag{4}$$

$$O_h = Attention(Q_h, K_h, V_h) = softmax\left(\frac{Q_h K^T_h}{\sqrt{d_K}}\right) V_h, \tag{5}$$

$$G_{1:T} = Concat(O_1, \cdots O_H) W^O, \tag{6}$$

$$I_{1:T} = LN\left(G_{1:T} + x^p_{1:T}\right), e_{1:T} = LN(FFN(I_{1:T}) + I_{1:T}) \tag{7}$$

Equation (4) maps $x^p_{1:T}$ to the query, key, and value matrices using the network parameter matrices W^Q_h, W^K_h, and W^V_h, respectively. The number of attention heads in the multi-head self-attention mechanism is denoted as H, and O_h represents the output of the $h(h \in [1, H])$ self-attention head. The number of rows in the key matrix is denoted as d_K. The operation $Concat(.)$ represents the concatenation of matrices, while W^O denotes the network parameters. The features $e_{1:T}$ extracted using the residual network, as formulated in Equation (7), are then computed. Here, $LN(.)$ represents layer normalization, and $FFN(.)$ denotes a feedforward neural network. The approximate posterior distribution generated by the inference network at time t is expressed as follows:

$$q_\phi(z_t|z_{t-1}, x_{1:T}) = N\left(z_t \big| \psi_\mu(z_{t-1}, e_t), diag(\psi_\sigma(z_{t-1}, e_t))\right), \tag{8}$$

where $\psi_\mu(.)$ represents a neural network composed of MLP and linear layers, which is used to generate the mean value of the latent variable. The function $\psi_\sigma(.)$ is implemented as an MLP and a fully connected layer with a Softplus activation function, which generates the standard deviation of the latent variable. This formulation introduces a Markovian dependency in the latent space, where each latent variable z_t is conditioned on the previous latent state z_{t-1} and the time-dependent feature representation e_t extracted by the transformer encoder. By explicitly incorporating z_{t-1} into the approximate posterior distribution, temporal correlations among latent variables can be continuously propagated within the latent space, thereby enhancing the model's ability to capture complex temporal dynamics.

The generative network is composed of a multi-layer perceptron and is designed to reconstruct the ship trajectory sequence in a step-by-step manner based on the latent variables and encoded features. Unlike traditional VAE models that rely solely on latent variables for reconstruction, the proposed model integrates the temporal hidden representations output by the transformer encoder in the inference network. This design enables the reconstruction process to jointly leverage the global temporal structure of the trajectory and the latent representations. To enhance the model's capacity for nonlinear representation, ReLU activation functions are applied between hidden layers, thereby improving the ability to capture complex trajectory patterns. Finally, a linear output layer maps the high-dimensional representation back to the original trajectory features, including longitude, latitude, SOG, and COG. The sampled latent variables are computed according to Equation (9), and the reconstructed values at each time step are obtained using Equation (10):

$$z_t = \mu_{z_t} + \sigma_{z_t} \bigodot \varepsilon, \tag{9}$$

$$x'_{1:T} = W_3 \cdot ReLU[W_2 \cdot ReLU(W_1 \cdot (W_0[e_t, z_t] + b_0) + b_1) + b_2] + b_3, \tag{10}$$

where $\varepsilon \sim N(0, I)$, μ_{z_t} and σ_{z_t} represent the mean and standard deviation of the approximate posterior distribution generated by the inference network.

The model utilizes reconstruction error and KLD as the loss function to optimize the parameters of the inference network and generative network.

$$LOSS = \sum_{t=1}^{T}\left[\left(x_t - x_t'\right)^2 + \beta D_{KL}\left(q_\phi(z_t|z_{t-1}, x_{1:T})\middle|\middle|p_\theta(z_t|z_{t-1})\right)\right] \tag{11}$$

In Equation (11), the coefficient β balances the reconstruction loss and the KLD. A larger β emphasizes the regularization of the latent space, while a smaller β focuses on accurate input reconstruction. In this study, β is set to 0.5 to achieve a balanced trade-off between these two objectives.

The network architecture diagram of the transformer–VAE anomaly detection model is shown in Figure 2.
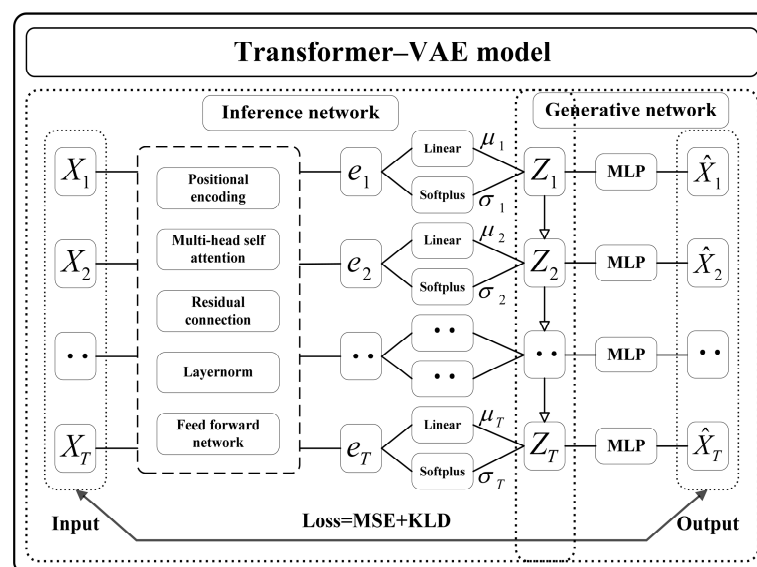
Figure 2. Architecture of the transformer–VAE model.

4. Case Study

4.1. Datasets and Model Parameters

This study focuses on the main navigational span of the Jintang Bridge and its adjacent waters as the designated research area. The Jintang Bridge is located in the coastal waters of Zhejiang Province, China. Spanning approximately 26 km, it connects Jintang Island with the mainland, serving as a major transportation corridor for both land and maritime traffic. The main navigational span, situated above one of the busiest shipping lanes in the region, accommodates frequent two-way passage of various vessel types. Due to the convergence of high vessel density, complex hydrological conditions, and limited navigational space beneath the bridge, the area presents significant challenges for maritime traffic management and safety monitoring. These characteristics make it an ideal and representative case study for evaluating the effectiveness of trajectory anomaly detection methods in high-risk, high-traffic maritime environments.

After data preprocessing, a total of 1400 northbound vessel trajectories were collected, as illustrated in Figure 3. To ensure that the anomaly detection model accurately learns normal vessel navigation patterns, the dataset was meticulously refined through manual screening combined with expert knowledge from maritime authorities. As a result, a

training set consisting of 850 normal trajectories was constructed, while the remaining trajectories were used as a test set for anomaly detection.
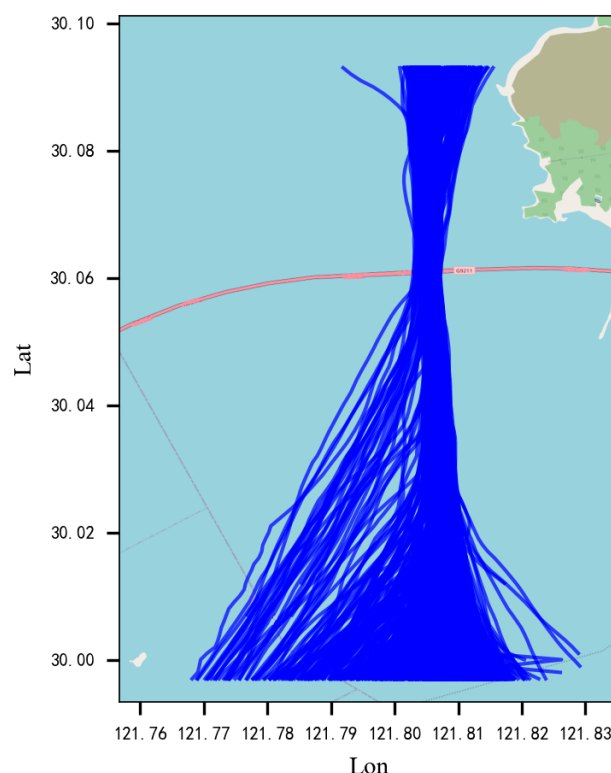
**Figure 3.** AIS trajectories of vessels navigating through the Jintang Bridge.

The input data to the model is structured as a tensor with the shape (N, T, F), where N represents the number of trajectories, T = 95 denotes the number of time steps (i.e., trajectory points per trajectory), and F = 5 indicates the number of features at each time step. The dataset is normalized using min–max scaling. Considering the periodic nature of the COG feature, this study transforms its original range from $[0°, 360°]$ to $[-180°, 180°]$ in order to avoid abrupt numerical transitions during normalization. This transformation preserves the continuity of angular values after normalization and effectively prevents the artificial jump between 0 and 1 that would occur if $0°$ and $360°$ were directly normalized. As a result, the stability of the model in learning periodic features is significantly improved. The output dimension of the transformer encoder was set to 128, with eight self-attention heads. The batch size was set to 64, and the number of training epochs was set to 100. The model was trained using the Adam optimizer with a learning rate of r = 0.001. Detailed model parameters are summarized in Table 1.

Table 1. Parameter settings.

Parameter Name	Parameter Size
Input dim	5
Batch size	64
Hidden dim	128
Latent dim	64
Encoder layers	6
Dropout	0.1
Learning rate	0.001

To evaluate the robustness of the transformer–VAE model with respect to architectural hyperparameters, a sensitivity analysis was conducted on four key parameters: latent dimension (latent_dim), hidden layer dimension (hidden_dim), number of encoder layers (num_layers), and number of attention heads (num_heads). In each experiment, only one parameter was varied while the others were held constant. The average reconstruction error on the test set was used as the evaluation metric.

As shown in Figure 4, the reconstruction errors remained within a narrow range across all configurations, with variation amplitudes generally within 0.02. This indicates that the model exhibits good stability and robustness under a variety of structural configurations. Specifically, changes in latent_dim had a negligible effect on reconstruction accuracy, indicating insensitivity to the size of the latent space. Increasing the hidden_dim slightly improved performance, although the differences were not substantial. When num_layers was set to 6, the model achieved the best performance, while using four or eight layers led to only minor fluctuations. Variations in num_heads had minimal impact on reconstruction accuracy, reflecting strong adaptability in the attention structure. These findings confirm that the transformer–VAE model maintains stable performance under different parameter settings and demonstrates high structural robustness and practical applicability in trajectory anomaly detection tasks.
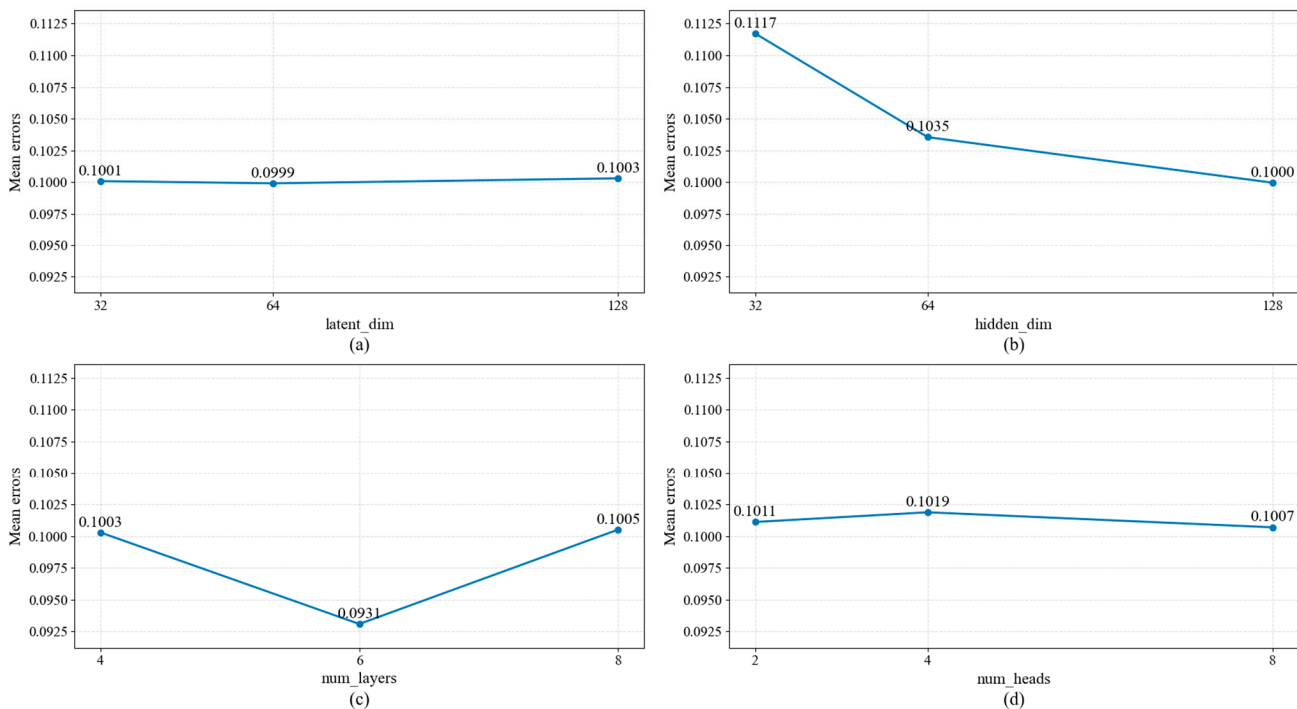

Figure 4. Sensitivity analysis of the transformer–VAE model with respect to (**a**) latent dimension, (**b**) hidden dimension, (**c**) number of encoder layers, and (**d**) number of attention heads.

To evaluate the transformer–VAE model's ability to learn normal trajectory patterns and reconstruct trajectories, comparative experiments were conducted with two baseline models: the traditional VAE and the LSTM–VAE. The traditional VAE uses fully connected neural networks (MLPs) as the encoder and decoder for trajectory data. The LSTM–VAE, a variational autoencoder based on long short-term memory networks, employs LSTM units as both encoder and decoder. During the training process of all three models—VAE, LSTM–VAE, and transformer–VAE—the same set of hyperparameters was used. The reconstruction performance was evaluated using RMSE, MSE, and MAE as the evaluation metrics.

4.2. Experimental Validation of Trajectory Reconstruction

Figure 5 presents a comparison of the reconstruction performance of three models—transformer–VAE, LSTM–VAE, and conventional VAE—on four trajectory features: longitude (LON), latitude (LAT), SOG, and COG, based on a randomly selected test trajectory sample. In the figure, the red curve represents the original observed values, while the other colored curves correspond to the reconstruction results produced by the different models. As shown in the figure, transformer–VAE achieves the best reconstruction performance across all feature dimensions, demonstrating superior modeling capability and higher fitting accuracy compared to both LSTM–VAE and the conventional VAE.
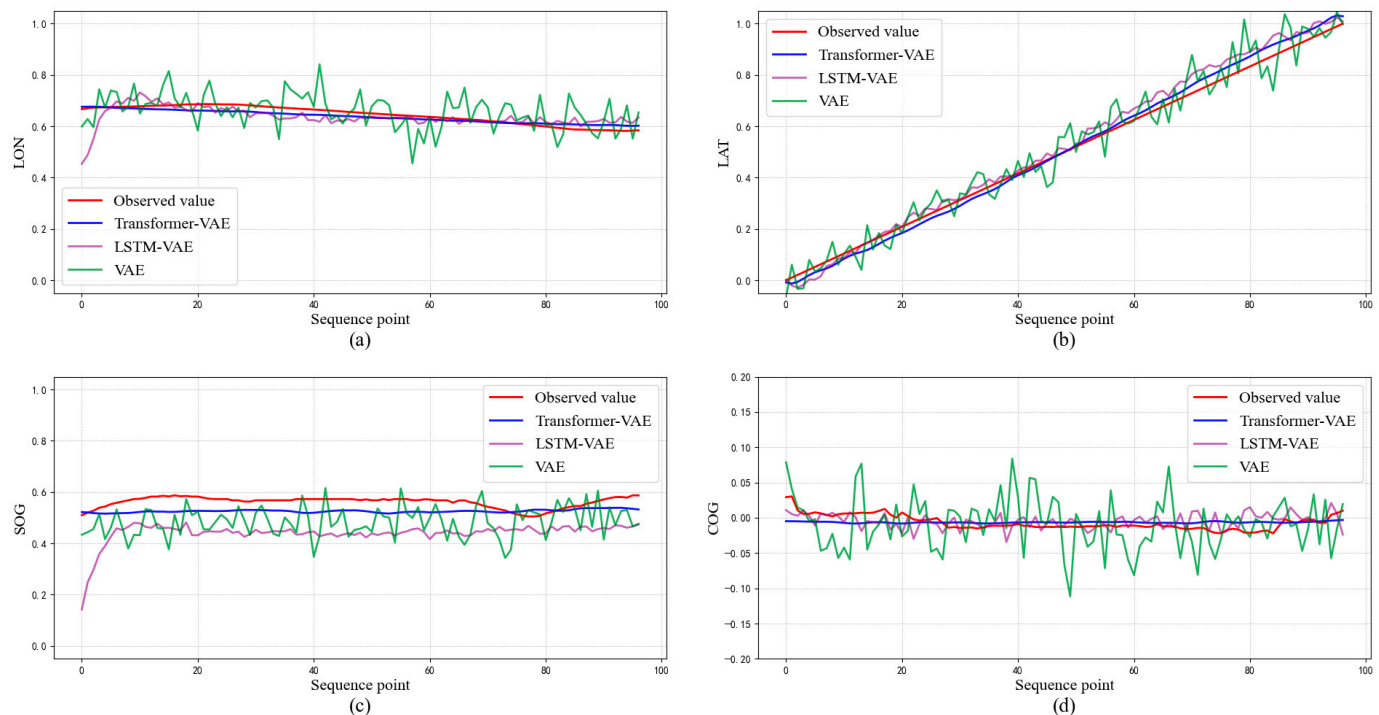
Figure 5. Comparison of different model reconstruction performances, where (**a**) LON, (**b**) LAT, (**c**) SOG, and (**d**) COG show the reconstructed results of each feature.

For the LON feature, the reconstruction curve generated by the transformer–VAE closely aligns with the observed values, accurately capturing both the overall trend and local variations, indicating strong spatial modeling capability. The reconstruction performance of LSTM–VAE is slightly inferior; although the overall trend is generally consistent with the observations, local delays and deviations can be observed. The conventional VAE performs worst on this feature, with its reconstruction curve exhibiting substantial fluctuation and significant deviation from the ground truth, suggesting a limited ability to fit spatial sequences. As for the LAT feature, since vessels navigating through the Jintang Bridge area primarily travel along the north–south direction, the latitude generally follows a monotonically increasing pattern. Consequently, the reconstruction differences among the three models are relatively small. Nevertheless, the transformer–VAE still demonstrates the best fitting performance, followed by LSTM–VAE, while the reconstruction curve of the conventional VAE still presents a certain degree of fluctuation.

For the SOG feature, transformer–VAE effectively captures the trend of speed variation, with reconstruction results that are smooth and closely aligned with the ground truth, demonstrating a strong capability in modeling dynamic features. In contrast, the reconstruction output produced by the LSTM–VAE is consistently lower than the original values, while the reconstruction curve of the conventional VAE exhibits substantial fluc-

tuations and fails to accurately reproduce the real speed variation process. Regarding the COG feature, the transformer–VAE maintains stable reconstruction performance and is capable of smoothly fitting the trend of heading variation. Although the reconstruction results of the LSTM–VAE generally follow the overall pattern of the ground truth, the curve exhibits noticeable fluctuations and fails to smoothly capture the periodic nature of heading changes, resulting in localized reconstruction instability. In contrast, the reconstruction produced by the conventional VAE shows significant oscillations and large reconstruction errors, making it incapable of accurately capturing the detailed variations in vessel heading.

To comprehensively evaluate the performance of the transformer–VAE, LSTM–VAE, and VAE in the trajectory reconstruction task, this study calculates the MSE, MAE, and RMSE of the three models on both the training and test datasets. The results are presented in Figure 6. On the training set, transformer–VAE achieves the lowest reconstruction errors across all trajectory feature dimensions, indicating its strong feature representation and trajectory pattern learning capabilities, which enable high-quality trajectory reconstruction. The test results further demonstrate that the transformer–VAE consistently maintains the lowest error levels across all metrics, reflecting its superior generalization performance. These findings further confirm that the transformer–VAE can effectively model complex spatiotemporal dynamics across different trajectory samples while maintaining a high level of reconstruction quality.

Based on the above trajectory reconstruction results, the transformer–VAE outperforms the baseline models in terms of reconstruction accuracy and feature fitting capability, exhibiting lower errors and higher consistency across trajectory features. This indicates that the proposed model can more accurately learn and reconstruct the spatiotemporal characteristics of normal vessel trajectories, providing a solid foundation for subsequent anomaly detection.
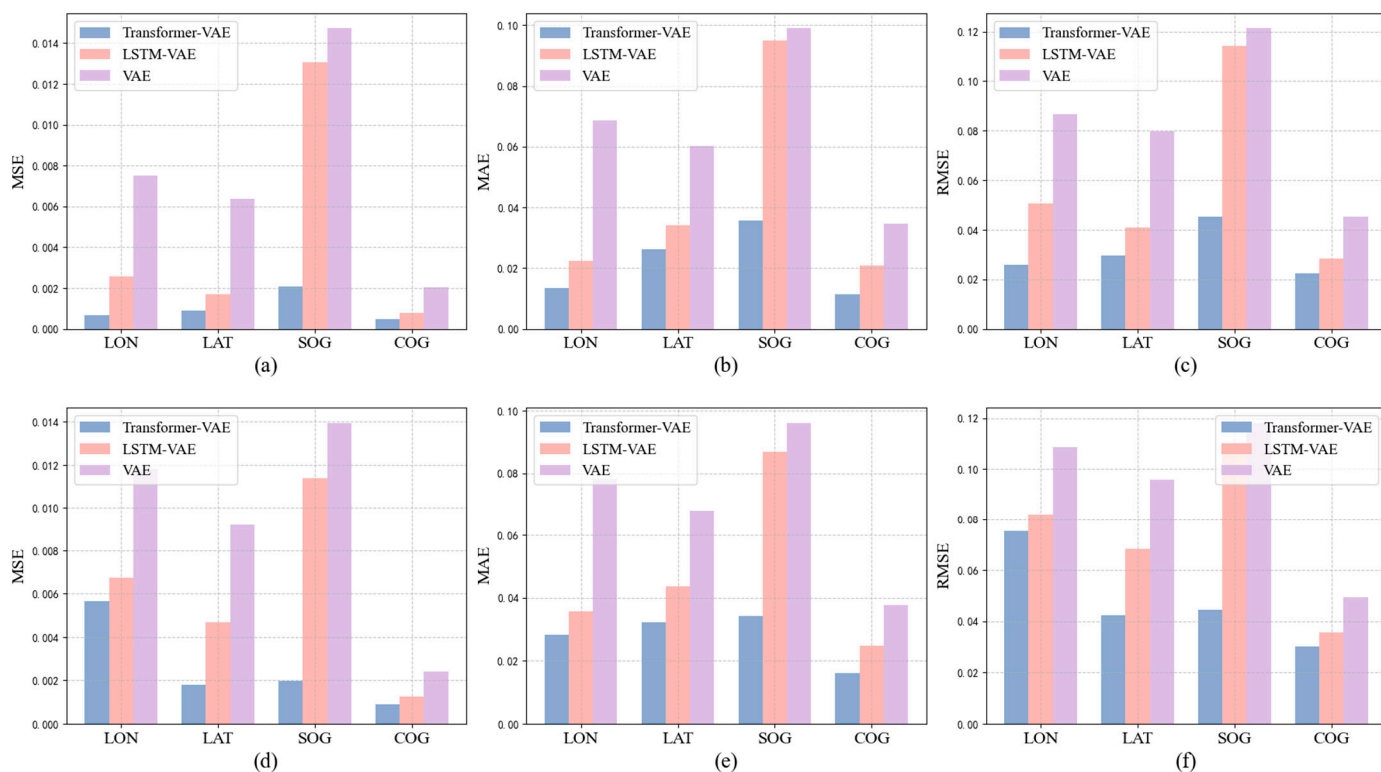
Figure 6. Comparison of reconstruction errors (MSE, MAE, RMSE) of different models. Subfigures (**a**–**c**) present the results on the training set, while (**d**–**f**) correspond to the test set.

In addition to reconstruction accuracy, computational efficiency is also important for evaluating model practicality. We compared the transformer–VAE and LSTM–VAE models under identical settings on a standard PC (Intel i5-7300HQ CPU, GTX 1050 GPU, 16 GB RAM) using Python 3.11. The transformer–VAE had 1,228,100 trainable parameters and required 3037 s to train for 200 epochs, while LSTM–VAE had 1,484,164 parameters and took 2820 s. Although the transformer–VAE has fewer parameters, this is due to its streamlined attention-based architecture compared to the gate-heavy structure of LSTM. The slightly higher training time is primarily attributed to the attention mechanism and the higher per-layer computational complexity of transformer blocks. Nonetheless, transformer–VAE remains suitable for large-scale offline anomaly detection. For real-time applications, lightweight optimization strategies can be considered.

4.3. Experimental Validation of Anomaly Detection

The transformer–VAE model distinguishes normal and abnormal trajectories by learning only the features of normal trajectories within the dataset. Specifically, the reconstruction error, or anomaly score, of a normal trajectory remains relatively low, as the model has learned its underlying patterns. Conversely, anomalous trajectories, which are not encountered during training, tend to exhibit higher reconstruction errors due to their deviations from learned patterns.

As shown in Figure 7, the distribution of reconstruction errors for the transformer–VAE model on both the training set and test set is illustrated. The distribution of training reconstruction errors exhibits a right-skewed pattern, with the majority of samples having reconstruction errors concentrated within the range of 0.0 to 0.01. The peak of the distribution is located in an extremely low-error region, approximately between 0.003 and 0.005. The sharp decline in error values beyond 0.01 indicates that the temporal features of most training samples have been stably modeled, with no significant deviations from the patterns learned by the model. This observation suggests that the model achieves a good fit on the training set, accurately learning the patterns of normal trajectories and enabling high-precision reconstruction. The distribution of test errors follows a similar overall trend to that of the training set but with notable differences. Although the test error distribution remains right-skewed, its range is broader, and some test samples exhibit significantly higher reconstruction errors compared to the training set. In particular, a noticeable long-tail distribution appears in the region above 0.02. Within the low-error range of 0.0 to 0.01, the test error distribution still exhibits a high peak, indicating that most test trajectories conform to the learned normal patterns and are reconstructed accurately. This confirms the transformer–VAE model's strong ability to generalize and reconstruct normal trajectories. However, beyond the 0.02 threshold, the tail of the test error distribution extends further than that of the training set, with the highest reconstruction error approaching 0.1. These high-error trajectories are likely to contain anomalous behavior, suggesting that the transformer–VAE model is capable of generating significantly higher reconstruction errors for abnormal trajectories under unsupervised conditions, thereby providing a reliable basis for anomaly detection.

In an unsupervised learning setting, the model has no prior exposure to abnormal trajectories; therefore, the core idea of anomaly detection is to identify trajectories with large reconstruction errors as anomalies based on the statistical distribution of reconstruction errors. To determine a reasonable threshold for anomaly detection, this study adopts a quantile-based approach to control the sensitivity of detection. Specifically, the 90th, 95th, and 98th percentiles are selected as threshold values for detecting anomalous trajectories within the test set. As shown in Figure 8, the number of detected anomalies decreases as the threshold increases. Under the 90th percentile threshold, the largest number of anoma-

lous trajectories is detected, though some normal trajectories may be falsely identified as anomalies. When using the 95th percentile threshold, fewer anomalies are detected, but most trajectories that deviate from normal patterns can still be captured. At the 98th percentile threshold, only trajectories with extremely large reconstruction errors are flagged as anomalous, reducing the false positive rate but potentially missing mildly abnormal trajectories. The results indicate that the choice of quantile threshold has a direct impact on anomaly detection outcomes. Lower thresholds (e.g., 90%) increase the recall rate but may result in more false positives, whereas higher thresholds (e.g., 98%) reduce false positives but may lead to missed detections. The 95th percentile threshold provides a relatively balanced trade-off between detection accuracy and robustness.
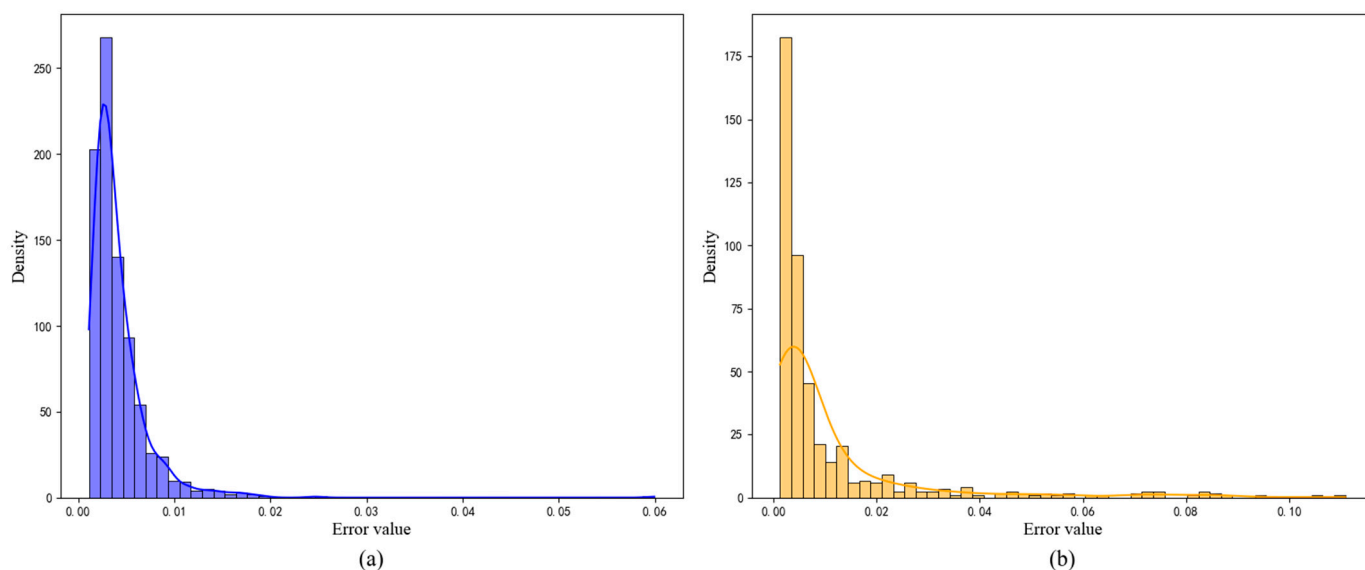
Figure 7. Reconstruction errors of the transformer–VAE model. (**a**) Training set; (**b**) Test set.
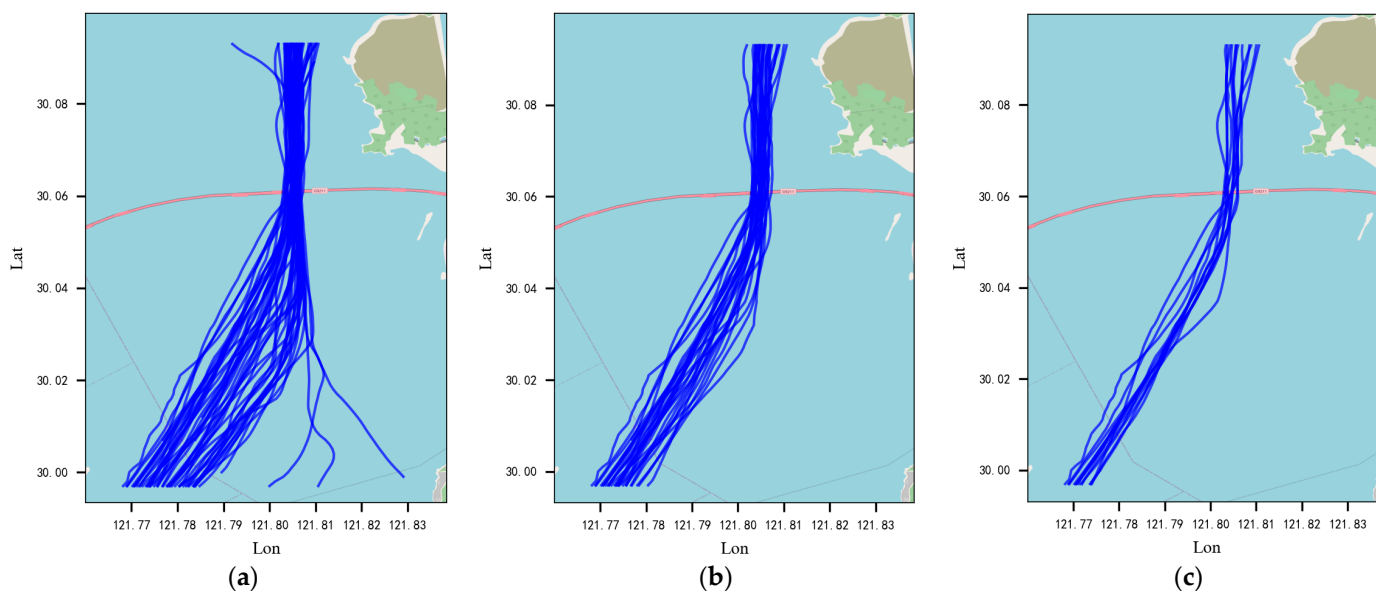
Figure 8. Anomaly detection results under different quantile thresholds. (**a**) Anomaly detection at 90% quantile threshold: 49 anomalous trajectories; (**b**) Anomaly detection at 95% quantile threshold: 27 anomalous trajectories; (**c**) Anomaly detection at 98% quantile threshold: 12 anomalous trajectories.

The choice of quantile threshold has a direct impact on the performance of anomaly detection. Lower quantile thresholds tend to increase the recall rate by detecting more

potentially anomalous trajectories, but they may also lead to higher false positive rates by misclassifying normal trajectories. In contrast, higher thresholds reduce false alarms but may miss moderately abnormal trajectories. To effectively evaluate the accuracy of anomaly detection under different quantile settings, it is essential to construct a reliable ground truth dataset of anomalous maritime trajectories.

In this study, representative anomalous vessel trajectories were identified and labeled by integrating multi-source data. Two primary sources were utilized:

(1)   Vessel traffic service (VTS) warning records, which capture real-time operational risk alerts flagged by maritime surveillance systems during navigation.

(2)   Administrative penalties and incident reports issued by maritime authorities, which provide authoritative documentation of historical violations or hazardous behaviors.

To further enhance the accuracy and representativeness of the labeled data, a manual verification process was conducted. A panel of three domain experts in maritime safety—including a certified captain, a VTS instructor, and a professor from a maritime academy—was invited to review the trajectories in the test set. The experts evaluated each trajectory in detail, taking into account real-world navigational contexts and vessel behavior characteristics, such as spatial deviation, abrupt speed changes, and anomalous course patterns. As a result, 33 anomalous trajectories were confirmed and used as ground truth labels to evaluate the performance of the proposed transformer–VAE model under different quantile thresholds. By comparing the detection results with expert-labeled ground truth, we further analyzed the model's performance in terms of false positive rate, false negative rate, and F1-score, thereby providing empirical guidance for threshold selection and detection accuracy optimization.

Figure 9 presents the confusion matrices of anomaly detection results under three quantile thresholds, while Table 2 summarizes the corresponding performance metrics, including accuracy, precision, recall, and F1-score. These metrics are calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F1\ score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recal} \tag{15}$$

where, *TP* (true positive) refers to cases where the model correctly predicts an anomalous trajectory, and *TN* (true negative) refers to cases where the model correctly predicts a normal trajectory. *FP* (false positive) occurs when the model incorrectly predicts an anomalous trajectory for a normal one, and *FN* (false negative) occurs when the model incorrectly predicts a normal trajectory for an actual anomalous one.

Table 2. Comparison of anomaly detection performance metrics under different quantile thresholds.

Quantile Threshold (%)	Accuracy	Precision	Recall	F1-Score
90	0.9586	0.5893	1.0000	0.7416
95	0.9874	0.9643	0.8182	0.8852
98	0.9622	1.0000	0.3636	0.5333
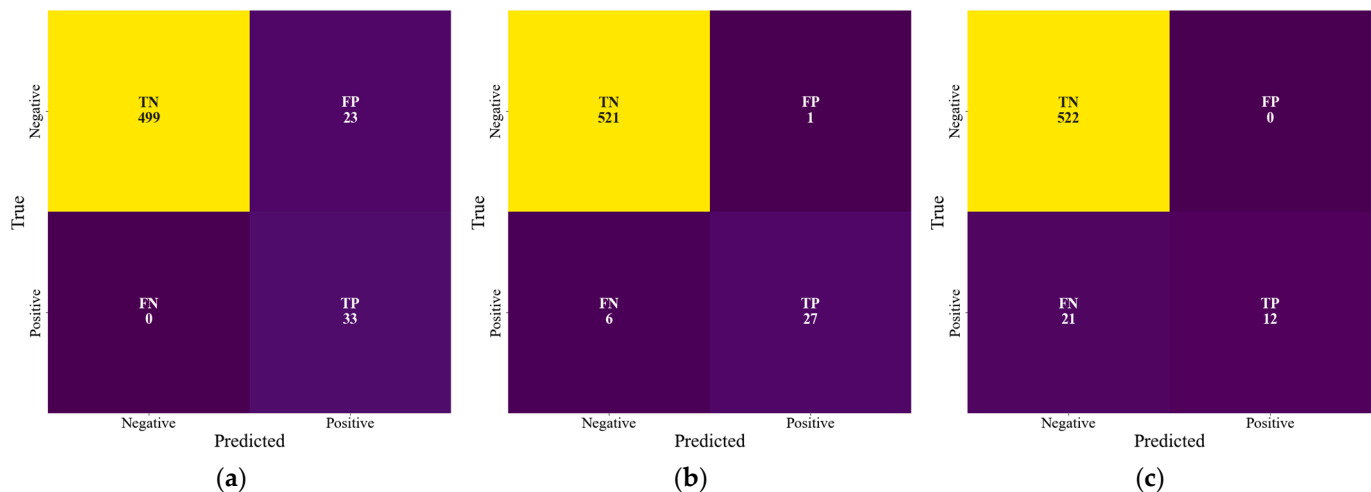
Figure 9. Confusion matrices of anomaly detection results at different quantile thresholds: (**a**) 90%, (**b**) 95%, and (**c**) 98%.

As illustrated in both the figure and the table, the choice of quantile threshold plays a critical role in shaping the model's detection performance. When using the 90% quantile threshold, the model achieved the highest recall (1.000), successfully identifying all 33 expert-labeled anomalous trajectories. However, due to the relatively low threshold, it also produced a high number of false positives (*FP* = 23), resulting in lower precision (0.5893) and overall accuracy (0.9586). This indicates that, under this setting, the model demonstrates strong sensitivity to anomalies but lacks sufficient discriminative ability for normal trajectories.

When the threshold is increased to 95%, the model shows a significant improvement in precision (0.9643) while maintaining a relatively high recall (0.8182), leading to the highest F1-score (0.8852). The number of false positives is reduced to only one, suggesting that this setting achieves an optimal balance between anomaly detection capability and false alarm control, with the highest overall accuracy (0.9874). However, when the threshold is further increased to 98%, the model becomes overly conservative. Although it attains perfect precision (1.0000), the recall drops sharply to 0.3636, and the F1-score declines to its lowest value (0.5333). The confusion matrix shows that most anomalous trajectories were misclassified as normal (*FN* = 21), indicating severe under-detection, which significantly limits the practical applicability of this threshold setting.

In summary, the 95% quantile threshold provides the best compromise between sensitivity and reliability, making it more suitable for real-world maritime traffic monitoring, where both accurate anomaly detection and low false alarm rates are essential for ensuring navigational safety and supporting effective traffic management.

To explore the feasibility of dynamic thresholding, a preliminary experiment based on extreme value theory was conducted as a methodological extension. Specifically, the POT method was applied. The initial threshold u was set at the 95th percentile of the reconstruction error distribution. All exceedances of $y = x - u$, where $x > u$, were extracted and modeled using a generalized Pareto distribution (GPD). The shape parameter (ξ) and scale parameter (σ) were estimated using maximum likelihood estimation (MLE). The final anomaly threshold was calculated by adding the target GPD quantile to u, resulting in a value of 0.047382, which closely matched the original 95th percentile threshold (0.047244). Both thresholds produced identical anomaly detection results on the test set. Although no additional anomalies were identified, this experiment validates the theoretical soundness of POT and highlights its potential applicability in dynamic maritime environments.

After completing the analysis of detection performance under different quantile thresholds, a representative normal trajectory and an anomalous trajectory were selected for detailed analysis, based on the fixed 95% quantile threshold.

Figure 10 presents the spatial visualization of the normal and anomalous trajectories identified by the transformer–VAE model. It is clearly observed from the figure that the vessel corresponding to the normal trajectory (blue line) traveled from south to north, strictly following the designated northbound navigational channel and smoothly passing through the bridge area. In contrast, the anomalous trajectory (red line) shows that the vessel did not comply with navigation regulations as it approached the bridge area, choosing instead to deviate significantly from the main navigation channel by taking a shortcut directly towards the bridge along its original heading. Such anomalous navigational behavior clearly violates the relevant regulations for bridge-area navigation and could potentially increase the risk of vessel-to-vessel collisions or severe vessel-bridge accidents.
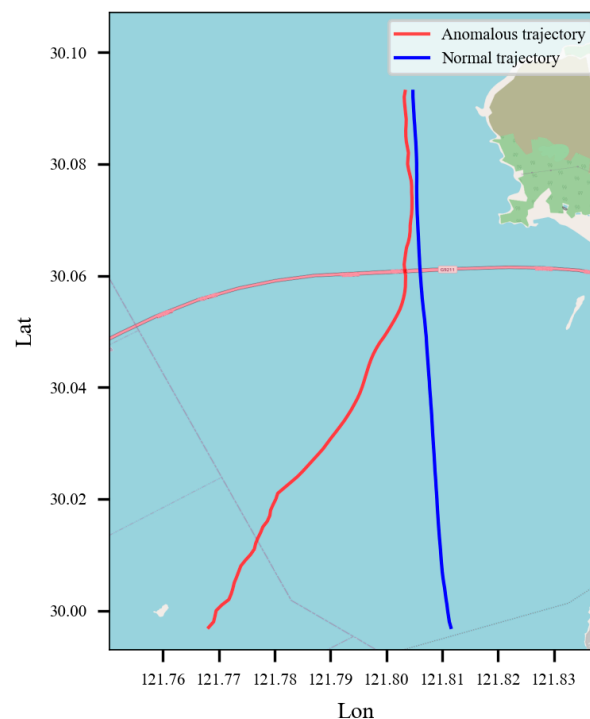
Figure 10. Comparison of the spatial distributions of normal and anomalous ship trajectories.

Figure 11 illustrates the reconstruction results of the normal trajectory features. As shown in Figure 11a,b, the original and reconstructed values of LON and LAT features of the normal trajectory exhibit high consistency, indicating that the vessel maintained a steady course along the designated navigational channel during actual sailing without any noticeable deviation in route or position. This trajectory behavior aligns well with the normal patterns learned by the model. Similarly, subplots of SOG and COG also show good reconstruction performance. Despite minor fluctuations observed locally, the reconstructed values closely match the original observed values. These results demonstrate that the transformer–VAE model, by learning from a substantial amount of normal trajectory data, successfully captures the inherent patterns of typical vessel movements, enabling accurate reconstruction of normal trajectories.
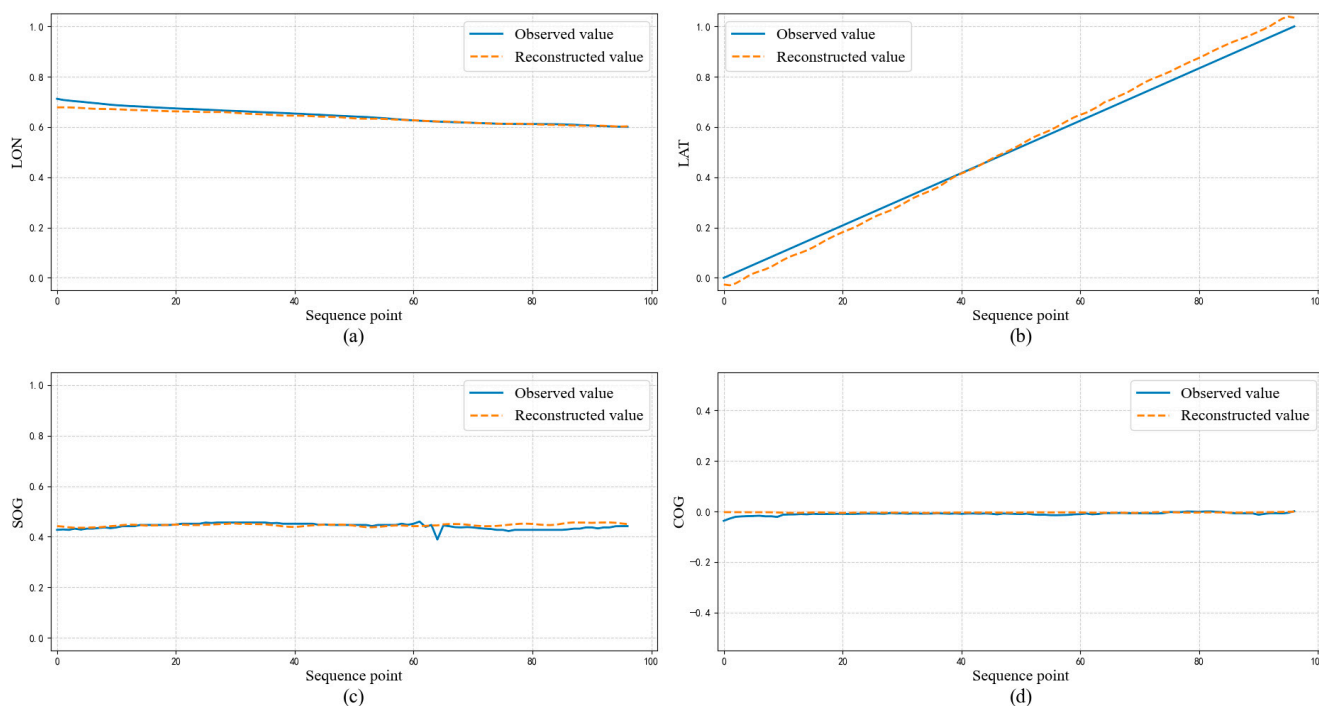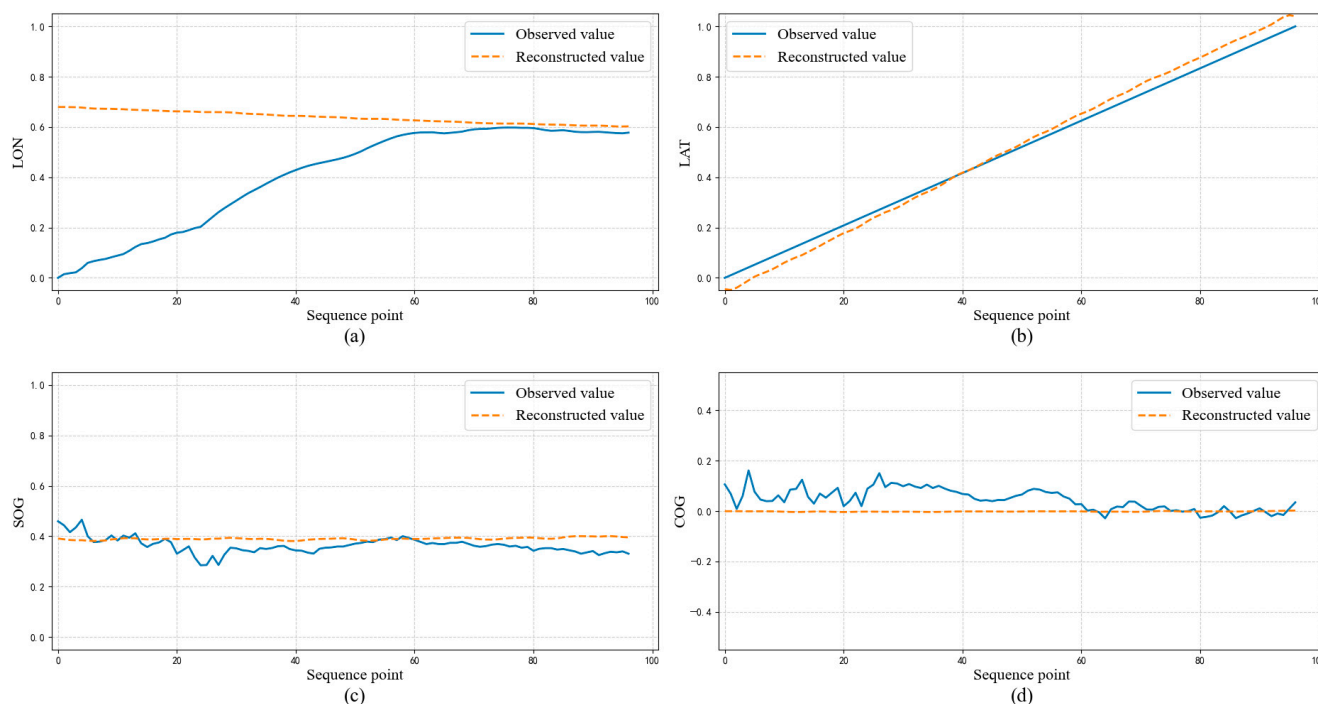
Figure 11. Identification of a typical normal trajectory. (**a**) LON, (**b**) LAT, (**c**) SOG, (**d**) COG.

A representative anomalous trajectory reconstruction result is shown in Figure 12. The original trajectory was identified as anomalous based on a quantile threshold, and the effectiveness of the transformer–VAE model can be verified visually. Specifically, the trajectory was flagged as anomalous due to significant reconstruction errors observed in three key features: LON, SOG, and COG. The spatial distribution of this trajectory significantly deviates from typical navigation patterns, particularly before the vessel enters the bridge area, where it significantly strays from the main navigational channel.

Prior to entering the bridge area, the actual longitude of the first 70 trajectory points (blue solid line) exhibits significant variation, clearly indicating that the vessel deviated from the standard navigational route. In contrast, since the transformer–VAE model has learned typical normal trajectory patterns, its reconstructed trajectory (orange dotted line) remains consistently aligned with the main navigational channel. As a result, the discrepancy between the actual trajectory and the model's reconstruction leads to a substantial increase in reconstruction error for the LON feature. Regarding LAT, the reconstruction performed well, primarily due to the anomalous trajectory's LAT remaining relatively stable and consistent with normal trajectories, enabling effective reconstruction based on learned normal patterns.

Regarding the SOG feature, vessels typically maintain stable speeds when navigating through cross-sea bridge areas under normal conditions. However, the trajectory under consideration reveals sharp fluctuations in speed characterized by brief bursts of acceleration and deceleration prior to reaching the bridge. Conversely, the transformer–VAE model, having learned stable speed patterns from mainstream trajectories, generates smoother reconstructed speed curves. This discrepancy leads to significant reconstruction errors in the SOG feature, further highlighting the anomalous nature of the actual trajectory. Similarly, for the COG feature, the anomalous trajectory showed frequent and irregular heading variations, whereas the model's reconstruction remained stable, reflecting learned normal heading patterns. This resulted in considerable reconstruction errors due to these irregular deviations.

Figure 12. Identification of a typical anomalous trajectory. (**a**) LON, (**b**) LAT, (**c**) SOG, (**d**) COG.

Overall, because the transformer–VAE model has effectively captured mainstream navigational patterns from normal vessel trajectories, its reconstructions consistently align with standard spatial (LON and LAT), speed, and heading behaviors. Thus, trajectories deviating from these typical patterns exhibit notable reconstruction errors, allowing the model to efficiently detect spatial deviations, abnormal speed fluctuations, and irregular heading changes. As an unsupervised anomaly detection method, the transformer–VAE demonstrates strong capability in identifying diverse types of anomalous trajectories.

4.4. Comparative Analysis of Anomaly Detection Performance Across Models

Building upon the previous section's evaluation of detection performance under different quantile thresholds and the in-depth analysis of representative normal and anomalous trajectories, this section further compares the proposed transformer–VAE model with two representative generative anomaly detection methods: LSTM–VAE and transformer–GAN. Both models adopt generative principles for anomaly detection and were evaluated on the same test set. Anomaly labels were based on the expert-reviewed ground truth established earlier, and the detection threshold for each model was uniformly set to the 95th percentile of its error distribution.

As shown in Table 3, the transformer–VAE outperformed both LSTM–VAE and transformer–GAN across all evaluation metrics, achieving an accuracy of 0.9874 and an F1-score of 0.8852. Compared to transformer–GAN, transformer–VAE demonstrated improvements in both detection sensitivity and false alarm control. While LSTM–VAE exhibited some detection capability, its relatively high miss rate indicated weaker performance. These results highlight the superior anomaly recognition ability and robustness of the proposed model.

Figure 13 presents the confusion matrices of the three models. The transformer–VAE achieved the most balanced results, with only one false positive and six false negatives. In contrast, transformer–GAN produced five false positives and ten false negatives, while the LSTM–VAE resulted in nine false positives and thirteen false negatives. These findings further confirm that the transformer–VAE exhibits stronger discriminative capability and

robustness in complex maritime environments. Overall, incorporating the transformer architecture into the variational autoencoding framework enables the model to more accurately learn the spatiotemporal characteristics of normal vessel trajectories and effectively identify anomalous deviations. This demonstrates its strong potential for unsupervised anomaly detection and real-world applicability in maritime traffic monitoring.

Table 3. Comparison of performance metrics for different models in anomaly detection.

Model	Accuracy	Precision	Recall	F1-Score
LSTM–VAE	0.9604	0.6897	0.6061	0.6452
Transformer–GAN	0.9766	0.8571	0.7273	0.7869
Transformer–VAE	0.9874	0.9643	0.8182	0.8852
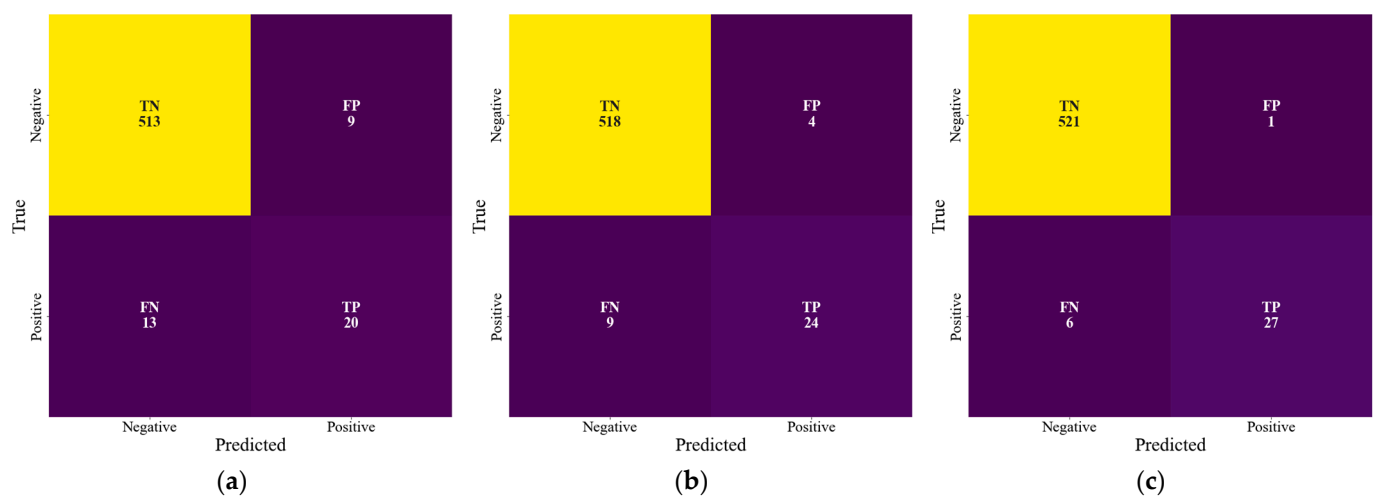
Figure 13. Confusion matrices of anomaly detection results for different models: (**a**) LSTM–VAE, (**b**) Transformer–GAN, and (**c**) Transformer–VAE.

5. Discussions

This study proposes an unsupervised vessel trajectory anomaly detection framework based on a transformer–VAE, validating its feasibility and effectiveness using AIS trajectory data collected from cross-sea bridge waters. The proposed framework aims to capture the underlying patterns of normal vessel trajectories and reconstruct vessel paths, subsequently identifying anomalies when trajectory reconstruction errors exceed predefined thresholds.

In trajectory reconstruction comparison experiments, the transformer–VAE demonstrates several distinct advantages over traditional VAE and LSTM–VAE models. First, the transformer–VAE exhibits a superior capability in modeling global temporal dependencies, primarily due to the incorporation of a multi-head self-attention mechanism in its encoder. Unlike recurrent neural networks, this mechanism allows simultaneous consideration of information across all time steps, significantly enhancing the model's capacity to capture long-term dependencies inherent in complex vessel movements. Second, the transformer–VAE achieves improved modeling of latent variables through the inclusion of temporal correlations explicitly in its latent space. Specifically, the integration of latent variables from preceding time steps facilitates richer representation of sequential trajectory patterns, thereby enhancing the accuracy and quality of trajectory reconstruction. Third, the transformer–VAE consistently yields lower reconstruction errors across various trajectory features, indicating its superior ability to learn normal vessel behaviors accurately, which in turn substantially enhances anomaly detection robustness. In addition, a sensitivity analysis on key architectural hyperparameters—such as latent dimension, number of attention

heads, and encoder layers—was performed. The results show that the transformer–VAE maintains stable performance across different configurations, confirming its structural robustness and practical implementation flexibility.

To enable practical anomaly detection without reliance on labeled data, this study employs reconstruction error quantiles of the 90th, 95th, and 98th percentiles as detection thresholds, thereby reflecting different sensitivity levels to potential anomalies. To systematically evaluate detection performance across these thresholds, a comprehensive ground truth dataset comprising anomalous trajectories was constructed from VTS alert records, official maritime incident/near miss reports, and expert validation. Detection results obtained under each quantile threshold were evaluated through standard performance metrics, including accuracy, precision, recall, and F1-score. The evaluation results indicate that the 95th percentile threshold achieves an optimal balance between detection accuracy and robustness, demonstrating practical applicability in real-world maritime settings. Building on this, considering that fixed quantile thresholds may not adapt well to varying maritime conditions, this study conducted a preliminary experiment using the POT method from extreme value theory to construct a dynamic threshold. Although the resulting threshold was numerically close to the 95th percentile and led to the same detection results in this case, the POT approach offers a theoretically sound and potentially adaptable framework. This preliminary test provides insight into future integration of dynamic thresholding mechanisms that incorporate real-time traffic patterns and environmental changes, further enhancing the model's adaptability and generalization in diverse maritime scenarios.

Further detailed analysis of trajectories classified as anomalous reveals two prominent behavioral patterns: (1) spatial deviations from the established main navigational channel, exemplified by vessels that unexpectedly detour or take shortcuts prior to entering the bridge area; and (2) abnormal variations in vessel speed and course, marked by substantial deviations from learned normal navigational patterns. The transformer–VAE model's strong spatiotemporal modeling capabilities effectively capture stable and representative patterns of typical trajectories. Consequently, when vessel behaviors deviate significantly from learned norms, the model generates notably higher reconstruction errors, facilitating accurate and timely anomaly detection.

To further assess the effectiveness of the proposed model, extended experiments were conducted to compare the anomaly detection performance of the transformer–VAE with two representative generative models, namely the LSTM–VAE and transformer–GAN, under consistent dataset and threshold settings. The results demonstrate that the transformer–VAE outperforms both baselines in terms of accuracy and F1-score, with improved detection sensitivity and false alarm control. The confusion matrix analysis further confirms the model's robustness and discriminative capability under dense maritime traffic scenarios. These findings further support the applicability and generalization capability of the proposed method in real-world maritime anomaly detection tasks.

Although the transformer–VAE performs well in both reconstruction and anomaly detection, its detection effectiveness remains influenced by the representativeness of normal trajectory patterns in the training data. Certain legitimate but uncommon navigational behaviors may be misclassified if absent from the training set. In this study, expert-guided screening and data cleaning were applied to build a high-quality training set, providing a stable foundation for learning. Nevertheless, some behavioral bias may still exist. Future research may incorporate anomaly-tolerant learning mechanisms or integrate automated cleaning and soft-labeling strategies to improve the model's robustness and practicality under more complex or partially contaminated data conditions.

Moreover, the applicability of the proposed approach is currently constrained by data limitations in two aspects. First, this study did not incorporate contextual environmental

features such as weather conditions, tides, or visibility, due to the unavailability of high-resolution environmental datasets that align temporally and spatially with AIS records. While this may affect detection accuracy in certain scenarios, the current work focuses on validating the structural effectiveness of the proposed model. Second, the current validation is limited to a representative cross-sea bridge navigation area characterized by narrow waterways, high traffic density, and tidal influences. These structural conditions are common in bridge-adjacent maritime zones, but empirical validation in other maritime domains—such as open-sea regions or port approaches—has not yet been conducted due to data unavailability. Future research may benefit from multi-modal integration and broader regional validation as more diverse datasets become accessible.

6. Conclusions

This study proposes a transformer–VAE designed specifically for unsupervised anomaly detection of ship trajectories in cross-sea bridge navigation scenarios. By leveraging a multi-head self-attention mechanism within the transformer architecture, the proposed model effectively captures complex, long-term spatiotemporal dependencies present in vessel trajectories. Consequently, it achieves high-accuracy reconstruction of normal vessel navigation patterns, significantly outperforming traditional VAE and LSTM–VAE models in terms of reconstruction quality and robustness. In addition, the model demonstrates strong structural robustness, as evidenced by its stable performance across different hyperparameter settings.

The identification of anomalies is conducted by evaluating reconstruction errors against quantile-based thresholds derived from the distribution of trajectories. Experimental analysis demonstrates that among the thresholds examined, the 95th percentile threshold provides an optimal balance between recall and precision. This threshold ensures effective detection of true anomalies while maintaining an acceptable rate of false alarms, thereby reflecting strong practical applicability. Preliminary experiments using the POT method further demonstrate the theoretical viability of dynamic thresholding, offering a promising direction for enhancing adaptability in complex maritime scenarios. Further examination of detected anomalous trajectories reveals two primary abnormal patterns: spatial deviations from designated navigation channels and atypical vessel dynamics, such as abrupt speed changes or unexpected course alterations. By effectively identifying these anomalies, the proposed model significantly contributes to enhanced intelligent monitoring and improved maritime traffic safety management. Comparative evaluations further confirm the superiority of the transformer–VAE over LSTM–VAE and transformer–GAN in terms of accuracy and detection balance.

Despite the model's demonstrated effectiveness, further efforts are needed to enhance its adaptability and generalization in practical deployments. Future research could prioritize the development of adaptive thresholding strategies that dynamically respond to real-time navigational contexts and variations in maritime conditions. Additionally, incorporating multimodal data sources, including AIS data, meteorological conditions, and tidal information, could enrich trajectory representations and enable a deeper understanding of complex maritime environments. Such integration would further enhance the generalization capabilities and explanatory power of the proposed anomaly detection framework.

Author Contributions: Conceptualization, J.H., H.Z. and P.Z.; methodology, J.H., H.Z., J.L., Y.Y. and P.Z.; software, J.H. and J.L.; validation, Y.Z., M.G. and Y.Z.; formal analysis, J.H., Y.Y. and P.Z.; investigation, J.H., H.Z., J.L. and Y.Y.; resources, P.Z.; data curation, J.H., H.Z. and J.L.; writing—original draft preparation, J.H. and H.Z.; writing—review and editing, M.G., Y.Z., Y.Y. and P.Z.; visualization, J.H.; supervision, P.Z.; project administration, H.Z. and P.Z.; funding acquisition, P.Z. All authors have read and agreed to the published version of the manuscript.

References

1. Zerveas, G.; Jayaraman, S.; Patel, D.; Bhamidipaty, A.; Eickhoff, C. A transformer-based framework for multivariate time series representation learning. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Singapore, 14–18 August 2021; ACM Press: New York, NY, USA, 2021; pp. 2114–2124.
2. He, Y.; Li, Z.; Mou, J.; Hu, W.; Li, L.; Wang, B. Collision-avoidance path planning for multi-ship encounters considering ship manoeuvrability and COLREGs. *Transp. Saf. Environ.* **2021**, *3*, 103–113. [CrossRef]
3. Sidibé, A.; Gao, S. Study of automatic anomalous behavior detection techniques for maritime vessels. *J. Navig.* **2017**, *70*, 847–858. [CrossRef]
4. Gao, X.; Dai, W.; Yu, L.; Yu, Q. Study on factors contributing to severity of ship collision accidents in the Yangtze River estuary. *Transp. Saf. Environ.* **2024**, *6*, tdae014. [CrossRef]
5. Mascaro, S.; Nicholso, A.E.; Korb, K.B. Anomaly detection in vessel tracks using Bayesian networks. *Int. J. Approx. Reason.* **2014**, *55*, 84–98. [CrossRef]
6. Xie, L.; Guo, T.; Chang, J.; Wan, C.; Hu, X.; Yang, Y.; Ou, C. A Novel Model for Ship Trajectory Anomaly Detection Based on Gaussian Mixture Variational Autoencoder. *IEEE Trans. Veh. Technol.* **2023**, *72*, 3456–3468. [CrossRef]
7. Smith, M.; Reece, S.; Roberts, S.; Rezek, I. Online maritime abnormality detection using gaussian processes and extreme value theory. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; pp. 645–654.
8. Le Guillarme, N.; Lerouvreur, X. Unsupervised extraction of knowledge from S-AIS data for maritime situational awareness. In Proceedings of the 16th International Conference on Information Fusion, Istanbul, Turkey, 9–12 July 2013; pp. 2025–2032.
9. Rong, H.; Teixeira, A.P.; Guedes Soares, C. Ship trajectory uncertainty prediction based on a Gaussian Process model. *Ocean Eng.* **2019**, *182*, 499–511. [CrossRef]
10. Rong, H.; Teixeira, A.P.; Guedes Soares, C. Data mining approach to shipping route characterization and anomaly detection based on AIS data. *Ocean Eng.* **2020**, *198*, 106936. [CrossRef]
11. Anneken, M.; Fischer, Y.; Beyerer, J. Evaluation and comparison of anomaly detection algorithms in annotated datasets from the maritime domain. In Proceedings of the Sai Intelligent Systems Conference, London, UK, 10–11 November 2015; IEEE: New York, NY, USA, 2015.
12. Nguyen, D.; Simonin, M.; Hajduch, G.; Vadaine, R.; Tedeschi, C.; Fablet, R. Detection of Abnormal Vessel Behaviors from AIS data using GeoTrackNet: From the Laboratory to the Ocean. In Proceedings of the 2020 21st IEEE International Conference on Mobile Data Management (MDM), Versailles, France, 30 June–3 July 2020; pp. 264–268.
13. Liu, W.; Yuan, W.; Chen, X.; Lu, Y. An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system. *Ocean Eng.* **2021**, *235*, 109435. [CrossRef]
14. Fu, P.; Wang, H.; Liu, K.; Hu, X.; Zhang, H. Finding abnormal vessel trajectories using feature learning. *IEEE Access* **2017**, *5*, 7898–7909. [CrossRef]
15. Djenouri, Y.; Belhadi, A.; Djenouri, D.; Srivastava, G.; Lin, J.C.-W. Intelligent deep fusion network for anomaly identification in maritime transportation systems. *IEEE Trans. Intell. Transport. Syst.* **2022**, *24*, 2392–2400. [CrossRef]
16. Lei, P.R. A framework for anomaly detection in maritime trajectory behavior. *Knowl. Inf. Syst.* **2016**, *47*, 189–214. [CrossRef]
17. Liu, B.; de Souza, E.N.; Matwin, S.; Sydow, M. Knowledge-based clustering of ship trajectories using density-based approach. In Proceedings of the IEEE International Conference on Big Data, Washington, DC, USA, 27–30 October 2014; IEEE: New York, NY, USA, 2014.
18. Nooshin, H.; Hamid, S. A fast DBSCAN algorithm for big data based on efficient density calculation. *Expert Syst. Appl.* **2022**, *203*, 117501.

19.   Li, K.; Guo, J.; Li, R.; Wang, Y.; Li, Z.; Miu, K.; Chen, H. The abnormal detection method of ship trajectory with adaptive transformer model based on migration learning. In Proceedings of the 2023 4th International Conference on Spatial Data and Intelligence, Nanchang, China, 13–15 April 2023; pp. 204–220.

20.   Kontopoulos, I.; Varlamis, I.; Tserpes, K. A distributed framework for extracting maritime traffic patterns. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 767–792. [CrossRef]

21.   Yang, J.; Liu, Y.; Ma, L. Maritime traffic flow clustering analysis by density-based trajectory clustering with noise. *Ocean Eng.* **2022**, *249*, 111001. [CrossRef]

22.   Chen, J.; Chen, H.; Chen, Q. Vessel sailing route extraction and analysis from satellite-based AIS data using density clustering and probability algorithms. *Ocean Eng.* **2023**, *280*, 114627. [CrossRef]

23.   Bai, X.; Xie, Z.; Xu, X.; Xiao, Y. An adaptive threshold fast DBSCAN algorithm with preserved trajectory feature points for vessel trajectory clustering. *Ocean Eng.* **2023**, *280*, 114930. [CrossRef]

24.   Singh, S.K.; Fowdur, J.S.; Gawlikowski, J.; Medina, D. Leveraging graph and deep learning uncertainties to detect anomalous maritime trajectories. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 23488–23502. [CrossRef]

25.   Zhao, L.; Shi, G. Maritime anomaly detection using density-based clustering and recurrent neural network. *J. Navig.* **2019**, *72*, 894–916. [CrossRef]

26.   Mantecón, T.; Casals, D.; Navarro-Corcuera, J.J.; del Blanco, C.R.; Jaureguizar, F. Deep learning to enhance maritime situation awareness. In Proceedings of the 2019 20th International Radar Symposium, Ulm, Germany, 26–28 June 2019; IEEE: New York, NY, USA, 2019; pp. 1–8.

27.   Nguyen, D.; Vadaine, R.; Hajduch, G. A multi-task deep learning architecture for maritime surveillance using ais data streams. In Proceedings of the 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, Turin, Italy, 1–3 October 2018; IEEE: New York, NY, USA; pp. 331–340.

28.   Karataş, G.B.; Karagoz, P.; Ayran, O. Trajectory pattern extraction and anomaly detection for maritime vessels. *Inter. Things* **2021**, *16*, 100436. [CrossRef]

29.   Rhodes, B.J.; Bomberger, N.A.; Zandipour, M. Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness. In Proceedings of the International Conference on Information Fusion, Quebec, QC, Canada, 9–12 July 2007; IEEE: New York, NY, USA, 2007.

30.   Xie, L.; Pi, D.; Zhang, X.; Chen, J.; Luo, Y.; Yu, W. Graph neural network approach for anomaly detection. *Measurement* **2021**, *180*, 109546. [CrossRef]

31.   Liang, Y.; Zhang, H. Ship track prediction based on AIS data and PSO optimized LSTM network. *Int. Core J. Eng.* **2020**, *6*, 23–33.

32.   Tang, H.; Yin, Y.; Shen, H. A model for vessel trajectory prediction based on long short-term memory neural network. *J. Mar. Eng. Technol.* **2022**, *21*, 136–145. [CrossRef]

33.   Maganaris, C.; Protopapadakis, E.; Doulamis, N. Outlier detection in maritime environments using AIS data and deep recurrent architectures. In Proceedings of the PETRA'24: Proceedings of the 17th International Conference on Pervasive Technologies Related to Assistive Environments, Crete, Greece, 26–28 June 2024; Association for Computing Machinery: New York, NY, USA, 2024; pp. 420–427.

34.   Sadeghi, Z.; Matwin, S. A computationally inexpensive method for anomaly detection in maritime trajectories from AIS dataset. In *Advances in Information and Communication*; Arai, K., Ed.; Springer Nature: Cham, Switzerland, 2024; pp. 304–317.

35.   Liang, M.; Weng, L.; Gao, R. Unsupervised maritime anomaly detection for intelligent situational awareness using AIS data. *Knowl.-Based Syst.* **2024**, *284*, 111313. [CrossRef]

36.   Chen, X.; Wu, H.; Han, B.; Liu, W.; Montewka, J.; Liu, R.W. Orientation-aware ship detection via a rotation feature decoupling supported deep learning approach. *Eng. Appl. Artif. Intell.* **2023**, *125*, 106686. [CrossRef]

37.   Chen, X.; Hu, R.; Luo, K.; Wu, H.; Biancardo, S.A.; Zheng, Y.; Xian, J. Intelligent ship route planning via an A* search model enhanced double-deep Q-network. *Ocean Eng.* **2025**, *327*, 120956. [CrossRef]