Scilight

Applied Mathematics and Statistics
https://www.sciltp.com/journals/ams

Article

Modeling Spatiotemporal Interactions between Ships for the Trajectory Prediction in Encounter Scenarios

Hao Yuan *, Fei Li, Xuan Jiang, Daiwei Yu and Mengjia Shen

The Ninth Research Laboratory, Wuhan Second Ship Design and Research Institute, Wuhan 430205, China
* Correspondence: 285967@whut.edu.cn

How To Cite: Yuan, H.; Li, F.; Jiang, X.; et al. Modeling Spatiotemporal Interactions between Ships for the Trajectory Prediction in Encounter Scenarios. *Applied Mathematics and Statistics* **2025**, *2*(2), 8. https://doi.org/10.53941/ams.2025.100008

Received: 25 November 2025
Revised: 16 December 2025
Accepted: 24 December 2025
Published: 29 December 2025

Abstract: Ship trajectory prediction is crucial for collision avoidance and maritime traffic management. In an encounter scenario, a ship must predict the future states of other ships to take effective collision avoidance actions. However, conventional models often fail to accurately capture the interactions between ships in converging waters, resulting in poor trajectory prediction. This paper proposes a new model, STETC, which combines a Transformer and a convolutional neural network (CNN) into a two-layer encoder–decoder structure. The motion path encoder employs a self-attention mechanism to extract motion features from historical trajectories. The encounter interaction encoder integrates a CNN and self-attention mechanisms to extract interaction features from an artificial potential field generated by ship dynamic parameters. The two decoders then use cross-attention to progressively establish the spatiotemporal relationships between motion and interaction features. Through training, the model learns the interaction patterns between ships and the dynamic development of encounter situations. Real ship trajectory data is used to validate the effectiveness of STETC. Comparison experiments with observation times of 5, 10, and 15 min demonstrate that the STETC model outperformed other models. The case study validates that the STETC model can precisely perceive the motion states of surrounding ships and, leveraging encounter scenarios, generates predicted trajectories that reflect the interactive behaviors of the ships more accurately.

Keywords: ship trajectory prediction; transformer; convolutional neural network; collision avoidance; spatiotemporal relationships

1. Introduction

The rapid expansion of the shipping industry in recent years has resulted in the progressive congestion of maritime trade routes [1]. This escalation in traffic has introduced additional challenges to the safety of maritime transport. Notably, ship collisions account for the majority of maritime accidents, with the most of them attributed to human error [2]. The primary causes include inadequate vigilance, erroneous risk assessments of potential collisions, and insufficient coordination in collision avoidance. In this context, trajectory prediction is critical as it offers foresight into future traffic scenarios. This enables maritime personnel to anticipate potential hazards and undertake preemptive measures for averting collisions, thereby mitigating the risk of conflicts.

Trajectory prediction tasks for vehicles and pedestrians have been studied extensively. Gao [3] used a graph neural network called VectorNet to fuse the relationships between high-precision map information and vehicle trajectories and achieved good prediction accuracy. Amir [4] extracted features from the physical environment using attention mechanisms and adversarial training, and Hao [5] proposed a hierarchical long short-term memory (LSTM) network incorporating social neighborhoods and scene layouts to predict pedestrian trajectories for foot traffic. Kosaraju [6] introduced Social-BiGAT, a recurrent generative adversarial network that uses graph attention network principles to predict multimodal trajectories realistically, thereby capturing pedestrian social interactions

within a scene. However, the motion characteristics and constraints of objects vary in different environments. For example, vehicles generally travel along lane lines and act according to signaling changes, whereas human trajectories must consider diverse influences such as individual and social interactions. Notably, vehicles and pedestrians can change their motion states within a short time interval. In contrast, ships cannot execute steering and deceleration operations quickly owing to large inertia, a certain safety distance needs to be maintained between ships in maritime navigation, and their motion patterns differ from those of vehicles and pedestrians. Therefore, trajectory prediction techniques developed for vehicles and pedestrians may not be directly applicable to predicting ship trajectories at sea. Moreover, ships navigate in more open spaces and are subject to fewer environmental restrictions. Accordingly, considering the distinct motion patterns of ships and their social interaction dynamics is essential. Although the International Regulations for Collision Avoidance at Sea (COLREGs) provide collision avoidance rules for various encounters, in practice, crew members often deviate from these rules according to the specific circumstances of an encounter, which increases the uncertainty of future trajectories. Therefore, historical trajectory data hold significant value, as they provide critical insights into ship behavior patterns under complex spatiotemporal conditions and reveal underlying interaction rules and navigation practices among ships. Fully utilizing and exploring historical data are crucial for improving the accuracy of trajectory prediction in encounter scenarios.

In order to address these challenges, we propose a trajectory prediction model that accounts for spatiotemporal interactions between ships. This model integrates a Transformer and a convolutional neural network (CNN) components, forming a complex two-layer encoder-decoder structure. This structure approach enables the model to capture both historical motion features and social interaction features across all timestamps, and generate socially aware future trajectories. Our main innovations and contributions are detailed as follows:

1. We present the novel STETC ship trajectory prediction model, which uses two encoders to capture ship motions and social features across different levels, from low-level behavior patterns to high-level contextual associations. The two decoders sequentially establish the spatiotemporal correlation between these features and future trajectories, thereby enabling the predicted trajectories to account for both basic motion details and comprehensive encounter interactions.
2. To capture the spatiotemporal interactions during ship encounters, an artificial potential field derived from the ship's dynamic parameters is employed to represent the interaction constraints between ships. In the encounter interaction encoder, a spatiotemporal attention comprising CNNs and self-attention mechanisms processes the continuous artificial potential field, effectively capturing spatial interaction features across different temporal scales during ship encounters. By integrating the long-term navigational behavior features previously extracted from the motion path encoder using the temporal self-attention module, the model is endowed with the ability to reason about the development of encounter scenarios.
3. Utilizing Automatic Identification System (AIS) data from the Zhoushan sea area, the proposed method was benchmarked against Transformer, Seq2Seq, and Social GAN models. Error analysis indicators, such as average displacement error (ADE) and final displacement error (FDE), indicate that the proposed method surpasses the other models in prediction accuracy. Furthermore, attention visualization and ablation studies substantiate the rationality of the model design.

2. Related Work

2.1. Machine Learning Models

Guo [7] proposed a ship trajectory prediction algorithm based on K-order multivariate Markov chains, which utilizes multiple navigation-related parameters to construct a state transfer matrix for trajectory prediction. This model addresses the uncertainty in ship trajectory prediction. Zhao [8] sought to enhance the prediction accuracy within constrained waterways. They utilized real-time estimation of system noise within the Kalman filter algorithm to predict ship trajectories in scenarios with limited ship AIS information. Liu [9] proposed a ship trajectory prediction model using support vector regression. They employed the adaptive chaos differential evolution algorithm to optimize the internal parameters, thereby enhancing convergence speed and prediction accuracy. Zhang [10] developed a maneuvering model for trajectory prediction based on aerodynamic parameters. They utilized Markov chain Monte Carlo simulations and Bayesian decision theory for development. This method can provide improved prediction accuracy compared to traditional extrapolation theory. However, it requires more comprehensive maneuvering patterns to be acquired, and this is often challenging in practical applications. Virjonen [11] employed the K-nearest neighbors (KNN) search method to predict ships' sailing times and positional coordinates. The KNN method involves comparing differences in sailing paths and speeds between new

and historical ships, identifying historical ships with behaviors similar to those of new ships, and utilizing the sailing times and positional coordinates of the most similar historical ships to predict the trajectories of new ships.

2.2. Deep Learning Models

With the widespread adoption of AIS and advancements in artificial intelligence technology [12,13], AIS data have become indispensable tools for water traffic control and regulation [14]. The AIS provides information such as the ship's unique identification number, name, latitude and longitude, speed, ground heading, and course information, along with extensive historical spatiotemporal data on ships, thus aiding researchers in exploring ship motion characteristics.

In recent years, several studies have utilized LSTM networks and corresponding deformers for ship trajectory prediction. Ding [15] predicted trajectory sequence characteristics, including longitude, latitude, speed, and heading of a ship for the next 5 to 20 min using a variational LSTM. Liu [16] addressed issues related to sequence coding distortion and spatiotemporal data discontinuity, proposing an optimized method for navigation dynamic prediction using an attention-enhanced LSTM network. Simulation analysis confirmed the accuracy and robustness of their model, which achieved high-accuracy predictions for ship latitude, longitude, heading, and speed. Chen [17] proposed a novel ship trajectory prediction framework using a bidirectional LSTM model, capable of feeding predicted trajectories back into the model to optimize trajectories when a ship's motion state changes significantly. Wang [18] introduced a ship trajectory prediction model, CNN-LSTM-SE, which integrates a CNN module, an LSTM module, and a squeeze-and-excitation (SE) module. The CNN module extracts relational data among variables such as longitude, latitude, speed, and ground course. The LSTM module captures temporal correlations, while the SE module adaptively adjusts the importance of channel features, prioritizing those significantly impacting future trajectories. This model accounts for potential correlations between variables and temporal features to achieve more accurate predictions. Bao [19] combined a multi-attention mechanism with gated recurrent units to develop a ship trajectory prediction model analogous to an LSTM network. Although LSTM networks have achieved satisfactory results in long-term trajectory prediction, they typically struggle to model complex temporal dependencies [20].

The Transformer, a deep learning model based on attention mechanisms, can capture long-distance data dependencies and inter-data interactions, tapping into internal data connections crucial for multimodal trajectory prediction tasks. Jiang [21] merged the LSTM structure with the Transformer algorithm, employing an LSTM module to capture temporal features of spatiotemporal data and integrating a self-attention mechanism from the Transformer to address LSTM limitations in capturing long-range sequence information. This method fully utilizes the long-range dependencies of spatiotemporal features and leverages the complementary advantages of a fused model. Zou [22] utilized a Transformer to fuse a dynamic scene composed of sensory LIDAR, camera, and combined inertial navigation system data for predicting the future trajectory of a target vehicle. Experimental results demonstrated that their Transformer model effectively fuses multimodal sensor information. Yang [23] introduced the long and short-term spatiotemporal aggregation (LSSTA) network, incorporating a Transformer network to handle long-term temporal dependencies and temporal convolutional networks to aggregate spatial and temporal features, successfully enabling multimodal prediction of pedestrian trajectories. To the best of our knowledge, a Transformer has not been used to fuse information from different types to extract the spatiotemporal relationships of ships in the ship trajectory prediction field. We propose a method combining a Transformer and CNN to process historical trajectory sequences and APF matrices, accurately predicting ship trajectories in encounter scenarios. The proposed method is expected to aid in route planning and ship collision avoidance.

3. Encounter Trajectory Prediction Model

In this study, AIS data were meticulously preprocessed before being input into the STETC trajectory prediction model to generate accurate predicted trajectories. As illustrated in Figure 1, anomalous data were removed from the raw AIS dataset to enhance model performance. In accordance with the COLREGs, pairs of ships involved in encounters were identified, and their trajectory data were extracted. Subsequently, the data were normalized to enhance the training efficacy and accelerate convergence. The proposed STETC model integrates Transformer and CNN features, efficiently extracting advanced motion features from historical data. It improves prediction accuracy by considering the influence of the sailing states of other ships in encounter scenarios.

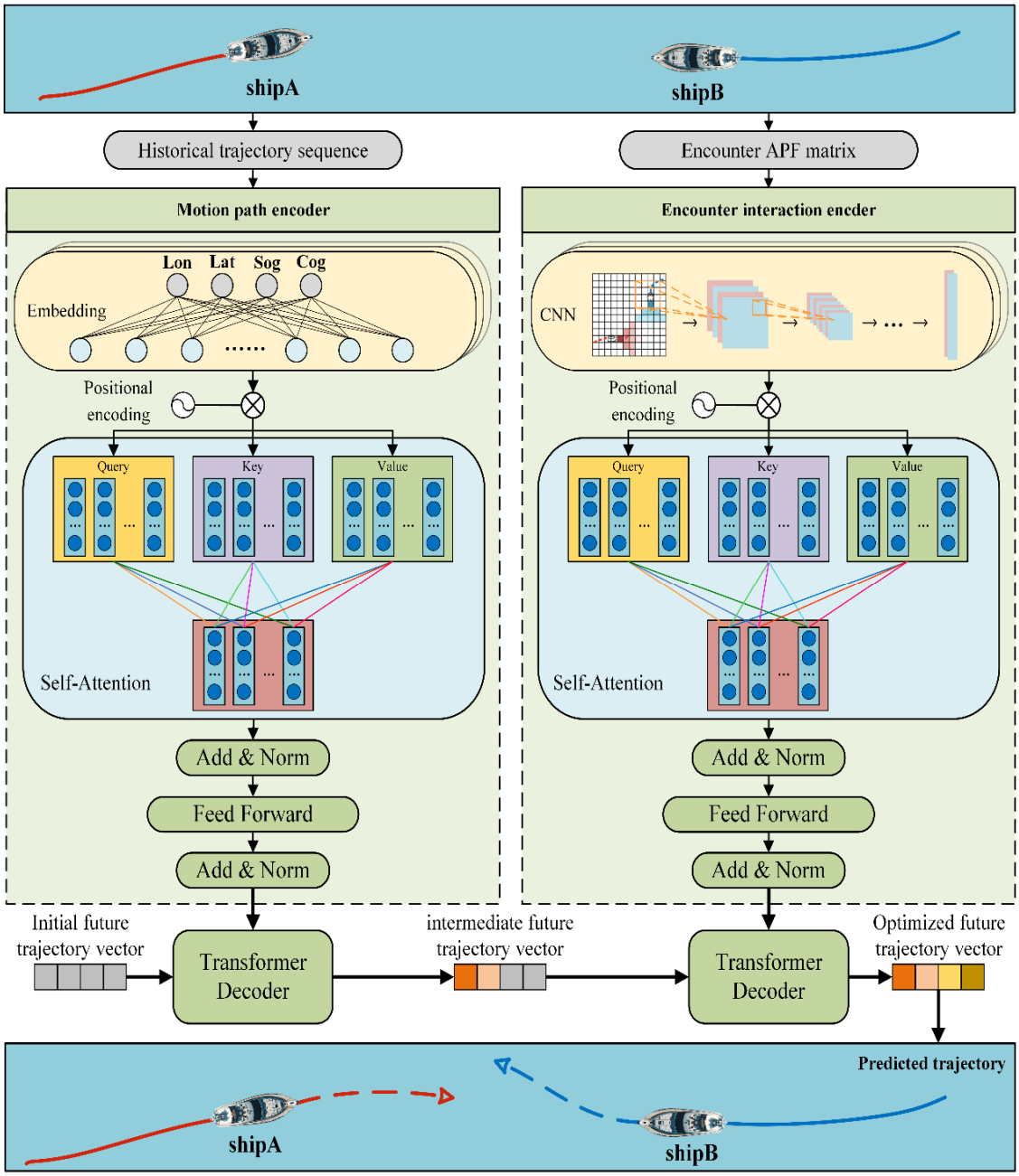

Figure 1. Overview of STETC.

The STETC model comprises two layers of encoder–decoder structures. It extracts trajectory information from the AIS data of encountering ships. This information is input into the motion trajectory encoder to extract the historical motion features of the two ships. The encounter potential field of the two ships is calculated using AIS data and subsequently input into the encounter interaction encoder to extract spatiotemporal relationship features. The features extracted by the two encoders are progressively fused through the two decoders. Each pass through a decoder updates the future trajectory vector. Ultimately, the future trajectories of the ships in the encounter scenario are inferred.

3.1. Problem Formulation

Assuming two ships in an encounter, the historical trajectories of the two ships near where the encounter occurred $X = \{X_1, X_2\}$ are considered as inputs and the future trajectories of the two ships $\hat{Y} = \{\hat{Y}_1, \hat{Y}_2\}$ are predicted. The historical trajectories of ships during the observation period are denoted as $X_i = \{Lat_i^t, Lon_i^t, Sog_i^t, Cog_i^t\}$, where $i \in \{1,2\}$ and $t \in \{1,\ldots,t_{obs}\}$, and the ground-truth trajectories for

a future period are denoted as $Y_t = \left\{ Lat_i^t, Lon_i^t \right\}$, where $t \in \left\{ 1, \ldots, t_{pred} \right\}$. The predicted trajectory of a ship is defined as $\hat{Y}_t = \left\{ \mathrm{Lat}_\mathrm{i}^\mathrm{t}, Lon_i^t \right\}$, where $t \in \left\{ 1, \ldots, t_{pred} \right\}$. Here, Lat is the latitude, Lon is the longitude, Sog is the ship-to-ground speed, and Cog is the ship-to-ground heading. Our goal is to predict the future trajectories of an approaching ship pair accurately.

3.2. Trajectory Processing for Encounters

Latitude, longitude, speed, and heading data were extracted from the AIS data to meet our research requirements. However, owing to multiple factors, including equipment malfunctions and the intricacies of maritime operational environments, AIS data often exhibit contamination by anomalies and instances of missing values during the acquisition, packaging, transmission, reception, and decoding stages. Given the challenges of using such data directly for model training and analysis, AIS data must be detected and repaired to meet the subsequent operational requirements. An extracted raw AIS dataset is defined as follows:

$$Traj_m = \left\{ Lat_t, Lon_t, Sog_t, Cog_t \right\} \tag{1}$$

where $Traj$ denotes the single-trajectory dataset, m is the index of the trajectory dataset, Lat is the latitude, Lon is the longitude, Sog is the ship-to-ground speed, Cog is the ship-to-ground heading, and t is the time step.

To clean the anomalous trajectories in AIS data, trajectory data that are too weak or non-navigational, AIS trajectories generated by anchorages, and drift data caused by the failure of onboard AIS equipment or poor signal transmission must be eliminated. Therefore, we designed a three-step data cleaning process. First, we performed offset detection by focusing on sudden changes in the positions of trajectory points. We compared the offsets between the current trajectory point and other trajectory points during a fixed time period, and if an offset exceeded a predefined threshold, the trajectory point was removed. Next, we performed velocity detection and eliminated trajectory data with velocities that were below a certain threshold. Finally, start-to-finish distance detection was performed to identify ships that exhibited a certain level of speed but were in an anchored gyration state. This type of trajectory exhibits a circular shape, resulting in an excessively small distance between the start and end points of the trajectory, and should be eliminated. The specific process used to implement these filtering steps is shown below (Algorithm 1).

Algorithm 1. Trajectory Cleaning

Input: Original set of traces $M = \left\{ Traj_i \right\}$
Output: Set of cleaned traces $M' = \left\{ Traj_i' \right\}$
For m in M
    For t in m
        If $t > 4$
            Calculate the average interval between the points m_{t-3} and m_t $d_{average}$
            Calculate the interval between the points m_{t-1} and m_t d_i
            If $d_i > 6 \times d_{average}$
                Delete track points t
        If $Sog_t < 2$ kn
            Delete track points t
    Calculate the distance between the points m_{Start} and m_{End} d_{se}
    If $d_{se} < 2$NM
        Delete Tracks m
Return the set of cleaned trajectories M'

The transmission frequency of a shipboard AIS ranges from 2 s to 3 min for ships with different sailing speeds. Varying transmission frequencies and missing data lead to difficulty in unifying time intervals when analyzing AIS data. Therefore, AIS trajectory data must be interpolated such that the AIS information on ships with different sailing statuses has the same frequency. Commonly used interpolation methods include linear,

segmented linear, Bessel curve, and polynomial interpolation. In this study, we employed the cubic spline interpolation method based on its established interpolation accuracy, as validated by Zhang [24].

Ships that satisfied the spatial and temporal constraints of an encounter scenario in the cleaned AIS dataset needed to be screened to obtain ship AIS trajectory data representing encounter scenarios. Specifically, we identified pairs of ships that overlapped in the time dimension and were close to one another in the space dimension.

To improve the computational efficiency, we used the sliding-window technique to split the entire pairwise trajectory set into multiple windows. The encounter trajectories were represented by several windows $(Traj_A, Traj_B) = [P^1, P^2, ...P^l]$, where $P^l = [(Lat_A^t, Lon_A^t, Sog_A^t, Cog_A^t, Lat_B^t, Lon_B^t, Sog_B^t, Cog_B^t)_{t=1}^n]$ and l denotes the number of encounter windows.

Attributes such as the latitude, longitude, heading, course, and speed of each ship were normalized to improve the training effectiveness and convergence speed. Owing to the significant dimensional differences among these attributes, attributes with larger dimensions dominate computations. By Normalizing, these attributes are adjusted to a uniform scale, with their values scaled to the range of [0, 1].

3.3. STETC

This paper presents a trajectory prediction model, namely STETC, designed for dual-ship encounter scenarios. This model achieves high-precision predictions by explicitly considering the spatiotemporal interactions between ships and integrating Transformer and CNN architectures. The model employs a two-layer encoder-decoder framework. The encoding modules consist of a Motion Trajectory Encoder and an Encounter Interaction Encoder, which are responsible for encoding temporal and spatial interactions, respectively.

The motion trajectory encoder employs a temporal self-attention mechanism to reveal hidden low-level motion patterns from comprehensive historical trajectory data. Multiple attention heads are used to concurrently learn relationships between different dimensions in the trajectory data, resulting in enriched motion trajectory features that establish the unperturbed baseline prediction. The encounter interaction encoder integrates CNN and self-attention mechanisms for generating spatiotemporal attention. It models ship interactions in encounter scenarios as an Artificial Potential Field (APF). The CNN layers process the APF at each timestamp, acting as feature detectors to efficiently extract the instantaneous spatial structure and localized conflict geometry. Given the temporal continuity of ship encounter interactions, the subsequent self-attention mechanism then establishes remote time dependencies and quantifies the dynamic, non-linear progression of the conflict, resulting in temporally continuous and comprehensive encounter interaction features. Each encoding module is followed by a corresponding decoding module. These decoders are connected in series, progressively fusing the hidden features extracted by the encoders. This fusion derives the spatiotemporal correlations between historical trajectories and interaction behaviors, thereby inferring the future trajectories of ships in encounter scenarios. The query in the decoding module is defined as the initial future trajectory vector, and each pass through the decoder updates this vector. As noted by Carion [25], this serial decoding structure can independently receive information from different encoders, hierarchically integrate various types of contextual information, and adapt to multi-input environments.

3.3.1. Ship History Movement Information

The motion trajectory encoder is used to encode the latitude, longitude, speed, and course of two ships for all observation times in the encountered scenario. In the motion trajectory encoder, the encounter trajectory pair P^l obtained following sliding-window processing is first embedded into the fixed-length vector e^l to obtain a high-dimensional spatial representation of the trajectory within the time window, as shown in Equation (2). Next, sine and cosine functions are used to encode embedding vector e^l to ensure that the data associated with each time step had a unique absolute position in the feature sequence.

$$e^l = emb(P^l) \tag{2}$$

where emb is a fully connected layer, e^l is the embedding vector.

Next, multi-head self-attention is employed to encode e^l, capturing the temporal dependencies of the motion trajectories. The model processes input trajectory sequences from multiple perspectives and subspaces through the parallel computations of multiple attention heads. Subsequently, the processed historical ship motion data is forwarded to the feedforward neural network, enabling the model to extract deeper features.

After the motion encoder encodes the historical trajectory data, the encoded information and the randomly initialized future trajectory vector are input into the motion decoder. In the decoder's self-attention layer [26], the

query(Q), key(K), and value(V) are obtained from linear transformations of the initialized future trajectory vector, it is used to learn the internal temporal relationships within the predicted trajectory sequence. In the cross-attention layer, the model maps general behavior patterns from the historical ship trajectories to the future trajectories. Finally, the feedforward neural network outputs the future trajectory vector.

3.3.2. Ship Encounter Information

During encounters, ships frequently interact based on the states of other ships. The interaction information embedded in historical trajectories is not adequately depicted in the motion trajectory encoder, and relying solely on the multi-attention module is insufficient to capture the interactions between ships in shared spaces. This paper proposes the use of the APF matrix to represent the interaction dynamics of ships during encounters. By utilizing the operational status information of ships, this method quantifies the influence on surrounding ships, thus providing a more accurate simulation of ship behavior in complex navigational environments and enhancing the model's performance. The APF, an essential tool for navigation and obstacle avoidance in maritime contexts, is based on the concept of potential energy in physics. The APF models a ship's motion in water as navigation within a virtual potential field. In the APF, the target point (typically the destination of a route) is considered a source of attraction, whereas obstacles (e.g., other ships, islands, and buoys) are treated as sources of repulsion. The surrounding space is represented as a numerical matrix of attractive and repulsive forces. The ship is attracted to the target point while simultaneously subjected to repulsive forces to avoid collisions.

An APF was designed that considers the repulsive forces between ships and accounts for the dynamic parameters of ships, including ship-to-ground speed, relative bearing angle, and relative distance. The APF matrix is obtained by gridding the area around a ship, as shown in Figure 2, and calculating the repulsive forces exerted on other ships at each grid coordinate. The areas with higher potential field values can be considered to exert greater repulsive forces on other ships. The potential field equation is defined as follows:

$$U(p) = v \times \cos(\alpha) / (1 + d_p^2) \tag{3}$$

where p is any point around the ship, α_{A} is the angle between the ship heading and the line connecting the ship to the point, v_{A} is the ship speed, and d_p is the distance between the ship and the point.

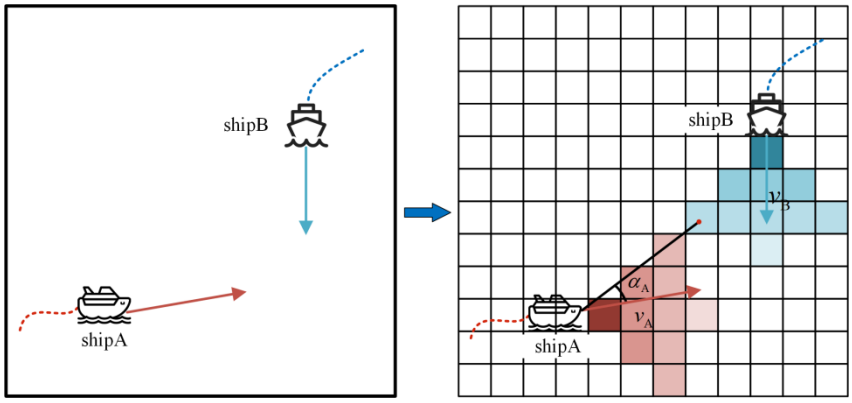

Figure 2. APF gridding.

In high-dynamic ship encounter scenarios, the primary ambiguity lies in predicting the mutually influencing maneuvers of the involved ships. The APF model effectively translates geometric proximity and relative motion states into a quantifiable field of potential conflict through kinematic-based repulsive force calculations. This transformation constitutes a non-linear geometric mapping, which is analogous to applying a high-pass filter to the raw data. This process effectively attenuates long-term navigational noise and isolates the immediate, localized conflict geometry into an image-like representation. Crucially, the potential field value at any grid location (as defined in Equation (3)) is more than a simple metric of distance; it serves as a quantitative indicator of the immediate risk and response urgency perceived by the navigating ships. Higher potential field values inherently signal an elevated requirement for collision avoidance maneuvers, a critical reactive signal decoupled from the ship's inertial motion characteristics.

Convolution operations in CNNs emulate the sensory field mechanisms of biological visual systems, facilitating the efficient capture of localized features within images. The convolutional kernels function as feature

detectors, identifying information such as edges, textures, and shapes within images. This localization assists a CNN in capturing the spatial structure of the APF matrix and accurately modeling the positional distribution and conflict geometry of two ships during an encounter. The convolutional kernels function as feature detectors, enabling the network to learn increasingly abstract representations of the APF structure through stacked layers. In the proposed method, the input to the CNN is the APF matrix generated by Equation (3). The matrix array undergoes convolution, activation, and pooling before being input into the fully connected layer to derive the encounter encoding vector. The CNN processing flow for the APF matrix is illustrated in Figure 3.

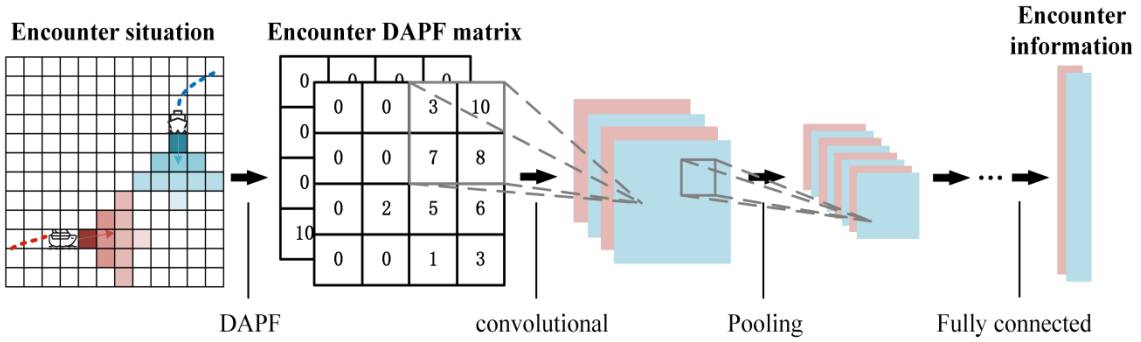

Figure 3. CNN processing of APF.

As shown in Figure 4, the input length of the encounter interaction encoder is consistent with that of the motion trajectory encoder processing the APF matrices sequentially across time steps. While the initial Convolutional Neural Network (CNN) layer excels at capturing the instantaneous, localized spatial geometry of the conflict, it inherently lacks the capacity to model the necessary global temporal dependencies. Therefore, a Self-Attention mechanism is employed immediately following the CNN to address the significant temporal correlations that exist between the spatial distribution features across different timestamps. The integration of these two components is designed to leverage their complementary strengths. The CNN first filters and extracts the immediate conflict features, and the subsequent Self-Attention mechanism then establishes remote time dependencies and quantifies the dynamic, non-linear progression of the encounter over the entire observation window. By weighting features across all time steps, the mechanism effectively learns to identify the critical moments where risk escalates or intent changes, thereby capturing advanced, time-dependent interaction dynamics essential for accurate trajectory prediction. The specific formulas are shown in Equations (4)–(6).

$$x^l = f\left(w^l \otimes x^{l-1} + b^l\right) \tag{4}$$

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{\mathrm{d}_k}}\right)V \tag{5}$$

$$\begin{cases} head_i = \text{Attention}(Q_i, K_i, V_i) \\ \text{MultiHead}(Q,K,V) = \text{Concat}(head_1, ..., head_h)W^O \end{cases} \tag{6}$$

For the convolution operation, x^l represents the output and w^l denotes the weight matrix of the l-th layer, with b^l as the bias and f as the RULE activation function. Within the attention mechanism, the Q, K, and V are obtained through linear transformations of x from the final convolutional layer. Here, h signifies the number of self-attention heads, with $i \in [1,h]$ as the head index. The dimension of the query and key is defined by $d_k = d_{model}/h$, and $W^O \in R^{hd_k \times d_{model}}$ is the output weight matrix.

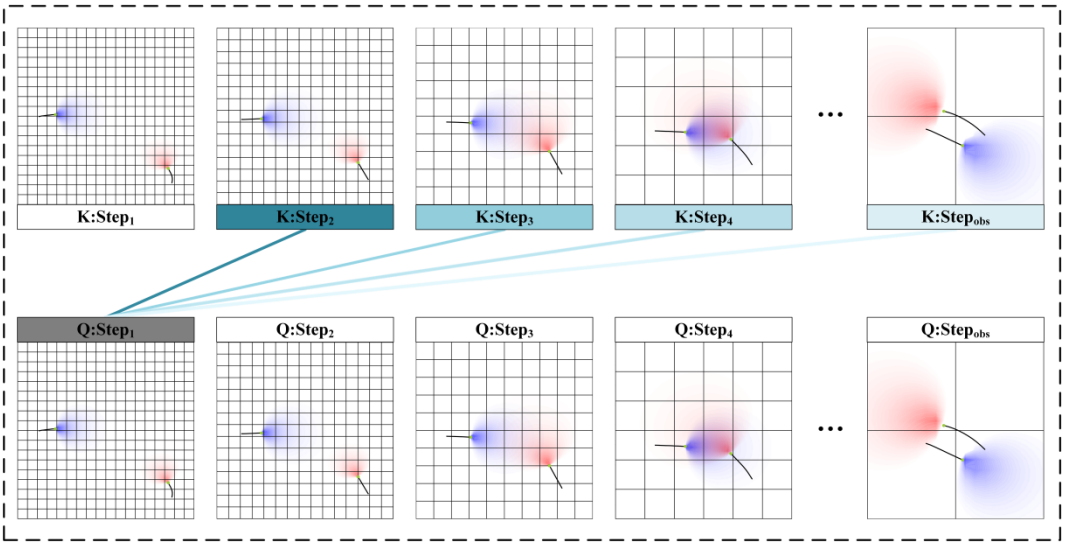

Figure 4. Relationships between the encounter information in the first time step and in the other time steps in the self-attention.

The final decoder receives outputs from the preceding decoder and the encounter interaction encoder, utilizing a cross-attention mechanism to integrate motion and interaction features. As illustrated in Figure 1, the output from the first decoder serves as the Query vector in the second decoder's Cross-Attention mechanism, while the interaction features are utilized as the Key and Value. Crucially, these interaction features are conceptualized as dynamic correction signals applied to the motion baseline trajectory, rather than as primary driving factors. This strategic design ensures that the trajectory is influenced by the interaction features only to the degree requisite for achieving effective collision avoidance. By computing precise correction weights between the unperturbed trajectory (Query) and the available interaction signals (Key/Value), the model performs fine-grained, temporally sensitive adjustments to the predicted trajectory coordinates. This targeted refinement process is instrumental in significantly enhancing prediction accuracy, particularly in high-dynamics encounter scenarios. Finally, a multi-layer perceptron, comprising three fully connected layers, reduces the dimensionality of the future trajectory vector to predict the trajectory coordinates:

$$\hat{Y}_t = mlp(S_i) \tag{7}$$

In summary, the model encodes historical trajectory information via the motion trajectory encoder, extracts dynamic encounter information using the encounter interaction encoder, and fuses ship trajectory and encounter scenario information through two decoders for sensing. The model establishes a spatiotemporal connection between historical trajectory information and encounter dynamic information through these two layers, enabling future trajectory vectors to update the future trajectories by gradually integrating different contextual information.

The fundamental rationale behind the proposed dual-encoder architecture is to achieve decoupled mapping in the feature space, thereby separating two orthogonal types of navigational information: historical trajectory information and encounter scenario information. The Motion Trajectory Encoder directly processes the raw sequential data, forcing the model to learn the inherent temporal dependencies and long-term inertial patterns of individual ships. Its output represents an unperturbed baseline predicted trajectory derived from historical momentum. In contrast, the Encounter Interaction Encoder first transforms the same raw data into an APF matrix. By feeding the APF matrix into a CNN and a self-attention module, the model is guided to capture only the short-term, spatially localized features required for collision avoidance. This strategy prevents the highly volatile, reactive signals from the APF from corrupting the smooth, predictable inertial features.

In summary, the model leverages this two-layer encoding structure to establish a robust spatiotemporal connection between historical trajectory and encounter dynamic information. This encoding separation subsequently enables the cascaded decoders to perform hierarchical feature fusion: first establishing the motion baseline, then applying necessary dynamic corrections based on the dedicated interaction features. This differentiated mapping mechanism, coupled with the hierarchical fusion process, constitutes the deep logic that allows the model to achieve superior prediction accuracy by gradually updating the future trajectory vectors with integrated contextual information.

3.4. Model Evaluation Indices

Two metrics, namely the ADE and FDE, were used to evaluate the predictive validity of the proposed model. The ADE and FDE are standardized metrics for assessing trajectory prediction models and are frequently employed as evaluation criteria in trajectory prediction studies [27]. The ADE represents the average Euclidean distance between the predicted and true trajectories, which can visually reflect the prediction accuracy of a model. The FDE represents the final Euclidean distance between the endpoints of the predicted and true trajectories, which is used to assess the model accuracy in terms of predicting the end position of a trajectory. These metrics are defined as follows:

$$\mathrm{ADE} = \frac{1}{t_{pred}} \sum_{t=1}^{t_{pred}} \left\| (Y_t - \hat{Y}_t) \right\|_2 \tag{8}$$

$$\mathrm{FDE} = \left\| (Y_{t_{pred}} - \hat{Y}_{t_{pred}}) \right\|_2 \tag{9}$$

where Y and $\hat{Y}$ denote the latitude and longitude coordinates of the real and predicted trajectories, respectively, and t_{pred} denotes the number of time steps in the window.

4. Experiments and Results

4.1. Experimental Settings

The experimental data were AIS data collected from sea areas with high traffic flows. The data were collected in September 2018, and the longitude coverage was from 122.03° E to 122.22° E, whereas the latitude coverage was from 29.78° N to 29.94° N. Preprocessing the AIS data and extracting ship trajectory pairs in encounter scenarios yielded a total of 2169 multi-ship sailing trajectory segments. The data visualization is presented in Figure 5. These track segments were classified into three types according to the COLREGs: 734 AIS track pairs representing pursuit scenarios, 956 AIS track pairs representing cross-encounter scenarios, and 479 AIS track pairs representing encounter scenarios. The sliding-window method was used to divide the dataset for the encounter scenarios, and 179,824 windows were obtained. The dataset was partitioned into training, validation, and test sets at a ratio of 6:2:2. The training set was employed to fine-tune the model parameters, including the weights and biases. The validation set was used to conduct an initial evaluation of the model's performance and fine-tune its hyperparameters. Following training, the test set was used to evaluate the generalization ability of the final model.

The STETC model utilizes a two-layer encoder–decoder structure. The two distinct encoders are the Motion Trajectory Encoder and the Encounter Interaction Encoder. They feed into their respective decoders, which are connected in a cascaded manner to facilitate hierarchical feature fusion. The input size for the Motion Trajectory Encoder is [64, 2, 10, 4], where 64 is the batch size, 2 is the number of ships, 10 represents the observation time length, and 4 denotes the feature dimension (e.g., latitude, longitude, speed, heading). The input size for the Encounter Interaction Encoder is [64, 2, 32, 32], where 64 is the batch size, 2 represents the number of APF (Artificial Potential Field) matrices, and 32 * 32 is the resolution of the APF matrix, each grid cell represents about 200 m. The detailed parameters of the model components are presented in Table 1.

Table 1. STETC network structure and parameters.

Parameter	Value	Parameter	Value
Model Dimension	64	Batch size	64
Number of Layers	2	Learning rate	0.0001
Number of Heads	8	Dropout Rate	0.4
Kernel Size	3	Optimizer	Adam
CNN Output Size	128	epoch	1000

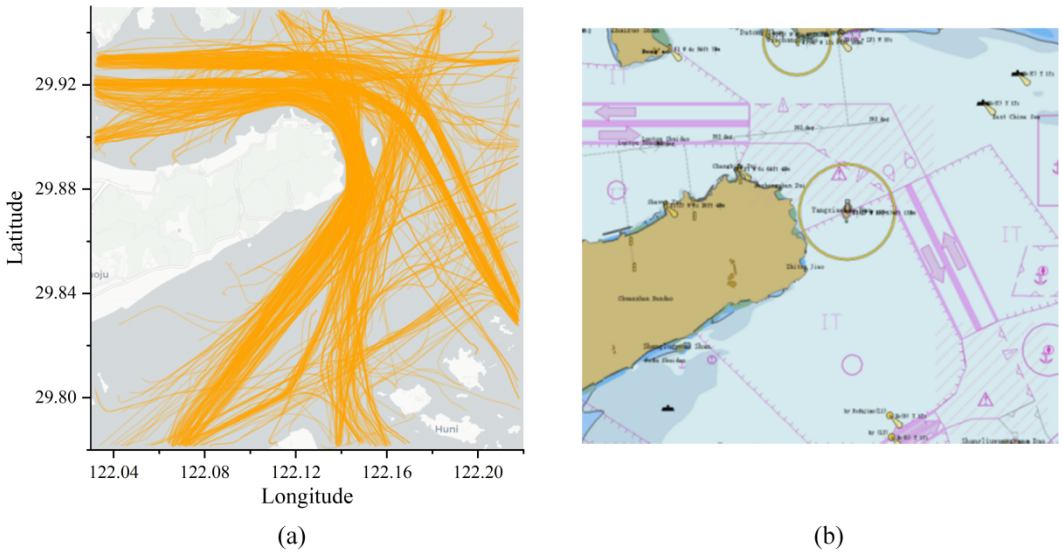

Figure 5. Study area. (**a**) Research dataset visualization. (**b**) Electronic chart of Zhoushan Luotou waterway.

4.2. Analysis of Experimental Results

4.2.1. Comparative Analysis of Experimental Results

The model subjected to experimental comparison in this study encompass the STETCT , TCN-Attention-GRU [28], Optuna-BILSTM [29], Transformer [26], Social GAN [27], and Seq2seq [30]. The observed time was set to 5 min and the predicted durations were configured as 5, 10, and 15 min to verify the predictive performance of STETC and comparative model across various prediction lengths. The time interval following AIS data preprocessing was set to 30 s. The proposed model was compared with the models described below over the different prediction lengths, with the ADE and FDE metrics employed to assess the efficacy.

The models described above were trained and tested on our collected dataset to predict trajectories from 5 to 15 min in the future using the historical trajectories at the observation times as inputs. Furthermore, when the prediction horizon was set to 5 min, the average inference time per trajectory for all models was calculated. Table 2 presents the performance contrasts between the proposed STETC model and the baseline models across various prediction lengths. The ADE and FDE values are provided, with bold font denoting the optimal ship trajectory prediction outcomes. The prediction errors of all the models increased as the prediction time increased. However, the performance of the STETC model consistently surpassed that of the other models across all prediction durations. Optuna-BILSTM, Seq2seq, and Social GAN are disadvantageous for long-distance prediction because the LSTM model may forget previous information for excessively long historical trajectory sequences, resulting in these models being unable to capture long-distance dependencies in the time series effectively. At a prediction duration of 5 min, the ADE of the STETC model was reduced by 8.9%, 11.0, 22.9 %, 16.0 %, and 12.8 % and the FDE was reduced by 15.5%, 10.6%, 19.6%, 20.7%, and 12.7% compared with those of the TCN-Attention-GRU, Optuna-BILSTM, Transformer, Social GAN, and Seq2seq models, respectively. This indicates that STETC effectively captures the encounter dynamics between ships based on their historical motion characteristics. However, for the FDE with prediction durations of 10 and 15 min, the values of STETC were only 0.4% and 7.1% smaller than those of the Transformer model, respectively. This is because ships in more distant positions may have already made it past the current intersecting waters, encounter interaction information in the APF that does not affect trajectories that are farther away, and the true endpoints of long-distance predictions are more affected by other environmental factors (e.g., meeting with other ships or turning in a channel). Overall, STETC is superior at predicting ship trajectories in encounter scenarios compared with the other models and can provide more effective reference information for maritime traffic safety warnings in intersecting waters.The comparison of the inference times reveals that STETC exhibits a relatively long processing time; however, the single-step inference time remains below 0.01 s, which is within an acceptable range for practical applications.

Table 2. Performance evaluation of STETC and baseline models.

Model	Metric (m)	5 min	10 min	15 min	Inference Time(s)
Transformer	ADE	92.64	151.46	248.09	0.0011
	FDE	176.16	306.10	515.66	
Social GAN	ADE	83.02	291.24	536.00	0.0010
	FDE	178.71	618.67	1142.83	
Seq2seq	ADE	79.98	177.75	265.05	0.0003
	FDE	162.26	362.36	561.19	
TTCN-Attention-GRU	ADE	76.57	151.46	238.14	0.0042
	FDE	167.73	306.10	495.71	
Optuna-BILSTM	ADE	78.32	162.76	263.07	0.0007
	FDE	158.49	323.08	559.82	
STETC	ADE	69.73	147.09	225.65	0.0091
	FDE	141.65	304.87	479.03	

4.2.2. Case Analysis

We selected pairs of ships of various encounter types from the test set, set both the observation and prediction times to 5 min, compared the predicted trajectories generated by STETC with those generated by the other models, and visualized the predicted trajectories and APF of the final observation time step to authenticate the trajectory prediction capability of the proposed model in authentic encounter scenarios further.

As shown in Figure 6, the two ships navigated into the intersecting water from the west and southeast directions, forming a cross-encounter scenario. In terms of the predicted trajectories, although the Social GAN, Transformer, TTCN-Attention-GRU, and Optuna-BILSTM accurately forecasted the direction of ship 2, it overlooked the impact of ship interactions, resulting in a notable deviation between the predicted trajectory of ship 1 and the actual trajectory. The predicted trajectories of STETC matched the real trajectories most closely, including the steering intention of ship 2, indicating that the proposed model can make reasonable predictions based on the perceived encounter scenarios. According to COLREG Article 15, a ship with another ship on its starboard side should give way to the other ship, meaning that ship 1 should give way in the example scenario. However, in reality, ship 2 took the initiative to avoid the other ship, leaving a sufficient safety distance. This demonstrates that STETC can flexibly capture information in encounter scenarios and learn the behaviors of ships outside the scope of the COLREGs, thus assisting ship operators in making effective safety decisions.

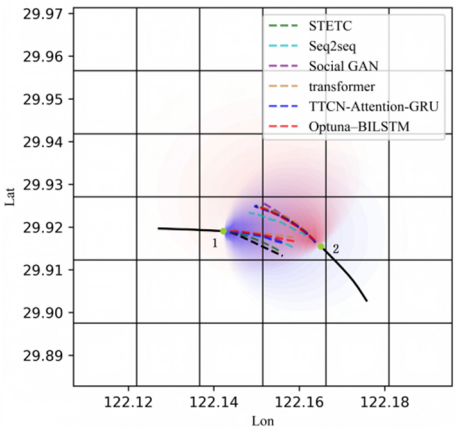

Figure 6. Schematic of the cross-encounter scenario. The solid black line represents the observed track and the dashed black line represents the true track.

As shown in Figure 7, two ships from the north and south entered a head-to-head encounter scenario. According to COLREG Article 14, when two motorized ships meet on opposite or nearly opposite courses to the extent that danger of collision exists, each ship should turn to the right to pass on the port side of the other ship. The future true trajectory shows that ship 1 and ship 2 both had a clear intention for a starboard turn. The predictions made by STETC were more closely matched with the real future trajectories compared with the other models. This indicates that the STETC model can learn ship motion laws from observed trajectories as well as capture the interactive relationships between ships from the APF through the encounter interaction encoder. This

information can be used to optimize future trajectories such that they match real trajectories in an encounter scenario more closely.

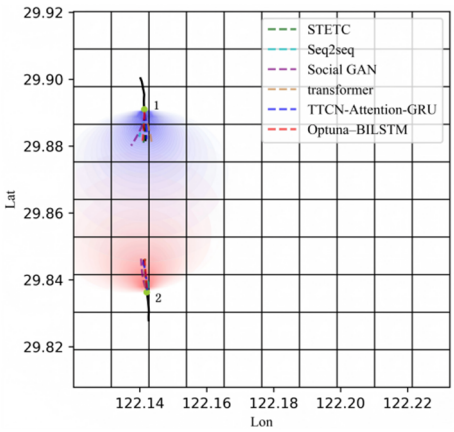

Figure 7. Schematic of the head-to-head encounter scenario. The solid black line represents the observed track and the dashed black line represents the true track.

Figure 8 presents an encounter scenario during an overrun chase, when the speed of ship 1 is greater than that of ship 2 and the two ships maintain a certain distance. For ship 1, the Social GAN and Seq2seq models were inferior to the other models in terms of capturing the global information of long sequences, and the predicted trajectory deviation was apparent. STETC had higher prediction accuracy than the other models. The enhanced efficacy of the STETC model can be ascribed to its adeptness in capturing the multilevel representations and intricate relationships inherent within the input sequence data. By utilizing the motion trajectory and spatial interaction encoders, STETC can encode the spatiotemporal and encounter features of ship trajectories efficiently and can achieve a clear perception of the ship in the overrun chase scenario, thus improving the prediction accuracy and reliability.

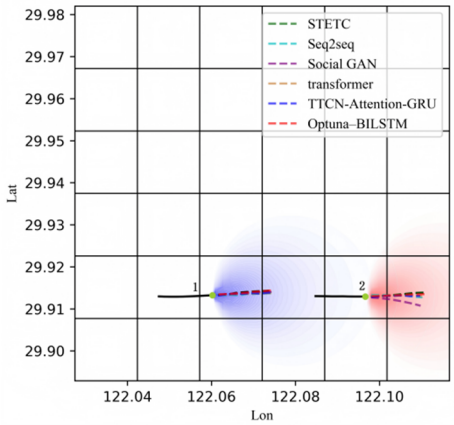

Figure 8. Schematic of the overrun chase encounter scenario. The solid black line represents the observed track and the dashed black line represents the true track.

Figure 9 presents a scenario in which two ships resume their respective routes after one ship gives way. In this scenario, the two ships did not enter an encounter scenario, and the influence of the APF on the future trajectories is small. However, our model could still make more accurate predictions compared with the other models.

However, the predictions made by STETC are not always optimal and may sometimes over-value collision avoidance behaviors in an encounter scenario. As shown in Figure 10, STETC predicted a trajectory for ship 2 that was not consistent with the true trajectory, exhibiting a large deviation compared with the predictions of the two models. It can be observed in the APF that STETC predicted a trajectory for ship 1 that deviated to the right to avoid the area with high repulsion in front of ship 2 and to leave a sufficiently wide space such that the two ships could safely pass through the intersecting water. COLREGs Article 8 stipulates that any change in course and/or speed to avoid collision shall, if circumstances at the time permit, be sufficiently large to be readily detectable by the other ship when observed visually or by radar. In practice, crews typically avoid other ships based on their own experience

or judgment, as is the case for the change of course of ship 1 in the diagram, which does not strictly comply with this rule. In this particular case, STETC placed heavy emphasis on the APF, resulting in an inaccurate prediction.

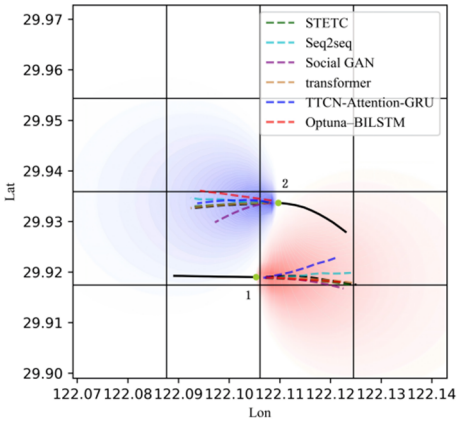

Figure 9. Schematic of the two ships restoring their course. The solid black line represents the observed track and the dashed black line represents the true track.

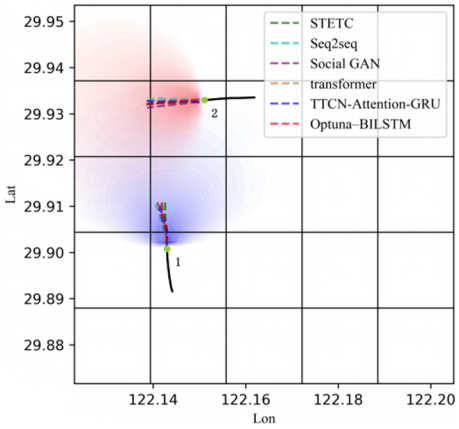

Figure 10. Schematic of the ship avoidance maneuvers. The solid black line represents the observed track and the dashed black line represents the true track.

4.2.3. Model Analysis

To explore the intrinsic behaviors of STETC when predicting trajectories, we visualized the attention weights of the second layer of the attention module in both decoders to understand the model's behavior. The state of the attention module for the encounter scenario in Figure 11, is visualized in Figures 12 and 13.

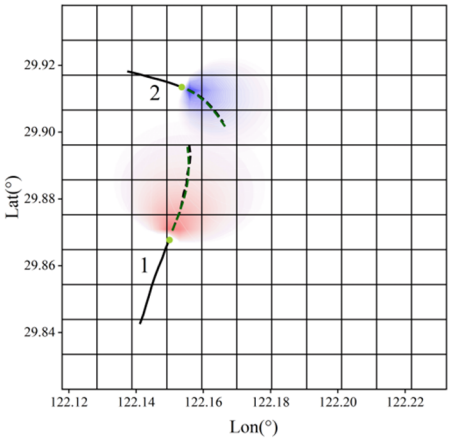

Figure 11. Schematic of the ship avoidance maneuvers. The solid black line represents the observed track and the dashed black line represents the true track.

Figure 12 presents a diagram of the attention mechanism within the first decoder, where the horizontal axis denotes the time steps of the observed trajectories. Steps 0 to 10 represent the observed trajectory of ship 1, and steps 10 to 20 represent the observed trajectory of ship 2. The vertical axis represents the time steps of the predicted trajectories, with the same order as that for the observed trajectories. It is evident that the predicted trajectories for ships 1 and 2 primarily reflected their own observed trajectories, with minimal influence from the observed trajectories of the other ship, indicating that the proposed model only considers each ship's own motion pattern in the first layer of the encoder–decoder structure, without accounting for the social interactions between ships. The model also emphasized the second half of the observed trajectory, indicating that the model perceives dynamic changes in trajectories and learns the spatiotemporal connections between observed trajectories.

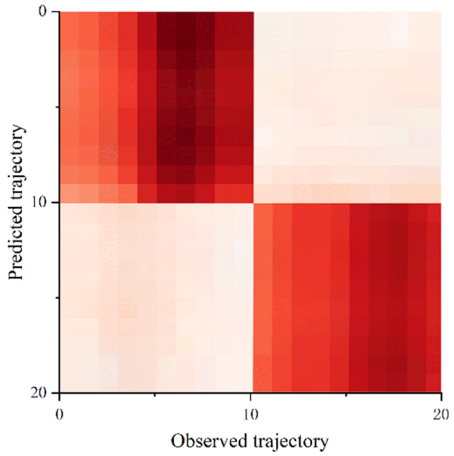

Figure 12. Visualization heatmap of the attention mechanism in the first decoder.

Figure 13 depicts the attention mechanism within the second decoder, with the horizontal axis denoting the time steps of the APF. Steps 0 to 10 represent the APF of ship 2 and steps 10 to 20 represent the observed trajectory of ship 1. The vertical axis indicates the forecasted trajectories, with the interval from 0 to 1 corresponding to the projected trajectory of ship 1 and the range from 1 to 2 indicating the anticipated trajectory of ship 2. Again, the proposed model emphasized the APF in the latter time steps. Note that ship 1 devoted significant attention to the APF of ship 2 in the final time step. Considering the results in Figures 4–6, this is likely because the spatial position of ship 2 interfered with the route of ship 1, and ship 1 had to take collision avoidance actions according to the encounter scenario, which was also verified by the real trajectory of ship 1. Therefore, it can be concluded that the encounter interaction encoder establishes a connection between the APFs of different time steps in the time dimension. In addition, the model can understand the connection between the APF and future trajectories using the second decoder, which provides accurate predictions based on the interactions between ships in an encounter scenario.

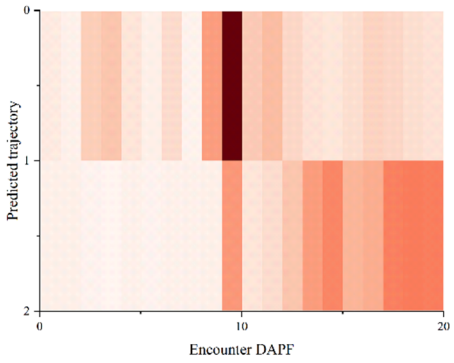

Figure 13. Visualization heatmap of the attention mechanism in the second decoder.

4.2.4. Ablation Experiments

In this section, we analyze the influence of each constituent element of the STETC model on the overall model efficacy. In Section 3, it was demonstrated that the proposed method improves the basic Transformer model. The STETC model consists of a two-layer of encoder–decoder structure. The first layer of the encoder–decoder structure is a primitive Transformer, which generates a preliminary trajectory by capturing the trajectory motion

features in a sequence of historical trajectories. The second layer processes the APF using a CNN and Transformer encoder, where the interactions between ships in an encounter scenario are modeled as an APF. The CNN is employed to analyze the APF matrix, extracting ship features in the spatial dimension within an encounter scenario, and the spatial features of the encounter are input into the transformer encoder to extract features in the time dimension. The encoder obtains continuous encounter interaction features in the time dimension and uses them to optimize future trajectories and improve the model prediction accuracy.

In our ablation experiments, we investigated the impact of the APF and CNN components on the overall STETC model. We individually incorporated the APF and CNN into the second layer of the encoder–decoder structure and assessed their effects on the ship trajectory prediction performance using the ADE and FDE metrics.

The results of the ablation experiments are presented in Table 3. After adding a Transformer encoder–decoder layer on top of the base Transformer to process the APF matrix, the ADE and FDE metrics for the ship trajectory prediction exhibited reductions of 14.6% and 9.8%, respectively. Subsequently, the integration of the CNN to analyze the APF facilitated the extraction of supplementary insights pertaining to the encounter scenario, thereby enhancing the overall performance. Incorporating the CNN led to a reduction of 11.8% in the ADE and 10.9% in the FDE for the ship trajectory prediction.

Table 3. Results of the ablation experiments.

Model	ADE (m)	FDE (m)
Transformer	92.64	176.16
Transformer + APF	79.07	158.89
STETC	69.73	141.65

5. Conclusions

We have proposed a novel model, STETC, which utilizes a two-layer encoder–decoder structure to accurately predict ship trajectories in complex encounter scenarios by integrating a Transformer and a Convolutional Neural Network.

The proposed STETC architecture implements a Dual-Encoder design to achieve decoupled feature space mapping. The Motion Trajectory Encoder constitutes the first layer, generating the unperturbed baseline trajectory by learning the ship's inertial motion features from historical sequences. The second layer, the Encounter Interaction Encoder, models the Artificial Potential Field to extract dynamic interaction features. These two sets of features are then fused hierarchically via cascaded decoders. The final decoder utilizes a Cross-Attention mechanism to apply the interaction features as dynamic correction signals to the baseline trajectory, thereby optimizing the future trajectory while maintaining temporal continuity.

A comprehensive comparison with other state-of-the-art deep learning models under various encounter scenarios highlighted the superior performance and robustness of STETC. The experimental findings validate the efficacy of STETC's approach in integrating ship inertial motion and leveraging the localized spatial interaction correlations within encounter scenarios. This method significantly enriches situational awareness, which is crucial for collision avoidance, optimized route planning, and improving overall maritime traffic efficiency. In practical navigation, the accurate predicted information provided by STETC can enable crews to navigate safely and efficiently in high-interference waters.

However, due to the integration of multiple complex structures, including Transformer blocks and a CNN, the training and inference processes of STETC require substantial computational resources. This limitation may present a challenge in resource-constrained applications or systems demanding high real-time performance. Future work will focus on the streamlining and optimization of the STETC architecture to reduce computational overhead and enhance operational efficiency for practical deployment. Furthermore, recognizing the complexity of real-world waterways, a critical avenue for future research is the explicit investigation and modeling of multi-ship encounter prediction, aiming to capture the interwoven dynamic interaction relationships among three or more vessels.

Author Contributions

H.Y.: Methodology, Conceptualization, Visualization, Writing—original draft. F.L.: writing—reviewing and editing, Validation. X.J.: Validation, Data curation. D.Y.: Supervision. M.S.: Software. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Wuhan Second Ship Design and Research Institute.

Data Availability Statement

The data that has been used is confidential.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

1. Liu, T.; Ma, J. Ship Navigation Behavior Prediction Based on AIS Data. *IEEE Access* **2022**, *10*, 47997–48008.
2. Deng, N.; Xu, J.; Deng, D. Reactive Collision Avoidance Planning for Intelligent Ships. *Ship Eng.* **2020**, *42*, 115–121.
3. Gao, J.; Sun, C.; Zhao, H.; et al. VectorNet: Encoding HD Maps and Agent Dynamics from Vectorized Representation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11525–11533.
4. Sadeghian, A.; Kosaraju, V.; Sadeghian, A.; et al. SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 1349–1358.
5. Xue, H.; Huynh, D.Q.; Reynolds, M. SS-LSTM: A Hierarchical LSTM Model for Pedestrian Trajectory Prediction. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 1186–1194.
6. Kosaraju, V.; Sadeghian, A.; Martín-Martín, R.; et al. Social-BiGAT: Multimodal Trajectory Forecasting Using Bicycle-GAN and Graph Attention Networks. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
7. Guo, S.; Liu, C.; Guo, Z.; et al. Trajectory Prediction for Ocean Ships Based on K-Order Multivariate Markov Chain. In *Wireless Algorithms, Systems, and Applications*; Springer: Cham, Switzerland, 2018; pp. 140–150.
8. Zhao, S.B.; De-Jun, S.W. Track Prediction of Ship in Controlled Waterway Based on Improved Kalman Filter. *J. Comput. Appl.* **2012**, *32*, 3247.
9. Liu, J.; Shi, G.; Zhu, K. Ship Trajectory Prediction Model Based on AIS Sensor Data and Adaptive Chaos Differential Evolution Support Vector Regression (ACDE-SVR). *Appl. Sci.* **2019**, *9*, 2983.
10. Zhang, K. Bayesian Trajectory Prediction for a Hypersonic Gliding Reentry Vehicle Based on Intent Inference. *J. Astronaut.* **2018**, *39*, 1258.
11. Virjonen, P.; Nevalainen, P.; Pahikkala, T.; et al. Ship Movement Prediction Using k-NN Method. In Proceedings of the 2018 Baltic Geodetic Congress (BGC Geomatics), Olsztyn, Poland, 21–23 June 2018; pp. 304–309.
12. Gasparin, A.; Lukovic, S.; Alippi, C. Deep Learning for Time Series Forecasting: The Electric Load Case. *CAAI Trans. Intell. Technol.* **2022**, *7*, 1–25.
13. Mukherjee, S.; Sadhukhan, B.; Sarkar, N.; et al. Stock Market Prediction Using Deep Learning Algorithms. *CAAI Trans. Intell. Technol.* **2023**, *8*, 82–94.
14. Tu, E.; Zhang, G.; Rachmawati, L.; et al. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey from Data to Methodology. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 1559–1582.
15. Ding, M.; Su, W.; Liu, Y.; et al. A Novel Approach on Ship Trajectory Prediction Based on Variational LSTM. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 27–29 June 2020; pp. 206–211.
16. Liu, H.; Liu, Y.; Li, B.; et al. Ship Abnormal Behavior Detection Method Based on Optimized GRU Network. *J. Mar. Sci. Eng.* **2022**, *10*, 249.
17. Chen, X.; Wei, C.; Zhou, G.; et al. Automatic Identification System (AIS) Data Supported Ship Trajectory Prediction and Analysis via a Deep Learning Model. *J. Mar. Sci. Eng.* **2022**, *10*, 1314.
18. Wang, X.; Xiao, Y. A Deep Learning Model for Ship Trajectory Prediction Using Automatic Identification System (AIS) Data. *Information* **2023**, *14*, 212.
19. Bao, K.; Bi, J.; Gao, M.; et al. An Improved Ship Trajectory Prediction Based on AIS Data Using MHA-BiGRU. *J. Mar. Sci. Eng.* **2022**, *10*, 804.

20. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271.
21. Jiang, D.; Shi, G.; Li, N.; et al. TRFM-LS: Transformer-Based Deep Learning Method for Ship Trajectory Prediction. *J. Mar. Sci. Eng.***2023**, *11*, 880.
22. Zou, B.; Li, W.; Hou, X.; et al. A Framework for Trajectory Prediction of Preceding Target Vehicles in Urban Scenario Using Multi-Sensor Fusion. *Sensors* **2022**, *22*, 4808.
23. Yang, C.; Pei, Z. Long-Short Term Spatio-Temporal Aggregation for Trajectory Prediction. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4114–4126.
24. Zhang, D.; Li, J.; Wu, Q.; et al. Enhance the AIS Data Availability by Screening and Interpolation. In Proceedings of the 2017 4th International Conference on Transportation Information and Safety (ICTIS), Banff, AB, Canada, 8–10 August 2017; pp. 981–986.
25. Carion, N.; Massa, F.; Synnaeve, G.; et al. End-to-End Object Detection with Transformers. In *Computer Vision-ECCV 2020*; Springer: Cham, Switzerland, 2020; pp. 213–229.
26. Vaswani, A.; Shazeer, N.; Parmar, N.; et al. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
27. Gupta, A.; Johnson, J.; Fei-Fei, L.; et al. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 2255–2264.
28. Lin, Z.; Yue, W.; Huang, J.; et al. Ship Trajectory Prediction Based on the TTCN-Attention-GRU Model. *Electronics* **2023**, *12*, 2556.
29. Zhou, Y.; Dong, Z.; Bao, X. A Ship Trajectory Prediction Method Based on an Optuna–BILSTM Model. *Appl. Sci.* **2024**, *14*, 3719.
30. Wu, W.; Chen, P.; Chen, L.; et al. Ship Trajectory Prediction: An Integrated Approach Using ConvLSTM-Based Sequence-to-Sequence Model. *J. Mar. Sci. Eng.* **2023**, *11*, 1484.