

Article

# An Improved Ship Trajectory Prediction Based on AIS Data Using MHA-BiGRU

Kexin Bao <sup>1,2,†</sup>, Jinqiang Bi <sup>1,2,3,†</sup> , Miao Gao <sup>3,\*</sup>, Yue Sun <sup>4</sup>, Xuefeng Zhang <sup>3</sup> and Wenjia Zhang <sup>1,2</sup>

<sup>1</sup> Tianjin Research Institute for Water Transport Engineering, Ministry of Transport of the People's Republic of China (M.O.T.), Tianjin 300456, China; baokx@tiwte.ac.cn (K.B.); bijq@tiwte.ac.cn (J.B.); 1120180117@mail.nankai.edu.cn (W.Z.)

<sup>2</sup> National Engineering Research Center of Port Hydraulic Construction Technology, Tianjin 300456, China

<sup>3</sup> School of Marine Science and Technology, Tianjin University, Tianjin 300192, China; xuefeng.zhang@tju.edu.cn

<sup>4</sup> Shenzhen Ansoft Huishi Technology Co., Ltd., Tianjin Branch, Tianjin 300210, China; sun20180315@outlook.com

\* Correspondence: gaomiao@tju.edu.cn

† These authors contributed equally to this work.

**Abstract:** According to the statistics of water transportation accidents, collision accidents are on the rise as the shipping industry has expanded by leaps and bounds, and the water transportation environment has become more complex, which can result in grave consequences, such as casualties, environmental destruction, and even massive financial losses. In view of this situation, high-precision and real-time ship trajectory prediction based on AIS data can serve as a crucial foundation for vessel traffic services and ship navigation to prevent collision accidents. Thus, this paper proposes a high-precision ship track prediction model based on a combination of a multi-head attention mechanism and bidirectional gate recurrent unit (MHA-BiGRU) to fully exploit the valuable information contained in massive AIS data and address the insufficiencies in existing trajectory prediction methods. The primary advantages of this model are that it allows for the retention of long-term ship track sequence information, filters and modifies ship track historical data for enhanced time series prediction, and models the potential association between historical and future ship trajectory status information with the current state via the bidirectional gate recurrent unit. Significantly, the introduction of a multi-head attention mechanism calculates the correlation between the characteristics of AIS data, actively learns cross-time synchronization between the hidden layers of ship track sequences, and assigns different weights to the result based on the input criterion, thereby enhancing the accuracy of forecasts. The comparative experimental results also verify that MHA-BiGRU outperforms the other ship track prediction models, demonstrating that it possesses the characteristics of ease of implementation, high precision, and high reliability.

**Keywords:** ship trajectory prediction; AIS data; MHA-BiGRU; multi-head attention mechanism; bidirectional-RNN structure; GRU



**Citation:** Bao, K.; Bi, J.; Gao, M.; Sun, Y.; Zhang, X.; Zhang, W. An Improved Ship Trajectory Prediction Based on AIS Data Using MHA-BiGRU. *J. Mar. Sci. Eng.* **2022**, *10*, 804. <https://doi.org/10.3390/jmse10060804>

Academic Editor: Mihalis Golias

Received: 9 May 2022

Accepted: 6 June 2022

Published: 11 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent decades, as the shipping industry of China has grown by leaps and bounds, water transportation has been confronted with issues, such as an increase in ship traffic density, the frequency of water traffic accidents, and the increasing difficulty of maritime safety supervision, all of which pose obstacles to the sustainable growth of the shipping industry [1,2]. AIS data, whose primary information is spatiotemporal data consisting of ship location and time, provides ship trajectory data that can be used to analyze ship navigation behavior in real time, as well as provide critical supplementary information in the process of collision avoidance [3,4]. The target ship trajectory can be predicted based on known historical location information using full analysis and deep mining of AIS ship behavior data, which can provide a strong reference for the supervision of vessel traffic

services (VTS), allowing for the timely detection and resolution of abnormal and non-standard ship navigation problems [5,6]. Therefore, real-time and accurate ship trajectory prediction can contribute significantly to ensuring water traffic safety and enhancing the efficacy of water traffic guarantee. Methods for ship trajectory prediction can be broadly categorized into two types: kinematic modeling-based approaches and neural network modeling-based approaches, respectively [7].

Methods based on kinematic modeling are widely used in the ship trajectory prediction industry, with the most common being the Gaussian process regression models (GP) and the Kalman filter (KF). With time as the independent variable, Anderson measured the trajectory as a one-dimensional Gaussian process. This method determines the posterior distribution of the projected value by extracting the joint prior density and covariance matrix of the observed value and the anticipated value, as well as models smooth trajectory estimation with the aid of dynamical systems [8]. Rong et al. regarded the shipping local as a Gaussian distribution and used GP modeling to forecast the route of a ship [9]. Jiang proposed constructing a polynomial Kalman filter to suit the nonlinear system based on the classic Kalman filter theory, compensating for the lack of track location data information and sluggish update, and predicting the ship's trajectory based on the longitude and latitude data [10]. These aforementioned methods function effectively when the ship's navigation behavior state is somewhat steady. However, ship dynamics are typically sensitive to distinct environmental excitations in different areas, which may result in a non-stationary condition and render the prediction result less accurate in reality.

The widespread usage of neural networks has ushered in a new stage in ship trajectory prediction. Giulia et al. developed a radial basis neural network for the construction of a short-term vessel prediction [11]. Zhou et al. built a track prediction model based on a three-layer back-propagation (BP) neural network, the training and prediction results of which match the standards of the VTS for accuracy, real time, and universality. However, due to the fact that the hidden units of this model are fewer in number, its expressive capacity is constrained [12]. Liu et al. suggested a trace estimation method with support vector regression and used an enhanced differential evolution approach to optimize the parameters of this model [13]. However, these solutions cannot effectively overcome the problem of long-term sequence dependency.

Due to AIS data being typically time series data, it is required to evaluate not only the present time step's ship trajectory but also the previously observed trajectory data in order to anticipate future ship trajectory. A recurrent neural network (RNN) can be regarded as a representative neural network capable of predicting future data using time series information, despite gradient-vanishing and gradient-explosion problems [14,15]. To work out these gradient errors of RNN, long short-term memory (LSTM) introduces the memory unit and gate mechanism to replace the hidden layer unit in RNN [16]. Additionally, then, Ger et al. optimized the LSTM by introducing a forget gate, which enables the LSTM to learn to reset itself [17]. The gated recurrent unit (GRU) is an excellent variation on LSTM, in that it only requires an update and reset gate to regulate the information flow [18]. Thus, due to their effectiveness in time series prediction, RNN and its variant models have been applied to the field of ship trajectory prediction in recent years. Ferrandis et al. established the LSTM method to predict the ship trajectory and solve the problem of the gradient vanishing and gradient explosion of RNN owing to rising data length [19]. Agarap utilized the GRU method for time series prediction and proved this method has a good performance and is suitable for time series forecasting [20]. The bidirectional recurrent neural network structure enables the output layer to receive complete past and future information for each point in the input sequence [21]. Gao et al. and Siami-Namini et al. created a bidirectional structure to improve contextual relevance based on the RNN method, which improves the accuracy of the ship trajectory prediction compared to RNN alone [22,23]. It is worth mentioning that Stateczny et al. proposed the optimum dataset method, which contributes to comparative navigation and provides a model for big data set processing [24]. After the application of attention mechanism (AM) in the field of image recognition, Vaswani et al. used

this mechanism to replace the recurrent neural network modeling, provided a model for machine translation, and then, it became prevalent in regression problems [25]. Cheng et al. implemented AM in the area of ship trajectory prediction, with the attention modes enhancing the AIS data characteristics extracted by each block and the attention module classifying these characteristics [26].

However, although these deep-learning approaches based on AIS data performed reasonably well at predicting ship trajectory, there are still a few issues with insufficient accuracy and real-time enforcement. The primary reason for these issues is that the majority of existing approaches for mining AIS data are relatively isolated and overlook elements such as AIS data characteristics and ship track sequence information. Thus, a high-precision ship track prediction model based on a combination of multi-head attention mechanism and bidirectional gate recurrent unit (MHA-BiGRU) is developed to solve the issues mentioned above. The contribution of this model is briefly summarized below: Firstly, this model retains long-term ship track sequence information, filters and modifies ship track historical data for enhanced time series prediction, and models the potential association between historical and future ship trajectory status information with the current state, thereby increasing forecast accuracy. Secondly, an MHA mechanism based on BiGRU is introduced, which not only calculates the correlation between the characteristics of AIS information but also actively learns cross-time synchronization between the hidden layers of the output and input ship track sequences and assigns different weights to the result based on the input criterion, thereby improving the accuracy and robustness of the overall model. Finally, the comparative experimental results in this paper verify that MHA-BiGRU, which fully exploits the advantages of bidirectional RNN, multi-head attention mechanisms, and GRU, outperforms the other seven ship track prediction models, demonstrating that the MHA-BiGRU possesses the characteristics of ease of implementation, high precision, and high reliability.

## 2. Materials and Methods

Figure 1 depicts the framework of the proposed method, which consists of four components: data processing, MHA-BiGRU model proposal, MHA-BiGRU model training, and comparison experiments. Specifically, data processing, which includes ship trajectory extraction, missing value recognition and completion, and data cleansing, is a crucial step in deep learning, as the processed data enable an improved model performance. An easy-to-implement method that is suitable for quick and concise analysis is proposed by combining the advantages of bidirectional RNN, multi-head attention mechanism, and GRU, which enables the improvement of prediction efficiency and accuracy of the ship trajectory. Additionally, then, the structure, application principle, training method, and contribution of the MHA-BiGRU are presented in a step-by-step manner. Finally, in order to demonstrate the effectiveness of the proposed method, some other prediction methods are compared in this paper.

### 2.1. AIS Data Processing

The AIS is a critical component of modern ship navigation systems, which is installed and widely available for ships to reinforce the capacity to mark the location and identify targets. There are two major issues with trajectory prediction using AIS data: time interval inconsistency and measurement error. The former issue is caused by a variety of circumstances, including variability in the broadcast frequency and packet losses. The latter issue occurs when the received AIS data value does not match the true value of the sensor at the moment of measurement, and the deviation can be rather considerable [7,27]. These two issues may result in data loss, sparsity, and offset. Thus, processing data, such as ship trajectory extraction, missing value recognition and completion, and data cleaning, are vital stages in deep learning, as processed data enable model performance to be improved.

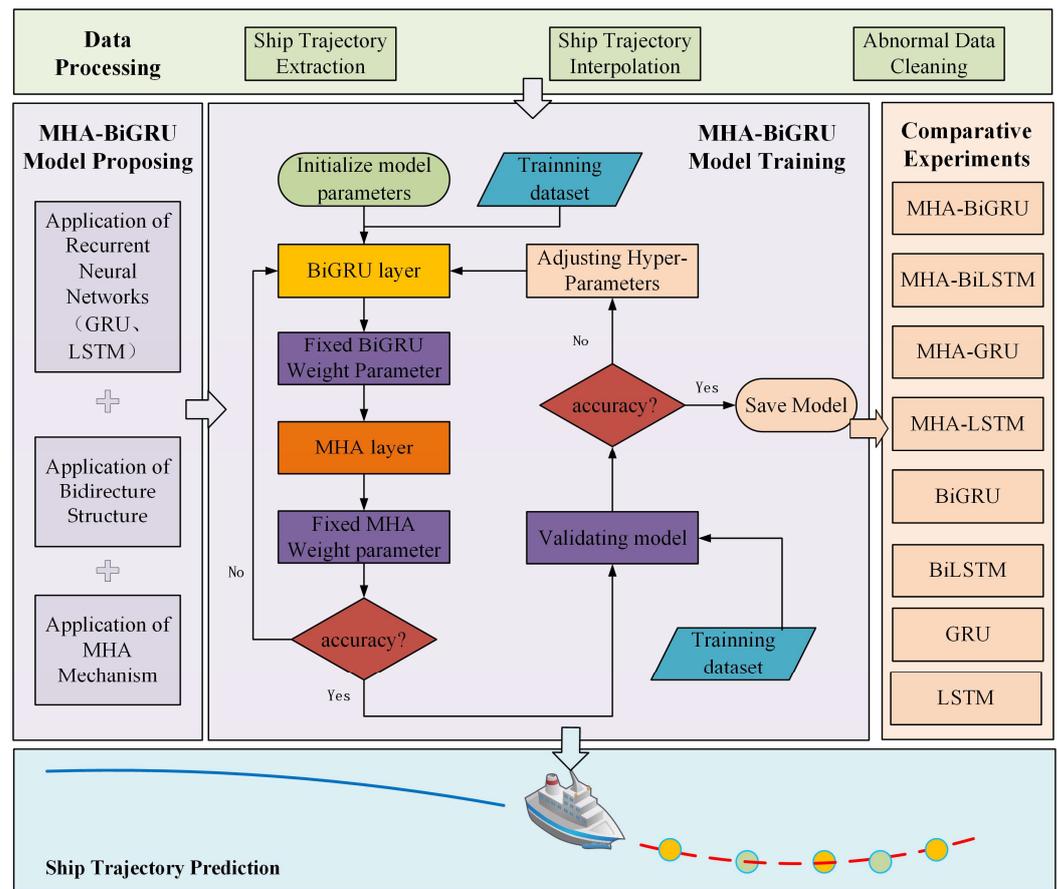


Figure 1. Flowchart of the ship trajectory prediction method.

AIS data are multidimensional and multiparametric in nature and are used to characterize ship behavior, such as the direction, position, and speed of the ship, as they change over time [28]. Each ship was classified based on its Maritime Mobile Service Identification (MMSI). After that, the ships were sorted according to their timestamps. To handle deficiency, deviation, and sparse AIS data from the original dataset, this section employs the following data processing techniques: ship trajectory extraction, deficiency value recognition, linear interpolation, and data cleaning.

The method for extracting the ship trajectory is based on time intervals and navigation speed. When the time interval between the ship trajectory points reaches 6 h, or the ship navigation speed reaches 0, the ship trajectory points are identified as tangent points to the trajectory sequence. Each track point contains information about the longitude and latitude positions, as well as its navigation speed and direction of ships.

Let the original data be  $s = \{p_1, p_2, \dots, p_N\}$  and the time interval between  $p_i$  and  $p_{i+1}$  be  $\Delta t_i$ . When  $\Delta t_i$  exceeds 10 min, the linear interpolation method is used to complete the missing value, with one deficiency value being completed every 5 min. If  $(t_j, p_j)$  is the deficiency data and  $(t_i, p_i)$  and  $(t_k, p_k)$  are the two data points closest to the deficiency data, then the completed data can be shown as follows [29]:

$$p_j^* = p_k + \frac{p_k - p_i}{t_k - t_i} (t_j - t_i) \tag{1}$$

Additionally, to address the ship trajectory deviation and sparse data, set  $p_i$  as the current track point. If the distance between the current track point  $p_i$  and its adjacent track point  $p_{i-1}, p_{i+1}$  is greater than the threshold, the adjacent track point  $p_{i-1}, p_{i+1}$  should be used as the observation point for linear fitting. When the track is too sparse and a significant amount of data are missing, the sparse ship trajectory is removed and no longer used.

### 2.2. Comparasion of GRU and LSTM

To work out the gradient-vanishing and gradient-explosion problems of RNN, LSTM introduces a memory unit and a gate mechanism to replace the hidden layer unit in RNN [15,16]. The LSTM modifies the current state of the memory cell and determines the output content via the forget gate, input gate, and output gate, which correspond to the writing and reading of the ship track reading characteristic data sequence and the reset operation of the previous state, respectively, in this paper. GRU is a great variation of LSTM, in that it requires only an update and reset gate to govern the flow of information. As a result of its smaller parameters compared to LSTM, it is extremely easy to train and enables it to respond more effectively to the implications of this information on current time inputs [18,29]. The comparison of the LSTM and GRU neural network structures can be seen in Figure 2 [17,18], and then, the following describes the concrete calculation process for these two models:

The following section details the precise calculation procedure employed by LSTM.

The amount of memory cell information used at the previous moment is controlled by the forget gate ( $f_t$ ).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

The input gate ( $i_t$ ) enables the control of the amount of information updated by the memory unit.  $\tilde{C}_t$  is a candidate vector produced by the tanh layer and will be added to the cell state. Additionally, then, it integrates the  $C_{t-1}$  with the  $C_t$  to update the cell units.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{5}$$

The output gate ( $O_t$ ) controls the amount of information output to the next hidden state. The output value is passed to the status value ( $h_t$ ) of the next unit to complete the training procedure.

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = O_t \odot \tanh(C_t) \tag{7}$$

The description of the concrete calculation process of GRU is as follows.

The reset gate ( $r_t$ ) enables the determination of how to combine the new input information with the previous memory. Additionally, when it is turned off, GRU cells can effectively forget the previous calculation and return to the state in which they are reading the first input sequence, so as to achieve the purpose of the reset.

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{8}$$

The update gate ( $z_t$ ) determines the activation status of GRU cells and the degree of update content.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{9}$$

The reset gate is applied to the  $h_{t-1}$  vector, and the obtained result is multiplied by  $x_t$  to form a splicing vector with  $x_t$ . The obtained result is transformed into a vector with elements between  $-1$  and  $1$  through the tanh function, and the candidate hidden state value is obtained. Through the above steps, the final hidden layer output information can be obtained.

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \tag{10}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{11}$$

where [ ] represents the multiplication of two vectors,  $\cdot$  means matrix multiplication,  $\odot$  shows that each element in the matrix is multiplied accordingly,  $W$  and  $b$  are the weight item of corresponding gates and bias items, respectively,  $\sigma$  is the sigmoid activation function.

Overall, as shown in Figure 2, GRU integrates  $f_t$  and  $i_t$  of the LSTM unit into  $z_t$ , and it also integrates the hidden state and unit state of the LSTM with the  $r_t$ , which can be used to control the extent of ignoring the states information of the previous time, so as to master the flow of vessel trajectory information. Based on this, GRU preserves the most critical data in order to avoid information loss during long-term propagation. Because the structure of GRU is simpler than that of LSTM, fewer parameters must be taught, and it also offers the benefit of quick training speed throughout the training process.

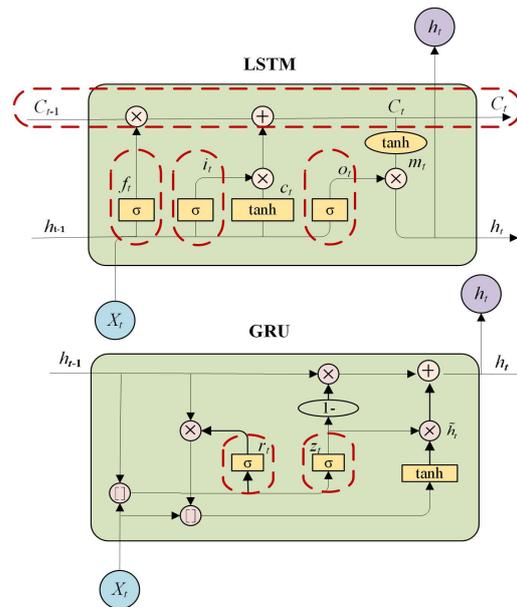


Figure 2. Comparison of LSTM and GRU neural network structures.

### 2.3. Application of Bidirectional RNN Structure

The bidirectional recurrent neural network structure enables the output layer to receive complete past and future information for each point in the input sequence. To be more precise, the forward RNN learns from previous data, while the reverse RNN learns from future data, so that each time step makes optimal use of upper- and lower-related data. Additionally, then, these two outputs are spliced together as the final output of the whole bidirectional RNN [21,30].

From this, BiGRU is a bidirectional RNN neural network that employs the GRU for each hidden node [31]. BiGRU divides GRU neurons into forward and backward layers that correspond to positive and negative time directions, respectively.

As shown in Figure 3 [21,29], the current statement of the hidden layer of BiGRU is determined by current input  $x_t$ , the hidden layer statement output of the forward layer  $\vec{h}_{t-1}$  and the backward layer  $\overleftarrow{h}_{t-1}$ . Since BiGRU can be regarded as two single GRU, the hidden layer state of BiGRU at time t can be obtained by the weighted sum of  $\vec{h}_{t-1}$  and  $\overleftarrow{h}_{t-1}$ , which can be shown as follows:

$$\vec{h}_t = \text{GRU}(x_t, \vec{h}_{t-1}) \tag{12}$$

$$\overleftarrow{h}_t = \text{GRU}(x_t, \overleftarrow{h}_{t-1}) \tag{13}$$

$$h_t = w_t \vec{h}_t + v_t \overleftarrow{h}_t + b_t \tag{14}$$

In conclusion, BiGRU enables the modeling of the potential association between historical and future ship trajectory status information with the current state, hence increasing forecast accuracy.

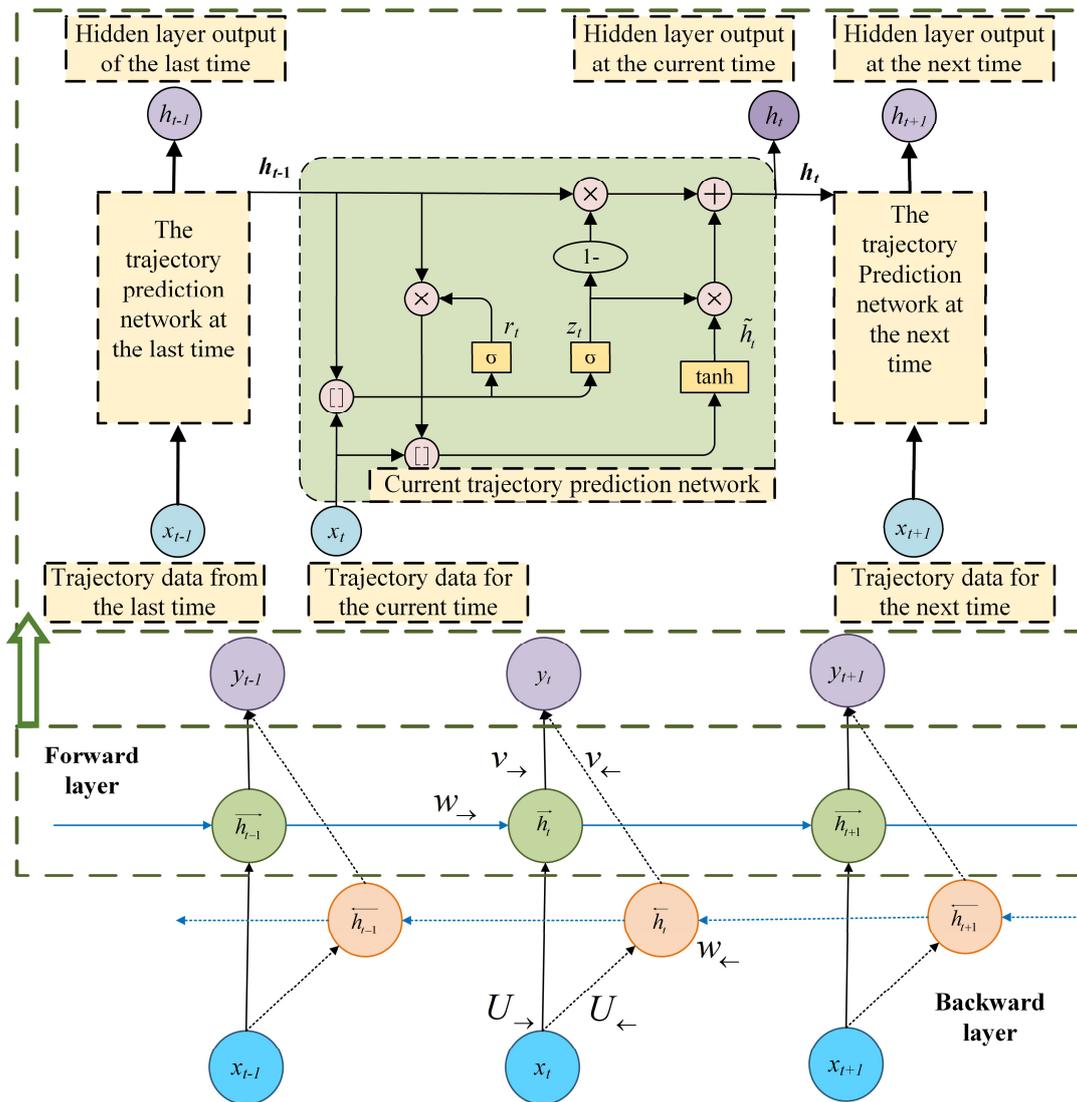


Figure 3. BiGRU structure.

2.4. Application of MHA Mechanism

The attention-based model originated in the field of image recognition and can now be used in place of RNN in the area of machine translation. By assigning a different weight to each factor in the input sequence, the attention-based model highlights the most significant influencing factors, thereby increasing the model’s accuracy. It is expressed as follows [26]:

$$f(x_i, y) = (W_1 * x_i, W_2 * y) \tag{15}$$

$$Attention = \sum_{i=1}^n softmax(f(x_i, y)) * x_i \tag{16}$$

Where  $x_i$  represents the input sequence, It is mapped in the (0, 1) interval through the normalized exponential function, which is “weight”. Additionally, dot product attention is the weighted combination of  $x_i$ .

With the attention-based model mechanism being widely used in image and natural language processing tasks, the multi-head attention (MHA) mechanism emerges as the situation requires [32]. An MHA is a combination of multiple self-attention structures. Using the query and key, the MHA mechanism calculates the weight coefficient of the relevant value and then performs weighted summation. MHA works by performing a

linear transformation on the query, key, and value and then inserting them into the zoom point to garner attention; this process is repeated a number of times. Additionally, each iteration’s linear transformation parameters  $W$  for  $Q$ ,  $K$ , and  $V$  are unique; they are not shared. Rather than using simple maximum or average pooling, MHA is used to process the data from the BiGRU output layer, as demonstrated by the following formula:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{17}$$

$$MultiHead(Q, K, V) = (head_1 \oplus head_2 \oplus head_3 \oplus \dots \oplus head_h)W^o \tag{18}$$

Thus, the multi-head attention mechanism, which is a combination of multiple attention-based models, can be regarded as a weighting scheme for information, which can assign weights to the hidden layer of BiGRU, so that they can make more rational use of information sources when making predictions.

### 2.5. MHA-BiGRU Model

By combining the advantages of bidirectional RNN, multi-head attention mechanism, and GRU, the MHA-BiGRU model is proposed as an easy-to-implement method suitable for quickly and succinctly analyzing ship trajectory. This model improves the prediction efficiency and accuracy of ship trajectory. This section introduces the MHA-BiGRU model in a hierarchical fashion and demonstrates the benefits of this method. Figure 4 vividly illustrates the structure of the proposed model.

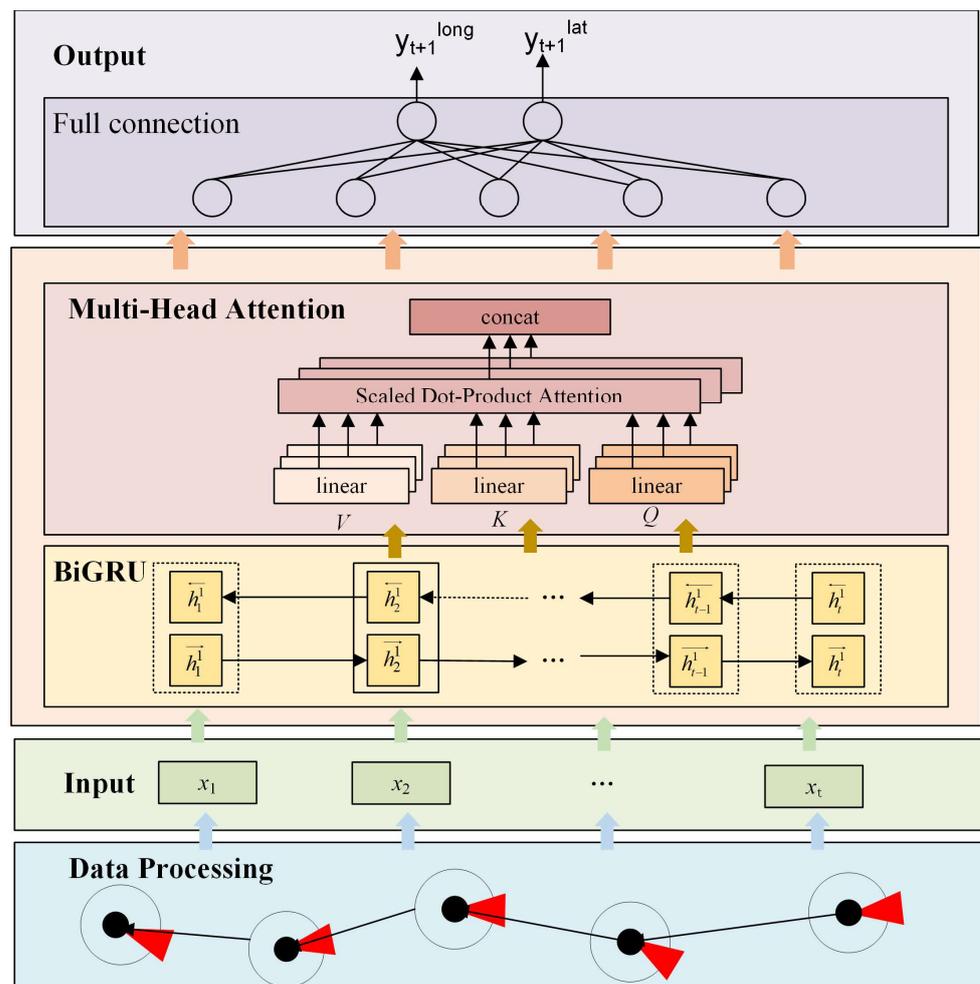


Figure 4. Schematic diagram of the MHA-BiGRU.

The MHA-BiGRU model retains long-term ship track sequence information, filters and modifies ship track historical data for enhanced time series prediction via GRU, and models the potential association between historical and future ship trajectory status information with the current state via the BiGRU structure, thereby increasing forecast accuracy.

Additionally, then, in order to resolve the common problems associated with RNN, which include AIS data being relatively isolated and overlooking elements such as AIS data characteristics and ship track sequence information, it is essential to implement an MHA mechanism based on the BiGRU structure. Firstly, this method allows for the calculation of the correlation between AIS information characteristics, such as time, latitude, longitude, speed, course, and heading, and the critical of the global impact. That is, a weighted representation is obtained by using attention sort and then put into a feedforward neural network to obtain a new representation that takes into account the correlation between various parameters.

Secondly, because the vector length is difficult to summarize with the complete track sequence information, and the information input after BiGRU will dilute the information of the previous vector to a certain extent, the accuracy of the fixed context vector response track data will gradually decrease. In addition, because the ship operation in the application scenario changes dynamically with time, to address the aforementioned issue, the MHA mechanism can actively learn the degree of cross-time synchronization between the hidden layers of the output and input sequences and assign different weights to the result based on the input criterion, thereby improving the accuracy and robustness of the overall model.

### 3. Comparative Experiments and Results

#### 3.1. Experimental Dataset

The following comparative experiments were performed using Python 3.7, Keras 2.1.4, and Pytorch 4.0 in a software environment. The original AIS data for August were collected in the coastal waters near the port entrance of Lianyungang, China. In order to avoid overfitting, the original AIS dataset characteristics of MMSI, Time, Latitude, Longitude, Speed, Course, and Heading were selected for deep learning. The experimental dataset was then created by completing data processing, including data extraction, missing value recognition and completion, data cleaning, and characteristics extraction from the original AIS dataset, which can be divided into the training set, the validation set, and the test set. Among them, the training set was utilized to train and determine the model's weight, bias, and other parameters. The validation set was employed to validate the model's performance and enhance its generalizability by adjusting the hyper parameters. After training, the test set was used to evaluate the final model [33]. The first 80% of this experimental dataset served as the training set, while the other 20% served as the test set. In order to accelerate the convergence of the model and enhance its precision, the maximum–minimum normalization was applied, so that all of the values are concentrated within the interval (0, 1).

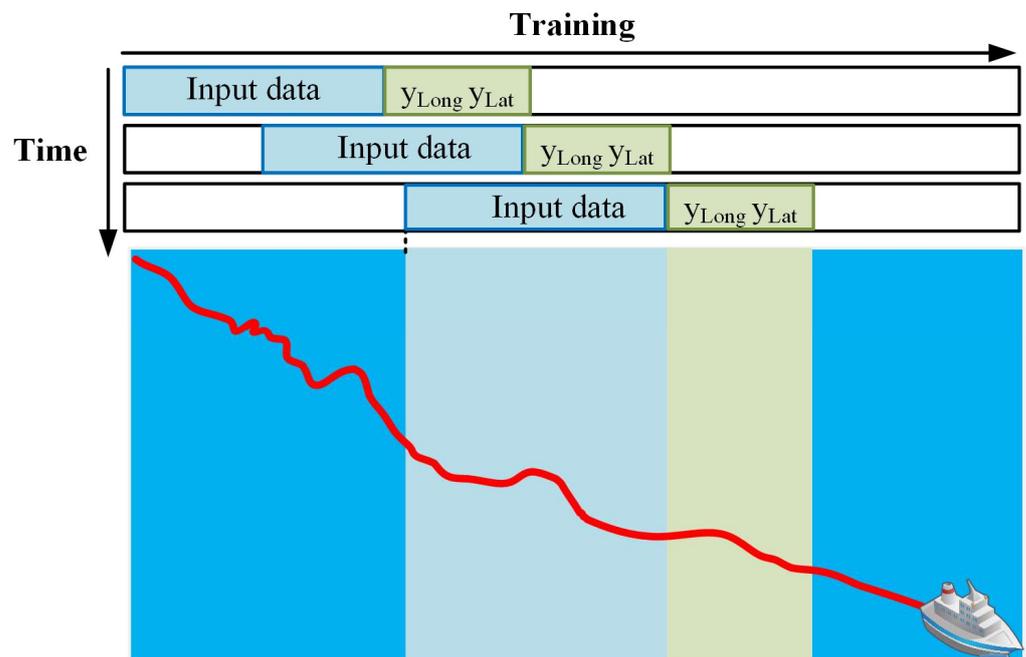
$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (19)$$

Figure 5 depicts the application of the sliding window method for data training in this experiment [34]. Through the sliding window, the final ship trajectory data for the current slider enable forecasting to be performed via sliding each unit until all of the data in the training set are traversed, thereby completing a training epoch.

#### 3.2. Hyper Parameters Setting

Parameter design plays a crucial role for recurrent neural networks. In this experiment, the Adam optimizer was chosen, which combines the benefits of the gradient descent algorithm with adaptive learning rate and the momentum gradient descent algorithm to not only adapt to sparse gradients but also mitigate the issue of gradient oscillation. The initial network learning rate was set to 0.002, and each training cycle's learning rate was reduced adaptively [35]. The model performs better if the learning rate is gradually

reduced. The training epoch ends when the loss value approaches 0, and the maximum training number is reached. Then, the MHA-BiRU was used to compare and validate the fundamental parameters, the number of units in the hidden layer and batch size in the network, in order to determine the optimal parameter combination. The selectable range of the former was {16, 32, 64, 128}, and the latter was {16, 32, 64, 128}. In particular, when the value of units in the hidden layer is large, the complexity of the model increases, and it is prone to overfitting, but when the value is small, the nonlinear fitting ability may be weakened. Additionally, when batch sizes are either oversized or undersized, the number of errors generated is excessive [36]. After several groups of hyper parameter selection comparative experiments based on MHA-BiGRU, the parameters selected in this experiment are as shown in Table 1. In order to effectively compare the prediction effect, the choice of other methods' parameters is consistent with the MHA-BiGRU.



**Figure 5.** Model training over time, where the blue block represents the input training data and the green one represents the prediction result for each batch of input.

**Table 1.** The hyper parameters selection of comparative experiments.

Model	Parameter Name	Optimal Parameters
MHA-BiGRU	Hidden_size	32
	Batch_size	64
	Activation function	Linear
	loss function	Mean Squared Error
	Learning rate	0.0001

### 3.3. Evaluation Index

This experiment selected mean squared error (MSE) as the loss function of the proposed model, allowing the overall error degree to be quantified. The more robust the model, the smaller the loss function. MAE and RMSE served as the evaluation indices for each method. The smaller the number, the more closely the predicted value matches the

actual value, and thus, the more accurate the prediction. These methods of computation are illustrated below.

$$loss(x, y) = \frac{1}{N} \sum_{i=1}^N |x - y|^2 \tag{20}$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \tag{21}$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |(y_i - \hat{y}_i)| \tag{22}$$

### 3.4. Results of Comparative Experiments

#### 3.4.1. Comparative Experiment Based on the Bidirectional RNN Structure

BiGRU, BiLSTM, GRU, and LSTM were used in the bidirectional RNN structure comparison experiment to compare prediction results. The model with the lower loss on new data has superior generalization performance and can alleviate the issue of overfitting. The prediction accuracy of these methods is illustrated below: BiGRU is superior to GRU, and BiLSTM is superior to LSTM, as shown in Figure 6.

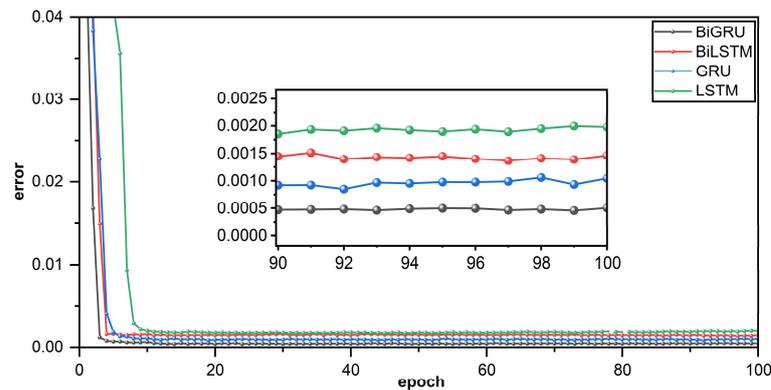


Figure 6. Comparative experiment loss value based on the bidirectional RNN structure.

#### 3.4.2. Comparative Experiment Based on the MHA Mechanism

Figure 7 vividly show comparative experiment results based on the MHA mechanism. The prediction accuracies of these algorithms are compared as follows: MHA-BiGRU > BiGRU, MHA-BiLSTM > BiLSTM, MHA-GRU > GRU, MHA-LSTM > LSTM. The results of these comparative experiments show that the combination of the MHA mechanism with both RNN and BiRNN is preferable to that without the MHA mechanism.

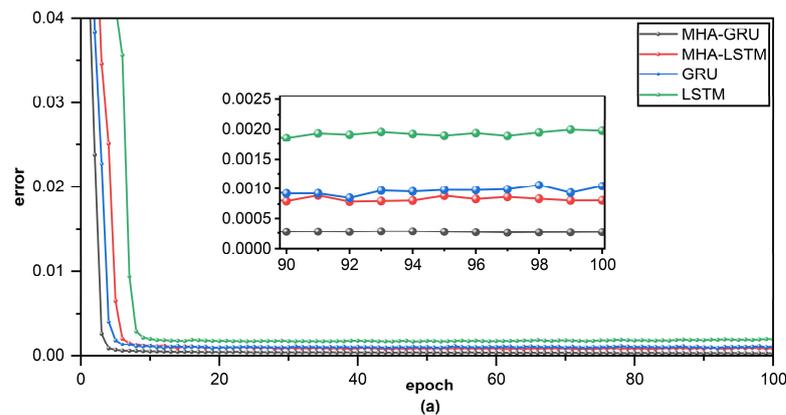
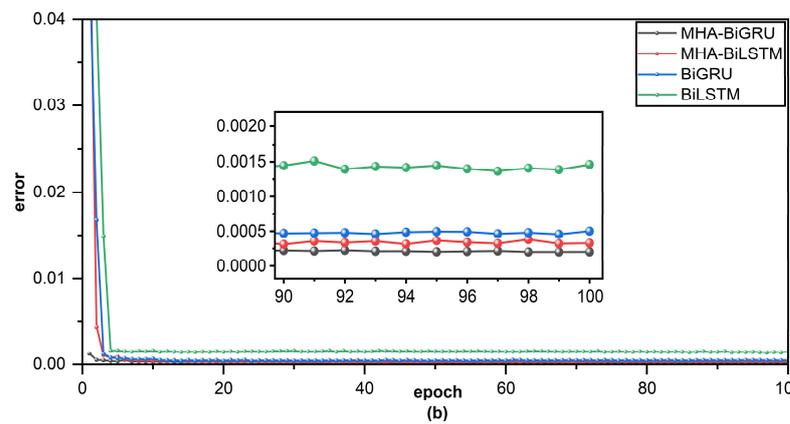


Figure 7. Cont.

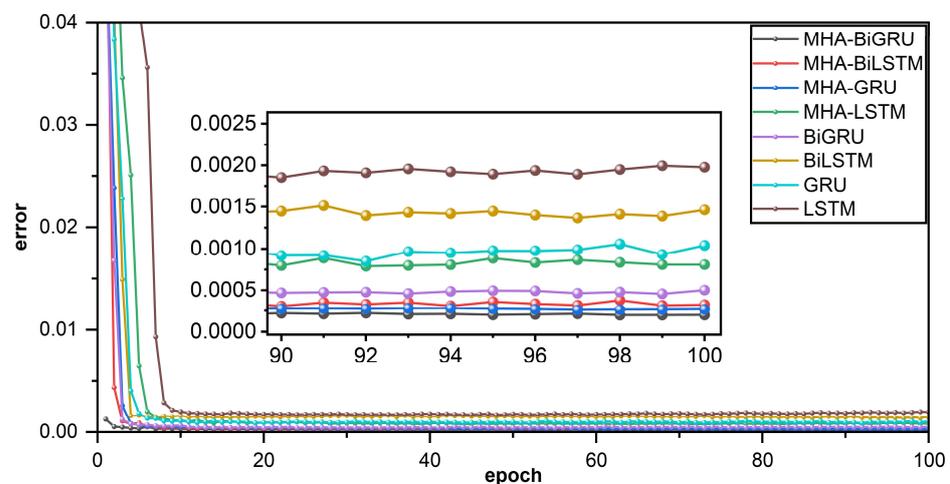


**Figure 7.** Comparative experiment loss value based on the MHA mechanism: (a) the combination of the MHA mechanism with RNN (b) the combination of the MHA mechanism with bidirectional RNN.

### 3.4.3. Overall Comparative Experiments

As illustrated in Figures 8–11 and Table 2, the prediction results of all eight methods on the test dataset perform well, and the ship trajectory prediction results are relatively accurate, demonstrating not only that the models were not fitted but also that all of these models can better deal with the problem of track prediction. The model with the lower loss value has the highest accuracy. As a result, the prediction accuracies of these algorithms are listed in the following order: MHA-BiGRU > MHA-GRU > MHA-BiLSTM > BiGRU > MHA-LSTM > GRU > BiLSTM > LSTM. These results show that GRU can outperform LSTM in this comparison scenario, both in terms of efficiency and accuracy, regardless of whether the model is combined with a two-way structure, the MHA mechanism, or neither. In addition, MHA-GRU is superior to BiGRU, and MHA-LSTM is superior to BiLSTM, which may indicate that the MHA mechanism contributes more to the model’s accuracy and robustness than the bidirectional structure.

Finally, and most importantly, the MHA-BiGRU model has the lowest loss, RMSE, and MAE values, indicating that its applicability, accuracy, and validity are superior to those of other comparison experiments. Additionally, as shown in Table 2 and Figure 10, MHA-BiGRU comes closest to the original ship’s path, which shows that this model’s prediction is the most accurate in this comparison.



**Figure 8.** Loss value of overall comparative experiments.

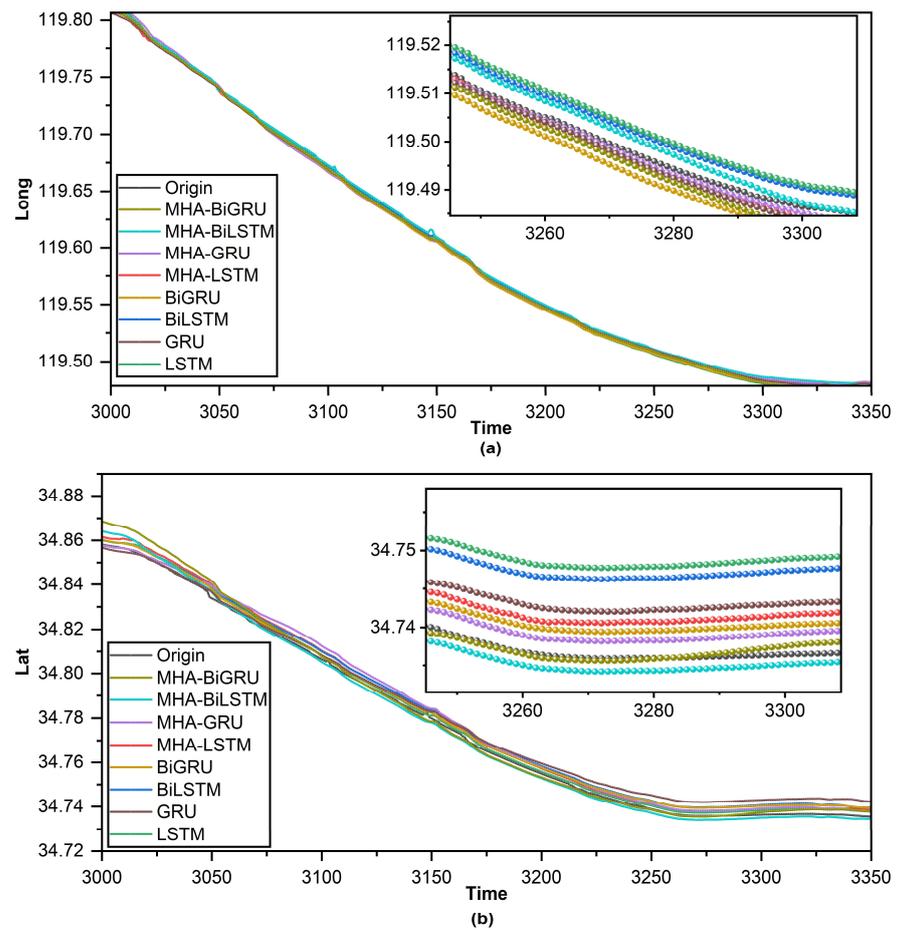


Figure 9. The comparative prediction results: (a) the predicted longitude (b) the predicted latitude.

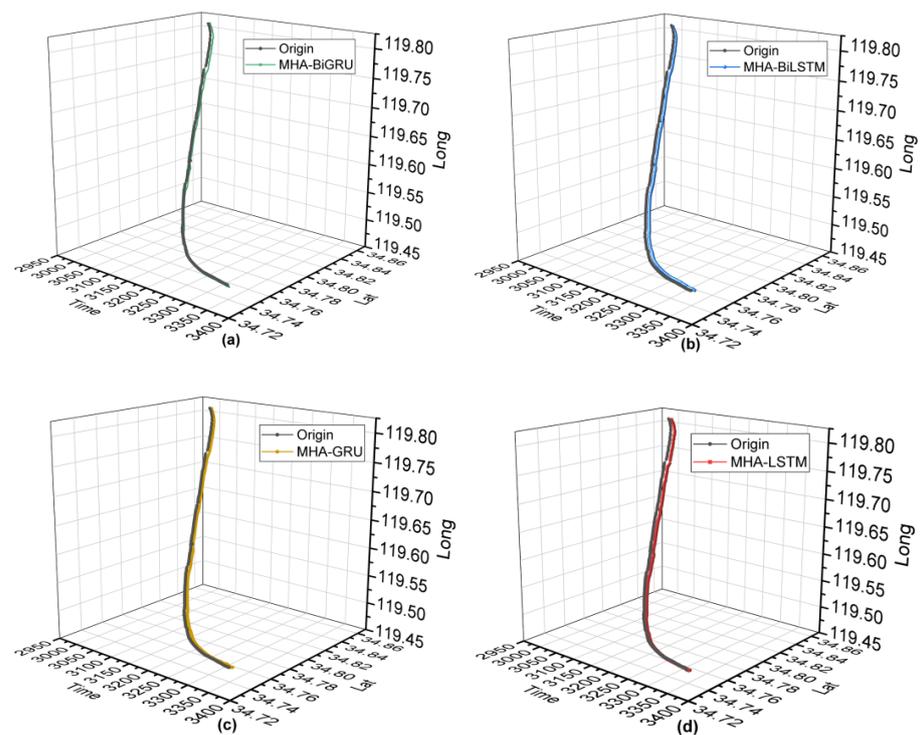
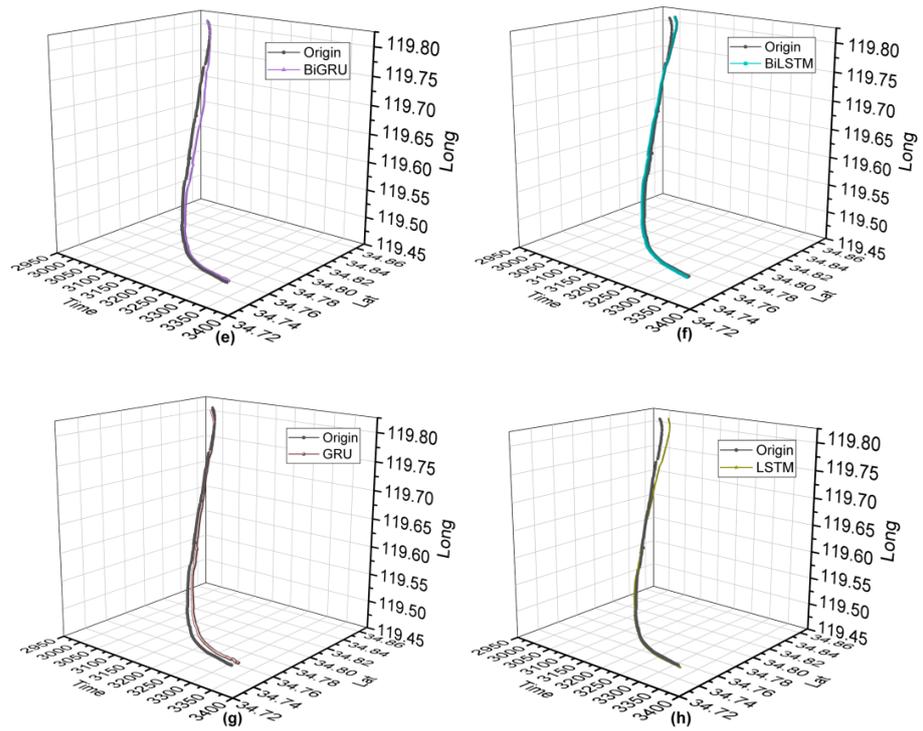
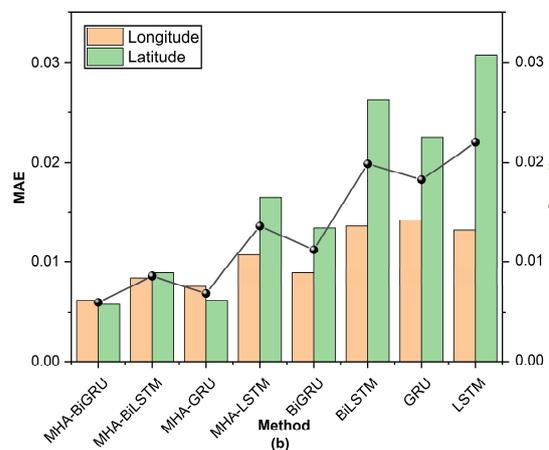
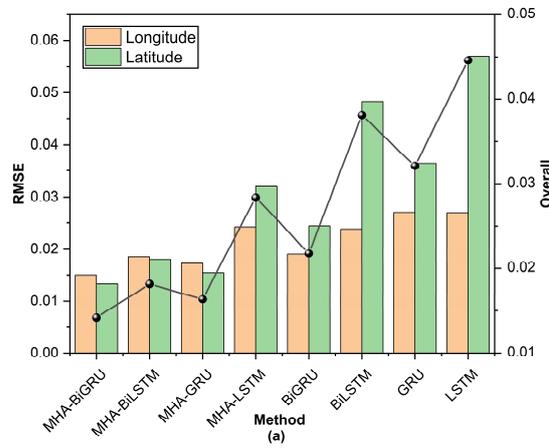


Figure 10. Cont.



**Figure 10.** Ship behavior prediction: (a) MHA-BiGRU (b) MHA-BiLSTM (c) MHA-GRU (d) MHA-LSTM (e) BiGRU (f) BiLSTM (g) GRU (h) LSTM.



**Figure 11.** RMSE and MAE value comparison among all models: (a) RMSE (b) MAE.

**Table 2.** RMSE and MAE value comparison among all models.

Method	RMSE			MAE		
	Longitude	Latitude	Overall	Longitude	Latitude	Overall
MHA-BiGRU	0.0150	0.0133	0.0142	0.0062	0.0058	0.0060
MHA-BiLSTM	0.0185	0.0180	0.0182	0.0084	0.0089	0.0086
MHA-GRU	0.0172	0.0154	0.0164	0.0076	0.0062	0.0069
MHA-LSTM	0.0242	0.0320	0.0284	0.0107	0.0165	0.0136
BiGRU	0.0190	0.0244	0.0218	0.0089	0.0134	0.0112
BiLSTM	0.0238	0.0483	0.0381	0.0136	0.0263	0.0199
GRU	0.0270	0.0364	0.0321	0.0142	0.0225	0.0183
LSTM	0.0269	0.0570	0.0446	0.0132	0.0308	0.0220

#### 4. Discussion

By gradually demonstrating the benefits of bidirectional RNN, multi-head attention mechanism, and GRU, the comparative experiment results demonstrated that MHA-BiGRU outperforms other models in terms of effectiveness and accuracy of ship trajectory prediction.

##### 4.1. The Contribution of the MHA-BiGRU Model

The LSTM and GRU, excellent variants of RNN, have a gate structure that not only preserves long-term sequence information but also filters and modifies ship track historical data for enhanced time series prediction. Additionally, in comparison with LSTM, the prediction task with GRU can be accomplished with fewer model parameters, but it can perform similarly to LSTM [17–19]. This experiment finds that GRU can outperform LSTM in this comparison scenario, both in terms of efficiency and accuracy, regardless of whether the model is combined with a two-way structure, the MHA mechanism, or neither. Although GRU outperforms LSTM in this experiment, there is no final conclusion on which is better or worse and which must be chosen based on specific tasks and datasets.

Gao et al. and Siami-Namini et al. proved that the use of a bidirectional structure to improve contextual relevance based on the RNN method improves the accuracy of ship trajectory prediction compared to RNN alone [23,24]. Whether combined with LSTM or GRU, this experiment further demonstrated that the bidirectional structure can improve the accuracy of ship trajectory prediction. As a result, this finding thoroughly demonstrates that the bidirectional RNN structure can simulate the prospective relationship between past ship trajectory status information and future ship trajectory status information with current state in order to increase prediction accuracy.

The MHA mechanism is frequently employed in image recognition and automatic translation. It was combined with a recurrent neural network in this experiment, from which significant conclusions are drawn. The most important results of the comparative experiments demonstrate the advantage of the MHA mechanism in combination with RNN and BiRNN. Additionally, when compared to bidirectional structures, the MHA mechanism contributes significantly more to the model's accuracy and robustness. Thus, the MHA mechanism not only calculates the correlation between the characteristics of AIS information but also actively learns cross-time synchronization between the hidden layers of the output and input sequences, and it assigns different weights to the result based on the input criterion, thereby improving the overall model's accuracy and robustness.

Overall, the most crucial advantage of MHA-BiGRU is that it enables the preservation of long-term sequence information, filters and modifies ship track historical data for improved time series prediction, models the potential relationship between historical and future ship trajectory status information and the current state via a bidirectional structure, and highlights critical ship trajectory prediction information in AIS characteristics and time series dimension via an MHA mechanism.

#### 4.2. The Limitations and Future Development

Experiments indicate that the MHA-BiGRU model has high prediction accuracy under normal navigation conditions, as well as good applicability and track prediction reliability. However, as the navigational status of each ship changes over time, the navigational status of other ships will have varying effects on the future course of the ship in inquiry. Additionally, the bad weather will impact the ship's navigation, leading to an abnormal ship trajectory. Moreover, in addition to using AIS data for ship trajectory prediction, it can also be supplemented with other system data, such as the radar system, to further increase the model's accuracy. Thus, in order to further investigate whether the model can correct and avoid ship collisions under abnormal conditions, it is necessary to combine other ship spatial information and bad weather information to verify the model's performance under abnormal circumstances.

#### 5. Conclusions

To improve the performance of vessel track prediction compared with some existing approaches, a high-precision method, which combines the advantages of bidirectional RNN, multi-head attention mechanism, and GRU is proposed in this paper. Through comparison experiments, the following conclusions can be drawn:

1. GRU is a great variation of LSTM, in that it requires only an update and reset gate to govern the flow of information. As a result of its smaller parameters compared to LSTM, it is extremely easy to train and outperforms in terms of efficiency and accuracy in this experiment.
2. Bidirectional RNN structure enables the modeling of the potential association between previous and future ship trajectory status information with the current state, so as to increase prediction accuracy.
3. The MHA mechanism not only calculates the correlation between AIS information characteristics but also actively learns cross-time synchronization between the hidden layers of the output and input sequences, and it assigns different weights to the result based on the input criterion, improving the accuracy and robustness of the overall model.
4. In general, all evaluation indicators show that the prediction accuracy of MHA-BiGRU is higher than that of other comparative experiments, implying that the proposed model can effectively improve ship trajectory prediction performance.

In the future, it will be necessary to combine other ship spatial information and bad weather information to verify the model's performance under abnormal conditions in order to further investigate whether the model can correct and avoid ship collisions in extreme circumstances.

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

**Funding:** This research was supported by the Fundamental Research Funds for the Central Public Welfare Research Institutes, grant number TKS20220306 and TKS20220304.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets analyzed or generated in this study are available from the corresponding author upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Perera, L.P.; Oliveira, P.; Soares, C.G. Maritime traffic monitoring based on vessel detection, tracking, state estimation, and trajectory prediction. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 1188–1200. [\[CrossRef\]](#)
2. Jiang, D.; Cheng, Z.Y.; Xue, J.; van Gelder, P.A. Towards a probabilistic model for estimation of grounding accidents in fluctuating backwater zone of the Three Gorges Reservoir. *Reliab. Eng. Syst. Saf.* **2021**, *205*, 107239. [\[CrossRef\]](#)
3. Praetorius, G. Vessel Traffic Service (VTS): A Maritime Information Service or Traffic Control System? Understanding Everyday Performance and Resilience in a Socio-Technical System under Change. Ph.D. Thesis, Chalmers University of Technology, Göteborg, Sweden, 2014.
4. Felski, A.; Jaskólski, K.; Banyś, P. Comprehensive assessment of automatic identification system (AIS) data application to anti-collision manoeuvring. *J. Navig.* **2015**, *68*, 697–717. [\[CrossRef\]](#)
5. Yang, R.; Shi, G.Y.; Li, W.F. Ship track prediction model based on automatic identification system data and bidirectional cyclic neural network. In Proceedings of the 2021 4th International Symposium on Traffic Transportation and Civil Architecture (ISTTCA), Suzhou, China, 12 November 2021. [\[CrossRef\]](#)
6. Wu, B.; Tang, Y.H.; Yan, X.P.; Soares, C.G. Bayesian Network modelling for safety management of electric vehicles transported in RoPax ships. *Reliab. Eng. Syst. Saf.* **2021**, *209*, 107466. [\[CrossRef\]](#)
7. Gao, D.W.; Zhu, Y.S.; Zhang, J.F.; He, Y.K.; Yan, K.; Yan, B.R. A novel MP-LSTM method for ship trajectory prediction based on AIS data. *Ocean Eng.* **2021**, *228*, 108956. [\[CrossRef\]](#)
8. Anderson, S.; Barfoot, T.D.; Tong, C.H.; Särkkä, S. Batch nonlinear continuous-time trajectory estimation as exactly sparse Gaussian process regression. *Auton. Robot.* **2015**, *39*, 221–238. [\[CrossRef\]](#)
9. Rong, H.; Teixeira, A.P.; Soares, C.G. Ship trajectory uncertainty prediction based on a Gaussian Process model. *Ocean Eng.* **2019**, *182*, 499–511. [\[CrossRef\]](#)
10. Jiang, B.; Guan, J.; Zhou, W.; Chen, X. Vessel trajectory prediction algorithm based on polynomial fitting kalman filtering. *J. Signal Processing* **2019**, *35*, 741–746. [\[CrossRef\]](#)
11. De Masi, G.; Gaggiotti, F.; Bruschi, R.; Venturi, M. Ship motion prediction by radial basis neural networks. In Proceedings of the 2011 IEEE Workshop On Hybrid Intelligent Models And Applications, Paris, France, 11–15 April 2011. [\[CrossRef\]](#)
12. Zhou, H.; Chen, Y.; Zhang, S. Ship trajectory prediction based on BP Neural Network. *J. Artif. Intell.* **2019**, *1*, 29–36. [\[CrossRef\]](#)
13. Liu, J.; Shi, G.; Zhu, K. Vessel trajectory prediction model based on ais sensor data and adaptive chaos differential evolution support vector regression (ACDE-SVR). *Appl. Sci.* **2019**, *9*, 2983. [\[CrossRef\]](#)
14. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [\[CrossRef\]](#)
15. Potter, C. RNN based MIMO channel prediction. In *Differential Evolution in Electromagnetics. Evolutionary Learning and Optimization*; Qing, A., Lee, C.K., Eds.; Springer: Berlin/Heidelberg, Germany, 2020; Volume 4, pp. 177–206. [\[CrossRef\]](#)
16. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
17. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078. [\[CrossRef\]](#)
19. del Aguila Ferrandis, J.; Triantafyllou, M.; Chryssostomidis, C.; Karniadakis, G. Learning functionals via LSTM neural networks for predicting vessel dynamics in extreme sea states. *arXiv* **2019**, arXiv:1912.13382. [\[CrossRef\]](#)
20. Agarap, A.F.M. A neural network architecture combining gated recurrent unit (GRU) and support vector machine (SVM) for intrusion detection in network traffic data. In Proceedings of the 2018 10th International Conference on Machine Learning and Computing, Association for Computing Machinery, New York, NY, USA, 26–28 February 2018. [\[CrossRef\]](#)
21. Agarap, A.F.; Grafilon, P. Statistical analysis on e-commerce reviews, with sentiments classification using bidirectional recurrent neural network (RNN). *arXiv* **2018**, arXiv:1805.03687. [\[CrossRef\]](#)
22. Gao, M.; Shi, G.; Li, S. Online prediction of ship behavior with automatic identification system sensor data using bidirectional long short-term memory recurrent neural network. *Sensors* **2018**, *18*, 4211. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The performance of LSTM and BiLSTM in forecasting time series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019. [\[CrossRef\]](#)
24. Stateczny, A.; Błaszczak-Bak, W.; Sobieraj-Żłobińska, A.; Motyl, W.; Wisniewska, M. Methodology for processing of 3D multibeam sonar big data for comparative navigation. *Remote Sens.* **2019**, *11*, 2245. [\[CrossRef\]](#)
25. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
26. Cheng, X.; Li, G.; Ellefsen, A.L.; Chen, S.; Hildre, H.P.; Zhang, H. A novel densely connected convolutional neural network for sea-state estimation using ship motion data. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 5984–5993. [\[CrossRef\]](#)
27. Tu, E.; Zhang, G.; Rachmawati, L.; Rajabally, E.; Huang, G.B. Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1559–1582. [\[CrossRef\]](#)
28. Sang, L.Z.; Wall, A.; Mao, Z.; Yan, X.P.; Wang, J. A novel method for restoring the trajectory of the inland waterway ship by using AIS data. *Ocean. Eng.* **2015**, *110*, 183–194. [\[CrossRef\]](#)

29. Suo, Y.; Chen, W.; Claramunt, C.; Yang, S. A ship trajectory prediction framework based on a recurrent neural network. *Sensors* **2020**, *20*, 5133. [[CrossRef](#)] [[PubMed](#)]
30. Zhang, G.; Tan, F.; Wu, Y. Ship motion attitude prediction based on an adaptive dynamic particle swarm optimization algorithm and bidirectional LSTM neural network. *IEEE Access* **2020**, *8*, 90087–90098. [[CrossRef](#)]
31. Wang, C.; Ren, H.; Li, H. Vessel trajectory prediction based on AIS data and bidirectional GRU. In Proceedings of the 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), Chongqing, China, 10–12 July 2020. [[CrossRef](#)]
32. Nguyen, D.; Fablet, R. TrAISformer-A generative transformer for AIS trajectory prediction. *arXiv* **2019**, arXiv:2109.03958. [[CrossRef](#)]
33. Sun, L.; Zhou, W. Vessel motion statistical learning based on stored ais data and its application to trajectory prediction. In Proceedings of the 2017 5th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2017), Beijing, China, 25–26 March 2017.
34. Zhang, L.; Zhang, J.; Niu, J.; Wu, Q.M.; Li, G. Track prediction for HF radar vessels submerged in strong clutter based on mscnn fusion with gru-am and ar model. *Remote Sens.* **2021**, *13*, 2164. [[CrossRef](#)]
35. Kingma, D.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980v9. [[CrossRef](#)]
36. Hu, Y.K.; Xia, W.; Hu, X.X.; Sun, H.Q.; Wang, Y.H. Vessel trajectory prediction based on recurrent neural network. *Syst. Eng. Electron.* **2020**, *42*, 871–877.