Ship Collision Risk Assessment Based on Collision Detection Algorithm

DONGDONG LIU AND GUOYOU SHI
Navigation College, Dalian Maritime University, Dalian 116026, China
Key Laboratory of Navigation Safety Guarantee of Liaoning Province, Navigation College, Dalian Maritime University, Dalian 116026, China
Corresponding author: Guoyou Shi (sgydmu@163.com)

This work was supported by the Fundamental Research Funds for the Central Universities under Grant 3132020134 and Grant 3132020139.

ABSTRACT Ningbo Zhoushan port handled 1.08 billion tons cargoes in 2018 which is considered as the one of biggest ports in the world. There are more than 1 000 ships enter or depart the port per day. Therefore, it is of importance to assess the collision risk for ships passing through the harbor area. In this article, a novel approach is initially proposed to assess ship collision risk in the harbor area based on collision detection technology of ship domain using automatic identified system (AIS) data. This study aims to build a unified framework of collision risk assessment which does not need to build different models in accordance with the ship domain we selected. To clean the historical motion data of ships, a method for anomaly detection of ship static information based on autoencoder (AE) is proposed. Based on the above proposed method, the ship collision frequency can be estimated, besides, the risk area can also be determined. The results obtained from the method could provide a reference on furthering enhance the navigational safety for the Maritime and Port Authority.

INDEX TERMS Waterway transportation, risk assessment, collision detection, ship domain, safety.

I. INTRODUCTION
The Ningbo Zhoushan port is one of the biggest ports all over the world, it transforms 1.08 billion tons of cargoes in 2018. The traffic density of the ship is high, where more than 1 000 ships pass through the port area daily. However, it is not wide enough for passing through due to it lies in the Zhoushan archipelago. With the increase in the volume of goods, traffic density is expected to increase continually. It is obvious the high traffic density would result in more accidents occurrence. Although the traffic separation scheme (TSS) can assure the navigation safety in the archipelago, however, the Maritime and Port Authority is still not known about the risk in the waters area precisely. Risk assessment can qualify to evaluate the collision risk and help the policymakers making effective strategies decision.

Many studies [1] – [4] about the navigational risk assessment had been proposed for improving maritime safety. Qu [1] assessed the risk with the degree of acceleration and deceleration, speed dispersion, and the number of ship domains overlapped. Weng [2], [3] applied the ship domain to build the collision risk frequency model, with which the ship collision risk in Singapore strait is assessed. In accordance with the critical situation theory, Chai [4] constructed the dynamic ship domain considering the ship type, ship scale, relative speed, encounter situation, and the location environment. Szlapczynski [5] derived the unified measure of collision risk from the concept of ship safety domain. Goerlandt F. and Kujala P. [6] utilized the ship collision probability model for traffic simulation.

In the field of research ship accident frequency, Fujii and MacDuff [7], [8] defined the occurrence frequency is equal to the product between the number of dangerous encounters and a causation probability. For evaluating the occurrence frequency, there are have two major categories, which are collision diameter (CD) theory and critical situation theory, respectively. In Europe’s waters, the CD model [9] was proposed by Pederson. Pederson model divided the ship collision into across collision model and straight collision model, and then evaluates and obtains the collision occurrence frequency respectively. Jakub [10], [11] replaced the collision diameter with the minimum collision distance (MDTC), divided the encounter situations into three classes, which are collision model, the encounter collision model, and across collision model, respectively. Cheekuang [12] assessed the collision risk in different encounter situations using AIS data.
Fowler and Sorgard [13] set the critical situation when the cross distance between ships is less than 0.5 nautical miles. However, the above-mentioned risk assessment methods using the ship domain are generally based on fixed diameter. Therefore, the collision risk assessment model would be different from the certain ship domain selected by the researcher. There is no existing study to unify the method for calculating collision frequency. Therefore, it is important to propose a method for collision risk assessment which is applicable to any ship domain.

AIS data has been widely applied to assess ship collision risk [16]. However, in the AIS history records, there are have many abnormal data caused at the receiving stage, transmitting stage, or sending stage. Due to AIS data receiving errors, human input errors, and sensor errors et al, some of the data will have missing fields or correlations that do not match, which is called abnormal data. Before conducting research, the AIS data needs to be pre-processed to clear the abnormal data in the data. Aims at tackling the anomaly data from the AIS history records, the pre-processing methods are proposed, which combine two parts: static information records pre-processing model and dynamic information records pre-processing model. In this article, the research data is from AIS which contains more than seven million records including the unique identification in terms of MMSI, LAT, LON, SOG, and COG at any reporting time (every 1–10 min) for every ship in the harbor area from Jan. 1st to Jan. 7th, 2018.

II. AIS HISTORY RECORDS PRE-PROCESSING

AIS is an automatic tracking system used on ships and by Vessel Traffic Services (VTS) for identifying and locating ships by electronically exchanging data with other nearby ships and VTS stations. The specific requirements of The International Maritime Organization’s (IMO) international convention for the Safety of Life At Sea (SOLAS) for the installation of an AIS are that all international navigation vessels of 300 gross tonnages and above, and non-international navigation vessels of 500 gross tonnages and above, and all passenger ships [14]. The system may automatically broadcast the ship’s name along with Maritime Mobile Service Identity number (MMSI), the latitude of position (LAT), the longitude of position (LON), speed over ground (SOG), course over ground (COG), Unix timestamp (UTC), and other navigation-related configurations in the form of an ASCII stream which was standardized by International Telecommunications Union [15]. The broadcast interval is from 2s to 6min, depending on their navigation status. The faster the speed of the ship, the shorter the broadcast interval, and vice versa.

Aim to construct the unified framework of ship collision risk assessment based on the different types of ship domain. In this article, there are have three parts in the unified framework. The first step is to pre-process the static information of track records, which was stored in the HeidiSQL database. The second step is to analyze the shape of most ship domains at present and solve the collision detection problem between different ship domains. The third step is to build a unified framework of collision risk assessment which does not need to build different models in accordance with the ship domain we selected.

![FIGURE 1. Marine traffic flow in Ningbo Zhoushan port area.](image1)

**FIGURE 1.** Marine traffic flow in Ningbo Zhoushan port area.

**A. STATIC INFORMATION PRE-PROCESSING**

In the ship static information anomaly detection research field, there are focusing on the stage of correlation analysis and data cleaning based on the defined rule [14]. Aiming at tackling the problem: the lack of anomaly detection method for ship static information data and the low data utilization rate, the static information anomaly detection method based on AE is proposed. AE is a special neural network including the input layer, hidden layer, and output layer. The special feature is that the number of features of the input layer and the output layer is the same, and the input and output are equal [17]. The structure of AE can be divided into encoder part and decoder part. The data processing is mainly divided into two stages, one is the encoding stage from the input layer to the hidden layer, and the other is the decoding stage from the hidden layer to the output layer. The weight of encoding and decoding in Figure 2 is shared, and the weight of the decoding stage is the transposition of the encoding weight. To make the input and output as similar as possible, after each output, it is necessary to carry out error backpropagation and continuously optimize the training parameters. The training process of AE is to minimizing the reconstruction error between the input vector and the reconstructed vector. In this article, we use the mean square error (MSE) formula:

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{x} - x)^2 \]

where \( x \) is the input vector, \( \hat{x} \) is the reconstructed vector, and \( n \) is the number of input samples.

![FIGURE 2. Structure of AE.](image2)

**FIGURE 2.** Structure of AE.
The static data is trained by the AE, and the reconstruction loss function is obtained to form an abnormal data detection model. The upper limit of the training sample loss function is determined by kernel density estimation, and the distribution characteristics are extracted by AE training [18]. The abnormal ship static information is mainly represented by the lack (NULL) or zero of certain information moulded draught (Draught), deadweight tonnage (DWT), length overall (LOA), beam (BM), and moulded depth (Depth) that we need to use in the representation of the ship’s static information matrix, we extract the static information moulded draught (Draught), deadweight tonnage (DWT), length overall (LOA), beam (BM), and moulded depth (Depth) that we need to use in the representation of the ship domain. For normal data that meet the normal epidemic distribution, the distribution characteristics are extracted by AE training [18]. The abnormal ship static information is sparse in data, that is, most of it is sparse data [19], which is mainly represented by the lack (NULL) or zero of certain element information data. In view of this, the data in the normal state are used as the upper limit of the loss function. If there is data in the test data that exceeds the upper limit of the loss function, it is determined as a singular point or a missing point, which is abnormal data. The process is shown as follow.

- First step: training the normal data by AE and obtaining the weight matrix of the encoder and the bias vector.
- Second step: evaluating the difference between the test data and the constructed data that serve as the loss error.
- Third step: determining the threshold of the AE model loss $R_e$ by evaluating the cumulative distribution function (CDF).
- Fourth step: testing the test static information sets by using the trained AE model and obtaining the construction error $R_e'$ of the testing data.
- Fifth step: comparing the relationship between $R_e'$ and $R_e$, determining the type of testing data: (1) if $R_e' \leq R_e$, the testing data is normal static information; (2) if $R_e' < R_e$, the testing data is abnormal static information.

### B. DYNAMIC INFORMATION RECORDS PRE-PROCESSING

Aim to assess the collision risk, some dynamic information in AIS records need to be extracted. In general, the factors considered in the ship domain are MMSI, LON, LAT, SOG, and COG. Although the reporting time and the ship records provided by AIS are highly reliable, inevitably there is an error occurring at the collecting end, the transmission end, or the receiving end. The noises would have a significant influence on the ship collision risk assessment. On the basis of Newton’s laws of motion, the speed equals to the ratio of distance and time. Therefore, the position and speed can be checked using a reasonable range of position and speed.

To tackle these inaccurate SOG, LON, and LAT data, the data can be checked using a reasonable range of position and speed. In order to minimize the effect from those inaccurate records, the pre-processing procedures are proposed as follows.

- First step: tackling the irrational speed.

Ordinarily, the mean speed $\bar{V}_{i,T_j}$ for a particular ship $i$ at a specific time interval $[T_j, T_{j+1}]$ can be estimated by

$$
\bar{V}_{i,T_j} = \arcsin \left\{ \frac{1}{2} \sin^2 \left[ \frac{(y_{i,T_{j+1}} - y_{i,T_j})}{2} + \cos(y_{i,T_{j+1}}) \right] \right\}^{1/2} \cos(y_{i,T_j}) \sin^2 \left[ \frac{(x_{i,T_{j+1}} - x_{i,T_j})}{2} \right]^{1/2} R/(T_{j+1} - T_j)
$$

where $x_{i,T_j}$ is the LON (degree) of ship $i$ at time $T_j$(hour); $y_{i,T_j}$ is the LAT (degree) of ship $i$ at time $T_j$(hour);

In accordance with Newton’s laws of motion, the mean speed of ship should be satisfied with the situation

$$
V_{i,T_j} - d_i (T_{j+1} - T_j) \leq \bar{V}_{i,T_j} \leq V_{i,T_j} + a_i (T_{j+1} - T_j)
$$

where $V_{i,T_j}$ is the SOG (km/h) of ship $i$ at time $T_j$(hour); $d_i$ is the maximum deceleration of ship $i$; $a_i$ is the maximum acceleration of ship $i$.

On the basis of Eq. (1) and (2), it can be checked whether the SOG at any point is within a reasonable range. If it is not, it is regarded as abnormal.
- Second step: tackling the irrational position data.

Refer to the same procedure, the position in terms of latitude and longitude can be calculated by

\[
\begin{align*}
x_{i,T_{j+1}} &= x_{i,T_j} + V_{i}^x \times (T_{j+1} - T_j) \\
y_{i,T_{j+1}} &= y_{i,T_j} + V_{i}^y \times (T_{j+1} - T_j)
\end{align*}
\] (3)

where \(x_{i,T_{j+1}}\) is the estimated LON (degree) of ship \(i\) at time \(T_{j+1}\) (hour); \(y_{i,T_{j+1}}\) is the estimated LAT (degree) of ship \(i\) at time \(T_{j+1}\) (hour); \(V_{i}^x\) is the decomposition speed of ship \(i\) at time \(T_j\) (hour) in longitude axis; \(V_{i}^y\) is the decomposition speed of ship \(i\) at time \(T_j\) (hour) in latitude axis. The estimated position should satisfy the situation

\[
2 \left[ (x_{i,T_{j+1}} - x_{i,T_j})^2 - (y_{i,T_{j+1}} - y_{i,T_j})^2 \right]^{1/2} \leq a_i (T_{j+1} - T_j)^2
\] (4)

On the basis of Eq. (3)-(4), it can be checked whether the estimated position at any point is within a reasonable range. If it is not, it is regarded as abnormal.

III. SHIP COLLISION RISK ASSESSMENT MODEL

A. SHIP SAFETY DOMAIN ANALYSIS

The ship safety domain refers to the safety distance of ships, and the safety distance of each model is different in all directions. The distance is usually related to the navigation environment, ship scale, navigation status et al. [21]. The ship safety domain is a term widely used in the fields of collision avoidance research [21], collision risk analysis [22],[23], and water safety analysis [25]. Scholars from various countries have conducted many detailed studies in the ship safety domain. Szlapczynski R. [26] summarized the models and applications in the field of ship safety domain in the literature. Traditionally, there are two definitions about ship domain, which were defined by Fujii [8] in 1971, and Goodwin [27] in 1975 respectively. Fujii [8] defined the ship domain is a two-dimensional area surrounding a ship which other ships must avoid. Goodwin [27] defined the concept of ship domain as the surrounding effective waters which the navigators of a ship want to keep clear of other ships or stationary obstacles. Goodwin [27] constructed the ship domain model of open waters that is composed of three different sectors. Davis [28], [29] made a more practical circular area of ship by smoothing the boundaries. C. J [30] thought there are existing a corresponding function between the speed and the scale of the ship. N. Wang [31] developed an analytical domain model which is flexible and can be controlled by four radii and index parameter \(k\). Many scholars have established many ship safety domain models on different scales to carry out a risk assessment study on the specific waters as shown in Figure 4.

Through analysis the above-mentioned ship domain, we can find that the ship domain is related to the ship type, speed, scale, and traffic environment. The shape of the ship domain proposed above is different, there are existing ellipse, polygon, and the combination of sectors. Therefore, there is not existing a unified standard domain model to assess the collision risk. The previous studies just can be applied to a specific ship domain model. On that basis, this article designed a method that can determining if there is a collision between ships by detecting the ship domain of ships whether are overlapped or not.

B. SHIP COLLISION DETECTION UNIFIED FRAMEWORK

The concept of conflict derives from the aviation field. Aviation safety is assessed by evaluating the probability of conflict. In reference to the collision theory of the aviation model and ship domain, ship conflict can be defined as one ship’s domain invaded by another ship. In reference to the work of MacDuff [7] and Mou [16] et al., in this article, the following formula is applied to calculating the ship collision frequency.

\[
f = Np_c
\] (5)

where:

\(f\) is the ship collision frequency, which represents the number of accident that occurs in a specific waters within a specific period.

\(N\) is the number of ship conflicts, which represents the number of potentially dangerous situations that occur in a specific waters within a specific period.

\(p_c\) is the casualty probability, which is the probability of failing to avoid a collision with the given-ship. It is a conditional probability that is determined based on casualty databases by statistical methods.

To calculate the number of ship conflicts, a ship conflict is defined as a critical situation where a ship invades another ship’s domain in a trajectory segment. We divided the trajectory into many trajectory segments using the partition rule of navigation water. The navigation water was divided into ten parts on the basis of upstream traffic (U1-U10), downstream traffic (D1-D10), and three alert areas (A1-A3) as shown in Figure 5. AIS history records are also grouped by the navigation water.

The number of conflicts of every ship is equal to the sum of the number of conflicts of every ship within every navigation water.
Here, in the process of calculation of the number of ship conflicts, the ship domain can be expressed as the model of Fujii [8], Goodwin [27] or Davis [28], [29]. The procedure of the calculation of ship conflict is as follows:

- Interpolating all the trajectory segments grouped by navigation water. And then, making the times are consistent for these trajectory segments that need to be determined if there are existing conflicts.
- Calculating all the distance between the trajectory segments based on the time consistency. And then, extracting the closest point of approach.
- Detect all the ship domains at the closest point of approach based on the collision detection algorithm and calculate the number of conflicts.
- Calculating the ship collision frequency on the basis of the number of conflicts and the causation probability in different encounter situations.

According to the ship’s MMSI code, each ship’s trajectory data is extracted from the database, and the ship’s trajectory is classified in chronological order. The time in which lack of the trajectory information is located, and then the trajectory data of the ship is interpolated to obtain the ship trajectory data at this time, mainly the speed, heading, and position of the ship. Through the above processing, it can ensure the comparison and processing of ship data information are at the same time. In this study, the process of ship trajectory interpolation is carried out by linear interpolation.

Select the two adjacent points on the trajectory to extract time \( t_1, t_2 \), ship speed \( v_1, v_2 \), heading \( c_1, c_2 \), ship position longitude \((x_1, x_2)\) latitude is \((y_1, y_2)\). The ship trajectory data at time \( t \) between \( t_1 \) and \( t_2 \) can be obtained by the following formula:

\[
\begin{align*}
\nu &= v_1 + (v_2 - v_1) (t - t_1) / (t_2 - t_1) \\
c &= c_1 + (c_2 - c_1) (t - t_1) / (t_2 - t_1) \\
x &= x_1 + (x_2 - x_1) (t - t_1) / (t_2 - t_1) \\
y &= y_1 + (y_2 - y_1) (t - t_1) / (t_2 - t_1)
\end{align*}
\]

Using the above-mentioned trajectory interpolation method, the trajectory segments are equally spaced with a time interval as shown in Figure 7.

According to the analysis result of the time interval, the time interval of trajectory data from AIS is often 2–3 s and most transition phase in the shape of the ship trajectory has more than 10 track points [32]. If the SOG is 10 n mile/h, the ship will sail approximately 50 meters in 10 seconds. This time interval is sufficient to ensure the reproduction of every critical maneuvering situation of the ship.

When comparing the encounter situation of two ship trajectories in the same time dimension, the ship trajectory points need to be aligned in the time dimension. To show the operation of each stage of the ship as much as possible when aligning the time dimension of the ship’s trajectory, the time alignment interval is set to 10 seconds in this article.

Then, the minimum distance point as the closest point of the two ships approach, which can be found by calculating the
distances at the same time point of different ship trajectory
segments.

2) COLLISION DETECTION
The most accurate collision detection algorithm is the pixel
 detection algorithm [33], [34], which tests each pixel of the
object. When the pixel overlap, it is a collision, but the cal-
culation of this algorithm is very large, which will seri-
ously decrease the running speed of the device. Therefore,
it is rarely used. When the accuracy requirement is not high,
the bounding sphere algorithm can be used which surrounds
the object with the circumscribed circle of the contour of the
object [35]. To test whether the two objects collide, only need
to calculate whether the distance between the two circles is
greater than the sum of the radius of the two circles. If the
distance between the two circles is greater than two circles,
which indicates there is no collision, and vice versa. Since
the tightness and simplicity of the surrounding ball are not
ideal in most cases, they are rarely used alone. Another type of
axially aligned bounding box (AABB) [30] is a bounding box
algorithm along the coordinate axis. The object is abstracted
into a rectangle whose sides are parallel to the coordinate
axes. Its simplicity is good, but the compact type is poor.
There is also an oriented bounding box (OBB) algorithm
[36], which is the best type of compact, but the amount of
calculation is larger. The general accuracy is fully competent,
and the amount of calculation is also within the range that
mobile devices can withstand, therefore it is more common.
The OBB bounding box collision detection algorithm gener-
ally uses a Separating Axis Theorem (SAT) [37], [38].

FIGURE 8. One-dimensional projection analysis.

The core of the SAT is to find an axis that can ensure
the projections of one object on the axis do not overlap
one another, then two objects are disjoint. As shown in Fig-
ure 8(a), the projections in the directions of 1/3, 5/7, 6/8 do not
overlap, and the projections in each direction in Figure 8(b)
are overlap. Therefore, the key to tackle the problem is how
to find this axis.

This algorithm is only applicable to convex polygons that
is to extend an arbitrary side of a polygon infinitely into a
straight line, if the other sides of the polygon are on the
same side of the line, such as triangle, quadrilateral, hexagon,
circle, etc. For non-convex polygons, we can decompose
them into multiple convex polygons. The algorithm can also
handle overlapping problems.

In the case of 2D, the normal vector of each edge of the
two polygons contains all possibilities of the axis. Therefore,
we only need to enumerate all the normal vectors of the edges
of the two polygons. The normal vector of a 2D vector is a
vector perpendicular to this vector. Vector \((X, Y)\) whose
normal vector can be expressed as \((Y, -X)\) or \((-Y, X)\).
Due to the direction of the normal vector does not effect.
Therefore, we can choose any one as the normal vector to
calculate the projection regions of all the vertexes of polygons
on this vector. If there is no overlap, then we can directly
determine that there is no collision on this projection, other-
wise, continue to judge the next normal vector. The following
are the general steps:

- First, calculate the normal vector of one side of the
polygon.

Let the two vertices on this side be \((x_1, y_1)\) and
\((x_2, y_2)\), then this edge can be represented by a vec-
tor as \((x_1 - x_2, y_1 - y_2)\), and the normal vector is
\((y_2 - y_1, x_1 - x_2)\).

- Second, calculate the projection of each edge of each
polygon on this normal vector, separately, and find the max-
imum and minimum values.

Suppose the vector of the edge is \((x_1, y_1)\) and the normal
vector is \((x_2, y_2)\), then the edge projection dot on the normal
vector is calculated as:

\[
\begin{align*}
    Temp_1 &= x_2^2 + y_2^2 \\
    Temp_2 &= x_1^2 + y_1^2 \\
    Dot_x &= Temp_2 x_2 / Temp_1 \\
    Dot_y &= Temp_2 y_2 / Temp_1 \\
    Dot &= Dot_x^2 + Dot_y^2
\end{align*}
\] (7)

Calculate the dot for each edge of the polygon and find the
maximum and minimum values.

- Third, compare the maximum and minimum values of
the two polygons. If there is an intersection, go to the first step
and continue to calculate the separation axis of the next edge.
If not, it means that an axis is found, so that the projections
of two objects do not overlap on this axis. It means that the
two objects do not collide and can end the calculation. If such an
axis is not found, it means that the objects overlapped.

FIGURE 9. Determination for projection overlap.
Suppose $C_j$ is a Convex polygon, and $j = 0, j = 1$. The vertex sets of the convex polygon are $\{V^j_i\}_{i=0}^{N_j-1}$. To determine whether two convex polygons collide, only the minimum and maximum values need to be calculated and checked if there is a gap between the polygon projection, as shown in Figure 9.

$$\begin{aligned}
\lambda_{\min}^j (D) &= \min_{0 \leq i \leq N_0} \left\{ D \cdot V^j_i \right\} \\
\lambda_{\max}^j (D) &= \max_{0 \leq i \leq N_0} \left\{ D \cdot V^j_i \right\}
\end{aligned} \quad (8)$$

And then, judging the inequality as follows:

$$\lambda_{\min}^0 (D) > \lambda_{\max}^1 (D) \text{ or } \lambda_{\max}^0 (D) < \lambda_{\min}^1 (D) \quad (9)$$

Because the circle has no edges, therefore, there is no obvious axis for projection [36]. In this case, the projection axis is defined as the line passing through the center of the circle and the vertices closest to the center of the circle as shown in Figure 10. The next step is to continue to determine each of the projection axes of another polygon as described above method. Then, we can determine the relationship between the projections. If there is overlap, these polygons are colliding, these ship domains collide.

It can be seen from Figure 4 that the ship domain contains various shapes including concave polygon, ellipse, circle, convex quadrilateral, triangle, and semi-ellipse. Since the separation axis theorem cannot be applied to ellipses and concave polygons, in this study, the concave polygon is decomposed into a series of convex polygons, making it suitable for the separation axis theorem. Therefore, the collision detection process between ship domains can be decomposed by collision detection between multiple unit bodies, which is the decomposition of the ship domain.

The following figure shows the process of decomposing the ship domain in the form of a concave polygon. During collision detection, we can detect the relationship between the decomposition parts and another ship domain or its decomposition, respectively. If any sector decomposed from concave polygon as shown in Figure 11 has a collision phenomenon with another ship domain or its decomposition, then there has a collision between the concave polygon and another, otherwise, no collision will occur.

Due to separating axis theorem is just suitable for convex polygon in 2D, therefore, it cannot be applied to these ship domains: Fujii model [8], Goodwin model [27] and Davis model [28], [29] et al. For tackling this problem, we need convert eclipses to be polygons or circle, in this study, these are converted to polygons as shown in Figure 12. These polygons can be used instead of the ellipse to detect collision with other objects.

Intersection test of ellipses with other polygons:

- If the polygon intersects the edges of the inscribed polygon or they contain each other, then the polygon intersects the ellipse.
- If the polygon does not intersect the edge of the circumscribed polygon, the polygon and ellipse must be separated. Otherwise, you need to divide the ellipse into the polygon that has more edges and iterate it step by step until a certain precision is satisfied.

C. SHIP COLLISION FREQUENCY ESTIMATION

According to the above proposed unified framework for the detection of ship collision, the number of conflicts can be obtained. However, the collision risk is different in the encounter situation, therefore the ship collision frequency is related to the encounter situation. The types of ship conflicts are defined as follows:

- Overtaking conflict. That means a conflict in which ships are lying on almost parallel courses and proceeding on the same route. The course difference of an
overtaking falls in the range $0^\circ$–$67.5^\circ$ or $292.5^\circ$–$360^\circ$. should not exceed $67.5^\circ$.

- Head-on conflict. That means a conflict in which the course difference of ships falls in the range from $175^\circ$ to $185^\circ$.
- Crossing conflict. That is referred to as a conflict in which the course difference of ships falls in the range $5^\circ$–$175^\circ$ or $185^\circ$–$355^\circ$.

The ship collision frequency could be calculated based on the combination of all conflict types.

$$f = N_{ov}p^{ov}_{c} + N_{hd}p^{hd}_{c} + N_{cr}p^{cr}_{c}$$  \hspace{1cm} (10)

where:
- $N_{ov}$ is the number of overtaking conflicts.
- $p^{ov}_{c}$ is the causation probability of overtaking conflicts.
- $N_{hd}$ is the number of head-on conflicts.
- $p^{hd}_{c}$ is the causation probability of head-on conflicts.
- $N_{cr}$ is the number of crossing conflicts.
- $p^{cr}_{c}$ is the causation probability of crossing conflicts.

IV. RESULTS AND DISCUSSIONS

The static information of the ship contains a large number of abnormal samples, but the abnormal samples have many forms of expression, which is difficult to mark. The resulting sample imbalance has seriously affected the classification and recognition of traditional neural networks, and the shallow structure is also difficult to prepare. A complex function that combines high dimensions. To overcome the above effects, this article takes the static information of the inbound and outbound ships in Ningbo in China as an example, extracts 338 normal samples, and uses 80% of the total data as the training data set and 20% as the test data set to verify the model.

A. AIS DATA PRE-PROCESSING RESULTS

The static information data of the ship with abnormality is detected according to the model obtained by the training, and if the data has an abnormality, the reconstruction error is greater than the set threshold. The detection result is represented by 0 and 1, 0 means that the data is normal, and 1 means that the data is abnormal.

Using the above-mentioned ship static information anomaly detection process, the abnormalities of the data set are separately tested to verify the feasibility of the anomaly detection method proposed in this article. The detection effects are shown in the following three aspects.

1) DATA ATTRIBUTE VALUE REPLACEMENT

Data attribute value replacement refers to an abnormal situation in which some eigenvalues are converted in the data set, causing the data to seriously deviate from the overall state. It is difficult to discriminate manually for such anomalies. The occurrence of a displacement also causes a peak or glitch in the reconstruction error curve. In Figure 14, the LOA and Draught attribute values are replaced at the index of 0, 8, 16, 24, 32, 40, 48, 56 and 64 for the partial data in the normal test data. The modified data set is then tested. The test results are shown in Figure 14.

2) MISSING DATA ATTRIBUTE VALUE

Missing data attribute values means that most of the values in the dataset are missing or zero, which is often referred to as sparse data. The abnormality of static information data of such ships is generally caused by incomplete data input of AIS, interference of signal transmission, abnormality of sensors, etc., which presents the occurrence of spikes or burrs at the sampling points. In Figure 15, the data in the testing data is zeroed at the index of 0, 10, 20, 30, 40, 50 and 60, and then the modified data is detected. The test results are shown in the figure below.

3) DATA ATTRIBUTE VALUES ARE NOT RELEVANT

Data attribute values are not related to the fact that the overall correlation is not satisfied between some of the data attributes in the data set. The anomalies of static information data of such ships are relatively hidden and difficult to be discovered. It is difficult to carry out effective detection simply by relying on manually set rules. In Figure 15, the partial data in the testing data is at the index 0, 10, 20, 30, 40, 50 and 60, and the attribute values of the other BM are equal to the attribute
values of the LOA, and then the modified data is detected. The test results are shown in Figure 16.

Combined with the static data of Ningbo Port’s inbound and outbound ships throughout the year, the detection results reveal that the proposed method can effectively identify singular points and missing information. By analyzing the static information detection results of the ship under three kinds of abnormal conditions, it can be seen that the proposed method has better detection effect. Especially in the case that the attribute values are not relevant, it will be difficult for humans to discriminate them. The method used in this article solves this problem well. At the same time, the artificially set outliers are fully identified in each state. However, due to the uncertainty of the training model and threshold selection, the anomaly detection model still has misjudgment.

B. SHIP TRAFFIC ANALYSIS

This section shows that the analysis for ships involved in conflicts. The ships sailing in the Ningbo Zhoushan port consist of ten groups with length. Figure 17 provides three different distributions of relative speed for ships involved in the different conflict types. Figure 17(a) indicates that the relative speed distribution of ships with involvement in head-on conflicts where the major relative speed is 20 knots. Figure 17(c) indicates the relative speed distribution of ships involved in overtaking conflicts. The major relative speeds fall in the range 5-12.5 knots or 25-32 knots. Analysis Figure 17, we can find that the relative speeds of ships with involvement in head-on conflicts are generally larger than those involved in overtaking and crossing conflicts. Since higher relative speeds could cause more serious consequences, a head-on collision is more likely to be associated with a high level of severity.

Figure 18 shows the composition of the length of the ship involved in different ship conflicts. Figure 18(a) indicates the stable length distribution of ships with involvement in the head-on conflict where the percentage of each length of the ship is flat. Figure 18(b) and Figure 18(c) show that the majority of length in crossing conflict and overtaking conflict is around 150 meters.

Figure 19 presents the composition of the spatial distribution with involvement in different ship conflict types. From Figure 19, we can find that the most risk area is 6U due to the huge amount of conflicts. Analysis Figure 4, there exists many berths around 6U that make it concludes the south traffic flow and north traffic flow, therefore there is a high frequency of traffic joint in this area. Which results in a great deal of conflicts in any types. Comparing upstream and downstream ships, the upstream ships have a higher conflict frequency. In addition, this area has a big turn, which
further increase the traffic complexity. The increase in the number of ship maneuvering could lead to an increase of the number of conflicts. This indicates that the ships navigating in this area have the biggest probability of being involved in head-on conflicts, crossing conflicts and overtaking conflicts. Figure 19(a) have two peaks at 2U and 6U, respectively. The little higher frequency of head-on conflicts is mostly likely due to the traffic width in these area (2D/U) are narrower than others (see Figure 4) because of the existence of Xi Houmen bridge. From Figure 18, we can also find that the major conflicts appear in these areas: 7D/U, 8D/U and 10U. This may be due to the facts that ships from the nearby anchorage. For the alert area (joint area), A3 is more safe than others. Due to the existence of westbound lane in A1 and the entrance of ships merging from north in A2, that lead to a little higher frequency of conflicts.

To calculate the probability of conflicts in different types, the causation probability values and the number of conflicts are applied. In theory, the causation probability values should be different from the studied areas. However, there is still no evidence to support that their values could significantly differ from waters. Therefore, many previous researches on different waters have utilized similar causal probabilities. Refer to previous research results [1], [2], [4], [9], we utilize three causation probability values which have been given by previous research: 

\[ p_{ov}^{cr} = 1.30 \times 10^{-4} \]

for the overtaking conflicts, 

\[ p_{hd}^{cr} = 4.90 \times 10^{-5} \]

for the head-on conflicts, and 

\[ p_{cr}^{cr} = 4.90 \times 10^{-5} \]

for the crossing conflict.

### C. COLLISION FREQUENCY

Table 1 presents the impacts of the ship duty period on the quantification of collision at Ningbo Zhoushan Port. From Table 1, it is clear that fewer head-on, crossing, and overtaking collision occurs in the duty period of the second officer.
The high frequency of ship conflicts in 2U, attention should be further shown that the most risk head-on area is 2U. Due to the downstream traffic, the collision frequency during the duty period of the second officer is lower than others.

It should also be noted that this study is just the first step towards the collision detection of ships for the casualty risk assessment in the Ningbo Zhoushan Port. The next step should focus on the construction of a model for calculating the time of the collision.

Table 2 indicates that the ship collision frequency within the upstream area is higher than the frequency within the downstream area. On the alert area, the majority of conflict types are overtaking and crossing.

Table 3 gives analysis results based on the one-factor analysis of variance (One-Way ANOVA) [2], [4], which shows that the collision frequency during the duty period in different conflicts. Similarly, the collision frequencies in the head-on conflicts, crossing conflicts, and overtaking conflicts during the duty period of the second officer are lower than others.

V. CONCLUSION
This study provided the unified framework for calculating the collision frequency based on any ship domain. For detecting the static anomaly information of AIS data, the anomaly detection method based on autoencoder was proposed.

Analysis of the results, it is easy to find that the collision frequency occurring in the area 6U is the highest. It is also placed on tracking and managing the traffic in the Xi Houmen bridge. The most risk area involved in crossing conflicts and overtaking conflicts is around in the area 6U. Besides, the ship collision frequency during the duty period of the second officer is lower than others.

Table 2. Comparison of collision frequency in different partitions.

| Collision Type | Head-on | Crossing | Overtaking | Overall |
|----------------|---------|----------|------------|---------|
| Upstream       | 1.13×10⁴ | 1.94×10⁴ | 1.67×10⁴ | 3.72×10⁴ |
| Downstream     | 9.8×10⁴  | 6.89×10⁴ | 2.89×10⁴ | 9.88×10⁴ |
| Alert area     | 9.8×10⁴  | 4.29×10⁴ | 2.25×10⁴ | 6.64×10⁴ |

Table 3. Comparison of collision frequency in different encounter situations.

| Collision Type | Mean/day | Std/day | F   | p-value |
|----------------|----------|---------|-----|---------|
| Head-on        | 4.9×10⁻⁵ | 6.13×10⁻⁴ | 1.87 | 0.18    |
| Second officer | 2.16×10⁻⁵ | 2.6×10⁻⁴ |     |         |
| Third officer  | 8.72×10⁻⁵ | 2.3×10⁻⁴ |     |         |
| Crossing       | 2.25×10⁻⁴ | 1.85×10⁻⁴ | 2.80 | 0.08    |
| Second officer | 1.13×10⁻⁴ | 2.28×10⁻⁴ |     |         |
| Third officer  | 1.57×10⁻⁴ | 3.07×10⁻⁴ |     |         |
| Overtaking     | 1.18×10⁻⁴ | 1.29×10⁻⁴ | 2.48 | 0.10    |
| Second officer | 6.42×10⁻⁴ | 2.27×10⁻⁴ |     |         |
| Third officer  | 1.12×10⁻⁴ | 4.19×10⁻⁴ |     |         |
| Overall        | 3.48×10⁻⁴ | 3.15×10⁻⁴ | 2.86 | 0.08    |

REFERENCES
[1] X. Qu, Q. Meng, and L. Su, “Ship collision risk assessment for the Singapore Strait,” Accident Anal. Prevention, vol. 43, no. 6, pp. 2030–2036, Nov. 2011, doi: 10.1016/j.aap.2011.05.022.

[2] J. Weng, Q. Meng, and X. Qu, “Vessel collision frequency estimation in the Singapore Strait,” J. Navigat., vol. 65, no. 2, pp. 207–221, Apr. 2012, doi: 10.1017/S037346331000683.

[3] J. Weng and S. Xue, “Ship collision frequency estimation in port fairways: A case study,” J. Navigat., vol. 68, no. 3, pp. 602–618, May 2015, doi: 10.1017/S0373463314000885.

[4] T. Chai, J. Weng, and X. De-Qi, “Development of a quantitative risk assessment model for ship collisions in fairways,” Saf. Sci., vol. 91, pp. 71–83, Jan. 2017, doi: 10.1016/j.ssci.2016.07.018.

[5] R. Szlapczynski, “A unified measure of collision risk derived from the concept of a ship domain,” J. Navigat., vol. 59, no. 3, pp. 477–490, Sep. 2006, doi: 10.1017/S0373463306003833.

[6] F. Goerlandt and P. Kujala, “Traffic simulation based ship collision probability modeling,” Rel. Eng. Syst. Saf., vol. 96, no. 1, pp. 91–107, Jan. 2011, doi: 10.1016/j.ress.2010.09.003.

[7] T. Macduff, “Probability of vessel collisions,” Ocean Ind., vol. 9, no. 9, pp. 144–148, Sep. 1974.

[8] Y. Fujii and K. Tanaka, “Traffic capacity,” J. Navigat., vol. 24, no. 4, p. 534–552, Oct. 1971, doi: 10.1016/S0373-4633(00)022384.

[9] P. T. Pedersen, “Collision risk for fixed offshore structures close to high-density shipping lanes,” Proc. Inst. Mech. Eng., M, J. Marit. Maritime Environ., vol. 216, no. 1, pp. 29–44, Jun. 2002, doi: 10.1243/14750900232082121.

[10] J. Montewka, P. Goerlandt, and P. Kujala, “Determination of collision criteria and causation factors appropriate to a model for estimating the probability of maritime accidents,” Ocean Eng., vol. 40, pp. 50–61, Feb. 2012, doi: 10.1016/j.oceaneng.2011.12.006.

[11] J. Montewka, T. Hinz, P. Kujala, and J. Matusiak, “Probability modelling of vessel collisions,” Rel. Eng. Syst. Saf., vol. 95, no. 5, pp. 573–589, Mar. 2010, doi: 10.1016/j.ress.2010.01.009.

[12] C. Tam, R. Bucknall, and A. Greig, “Review of collision avoidance and path planning methods for ships in close range encounters,” J. Navigat., vol. 62, no. 3, pp. 455–476, Jul. 2009, doi: 10.1017/S0373463308005134.

[13] T. G. Fowler and E. Sorgard, “Modeling ship transportation risk,” Risk Anal., vol. 20, no. 2, pp. 225–244, Apr. 2000, doi: 10.1111/0272-4332.202022.

[14] L. Zhao, G. Shi, and J. Yang, “Ship trajectories pre-processing based on AIS data,” J. Navig., vol. 75, no. 5, pp. 1210–1230, Apr. 2018, doi: 10.1017/S0373463318000188.

[15] Technical Characteristics For An Automatic Identification System Using Time-Division Multiple Access in the VHF Maritime Mobile Band, Standard ITU-R M.1371, Feb. 2014. [Online]. Available: http://www.itu.int/rec/R-REC-M.1371/en

[16] J. M. Mou, C. V. D. Tak, and H. Ligteringen, “Study on collision avoidance in busy waterways by using AIS data,” Ocean Eng., vol. 37, nos. 5–6, pp. 483–490, Apr. 2010, doi: 10.1016/j.oceaneng.2010.01.012.

[17] C. Fan, F. Xiao, Y. Zhao, and J. Wang, “Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data,” Appl. Energy, vol. 211, pp. 1123–1135, Feb. 2018, doi: 10.1016/j.apenergy.2017.12.005.

[18] W. Yu and C. Zhao, “Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net,” IEEE Trans. Control Syst. Technol., vol. 28, no. 3, pp. 1083–1091, May 2020, doi: 10.1109/TCST.2019.2897946.
D. Liu, G. Shi: Ship Collision Risk Assessment Based on Collision Detection Algorithm

DONGDONG LIU received the B.Eng. degree in nautical science from Dalian Maritime University (DMU), Dalian, China, in 2016, where he is currently pursuing the Ph.D. degree in maritime science and technology. His research interests include traffic organization, abnormal detection, fundamental study of traffic flow on the sea, ship collision avoidance, and collision risk assessment.

GUOYOU SHI received the B.Eng., M.Eng., and Ph.D. degrees from Dalian Maritime University (DMU), Dalian, China, in 1993, 1996, and 2006, respectively. He was an Assistant, a Lecturer, and an Associate Professor with the Navigation College, DMU, where he is currently a Professor. His research interests include ship maneuvering, intelligent ship collision avoidance, and technologies for the maritime autonomous surface ship.

19] Z. Sun and H. Sun, “Stacked denoising autoencoder with density-grid based clustering method for detecting outlier of wind turbine components,” IEEE Access, vol. 7, pp. 13078–13091, 2019, doi: 10.1109/ACCESS.2019.2893206.

20] Z. Jingsong, W. Zhaolin, and W. Fengchen, “Comments on ship domains,” J. Navig., vol. 46, no. 3, pp. 422–436, Sep. 1993, doi: 10.1017/S0373463300011875.

21] A. Hörteborn, J. W. Ringsberg, M. Svanberg, and H. Holm, “A revisit of the definition of the ship domain based on AIS analysis,” J. Navig., vol. 72, no. 3, pp. 777–794, May 2019, doi: 10.1017/S0373463318000978.

22] W. Zhang, C. Kopca, J. Tang, D. Ma, and Y. Wang, “A systematic approach for collision risk analysis based on AIS data,” J. Navig., vol. 70, no. 5, pp. 1117–1132, May 2017, doi: 10.1017/S0373463317000212.

23] L. P. Perera and C. G. Soares, “Collision risk detection and quantification in ship navigation with integrated bridge systems,” Ocean Eng., vol. 109, pp. 344–354, Nov. 2015, doi: 10.1016/j.oceaneng.2015.08.016.

24] R. Szlapczynski and J. Szlapczynska, “An analysis of domain-based ship collision risk parameters,” Ocean Eng., vol. 126, pp. 47–56, Nov. 2016, doi: 10.1016/j.oceaneng.2016.08.030.

25] R. Szlapczynski and J. Szlapczynska, “A method of determining and visualizing safe motion parameters of a ship navigating in restricted waters,” Ocean Eng., vol. 129, pp. 363–373, Jan. 2017, doi: 10.1016/j.oceaneng.2016.11.044.

26] R. Szlapczynski and J. Szlapczynska, “Review of ship safety domains: Models and applications,” Ocean Eng., vol. 145, pp. 277–289, Nov. 2017, doi: 10.1016/j.oceaneng.2017.09.020.

27] E. M. Goodwin, “A statistical study of ship domains,” J. Navig., vol. 28, no. 3, pp. 328–344, Jul. 1975, doi: 10.1017/s0373463300041230.

28] P. V. Davis, M. J. Dove, and C. T. Stockel, “A computer simulation of multi-ship encounters,” J. Navig., vol. 35, no. 2, pp. 347–352, May 1982, doi: 10.1017/S0373463300022177.

29] P. V. Davis, M. J. Dove, and C. T. Stockel, “A computer simulation of marine traffic using domains and arenas,” J. Navig., vol. 33, no. 2, pp. 215–222, May 1980, doi: 10.1017/S0373463300035220.

30] J. Canny, “Collision detection for moving polyhedra,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 2, pp. 200–209, Mar. 1986, doi: 10.1109/TPAMI.1986.4767773.

31] N. Wang, “An intelligent spatial collision risk based on the quaternion ship domain,” J. Navig., vol. 63, no. 4, pp. 733–749, Sep. 2010, doi: 10.1017/S0373463310000202.

32] L. Zhao and G. Shi, “A method for simplifying ship trajectory based on improved Douglas–Peucker algorithm,” Ocean Eng., vol. 166, pp. 37–46, Oct. 2018, doi: 10.1016/j.oceaneng.2018.08.005.

33] P. M. Hubbard, “Collision detection for interactive graphics applications,” IEEE Trans. Vis. Comput. Graphics, vol. 1, no. 3, pp. 218–230, Sep. 1995, doi: 10.1109/2945.466717.

34] G. Baciu, W. S.-K. Wong, and H. Sun, “RECODE: An image-based collision detection algorithm,” J. Vis. Comput. Animat., vol. 10, no. 4, pp. 181–192, Nov. 1999. [Online]. Available: https://ieeexplore.ieee.org/document/732079, doi: 10.1109/PCCGA.1998.732079.

35] E. Rimon and S. P. Boyd, “Obstacle collision detection using best ellipsoid fit,” J. Intell. Robot. Syst., vol. 18, no. 2, pp. 105–126, Feb. 1997, doi: 10.1023/A:1007960531949.

36] J.-W. Chang and M.-S. Kim, “Efficient triangle–triangle intersection test for OBB-based collision detection,” Comput. Graph., vol. 33, no. 3, pp. 235–240, Jun. 2009, doi: 10.1016/j.cag.2009.03.009.

37] S. Gottschalk, “Separating axis theorem,” Dept. Comput. Sci., UNC Chapel Hill, Chapel Hill, NC, USA, Tech. Rep. TR96-024, 1996.

38] C. Liang and X. Liu, “The research of collision detection algorithm based on separating axis theorem,” Int. J. Sci., vol. 2, no. 10, pp. 110–114, 2015.

DONGDONG LIU received the B.Eng. degree in nautical science from Dalian Maritime University (DMU), Dalian, China, in 2016, where he is currently pursuing the Ph.D. degree in maritime science and technology. His research interests include traffic organization, abnormal detection, fundamental study of traffic flow on the sea, ship collision avoidance, and collision risk assessment.

GUOYOU SHI received the B.Eng., M.Eng., and Ph.D. degrees from Dalian Maritime University (DMU), Dalian, China, in 1993, 1996, and 2006, respectively. He was an Assistant, a Lecturer, and an Associate Professor with the Navigation College, DMU, where he is currently a Professor. His research interests include ship maneuvering, intelligent ship collision avoidance, and technologies for the maritime autonomous surface ship.

* * *