Journal of *Marine Science and Engineering*

MDPI

Article

Edge-Based Dynamic Spatiotemporal Data Fusion on Smart Buoys for Intelligent Surveillance of Inland Waterways

Ruolan Zhang [1], Chenhui Zhao [1], Yu Liang [2], Jingfeng Hu [1] and Mingyang Pan [1,*]

[1] Navigation College, Dalian Maritime University, Dalian 116026, China; zhangruolan@dlmu.edu.cn (R.Z.); zch2022@dlmu.edu.cn (C.Z.); hujf@dlmu.edu.cn (J.H.)
[2] Science and Technology Information Division, Guangxi Zhuang Autonomous Region Port and Navigation Development Center, Nanning 530029, China; femanly@163.com
* Correspondence: panmingyang@dlmu.edu.cn

Abstract: Increasing vessel traffic in narrow, winding inland waterways has heightened the risk of accidents, driving the need for improved surveillance and management. This study addresses the challenge of real-time processing and synchronization of voluminous video and AIS data for effective waterway management. We developed a surveillance method utilizing smart buoys equipped with sensors and edge computing devices, enabling dynamic spatiotemporal data fusion. The integration of AIS data with advanced computer vision techniques for target detection allows for real-time traffic analysis and provides detailed navigational dynamics of vessels. The method employs an enhanced Long Short-Term Memory network for precise trajectory prediction of AIS data and a single-stage target detection model for video data analysis. Experimental results demonstrate significant improvements in ship detection accuracy and tracking precision, with an average position prediction error of approximately 1.5 m, which outperforms existing methods. Additionally, a novel regional division and a Kalman filter-based method for AIS and video data fusion were proposed, effectively resolving the issues of data sparsity and coordinate transformation robustness under complex waterway conditions. This approach substantially advances the precision and efficiency of waterway monitoring systems, providing a robust theoretical and practical framework for the intelligent supervision of inland waterways.

Keywords: continuous trajectory prediction; AIS; target detection; waterway traffic monitoring; enhanced Kalman filter

check for updates

Academic Editor: Chung-yen Kuo

Received: 28 December 2024
Revised: 20 January 2025
Accepted: 22 January 2025
Published: 24 January 2025

Citation: Zhang, R.; Zhao, C.; Liang, Y.; Hu, J.; Pan, M. Edge-Based Dynamic Spatiotemporal Data Fusion on Smart Buoys for Intelligent Surveillance of Inland Waterways. *J. Mar. Sci. Eng.* **2025**, *13*, 220. https://doi.org/10.3390/jmse13020220

Copyright: © 2025 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

With the recovery of the global economy, the importance of maritime transport in global trade has steadily increased. According to statistics, global maritime trade reached 12.292 billion tons in 2023, marking a 2.4% growth compared to the previous year [1]. However, the rapid development of shipping has led to a significant rise in traffic density and vessel flow. Traditional monitoring methods, such as manual observation and basic radar surveillance, are no longer sufficient to address the increasingly complex and dynamic shipping scenarios. To ensure the efficiency and safety of waterway navigation, regulatory approaches urgently need to transition from traditional experience-based and manual judgment methods to intelligent management using technologies such as the Internet of Things (IoT), big data, and artificial intelligence (AI).

Intelligent regulation aims to enable data-driven decision making, allowing waterway management authorities to acquire near real-time dynamic information about waterways and to predict potential risks in advance. This enhances the scientific rigor and timeliness

of management decisions. Furthermore, intelligent management efficiently processes and analyzes large-scale data, alleviating the fatigue and delays associated with manual monitoring. It meets the growing demand for digitalization and intelligent transformation in the maritime sector, laying a solid foundation for the long-term development of digital waterways and intelligent shipping.

With the continuous advancement of China's "Digital Waterway" initiative, digital technology has become a key driver of economic development [2]. Big data and artificial intelligence (AI) play a critical role in enabling the digital transformation of maritime trade [3]. Monitoring methods for waterways have gradually transitioned from traditional manual observation to data-driven intelligent regulation, with digital collection and processing of waterway operation data being realized in key regions and busy segments such as the Yangtze and Pearl Rivers. However, existing intelligent management systems still face challenges, often adding complexity to waterway management tasks. For instance, the level of automation in video data analysis remains low, frequently requiring human involvement, which fails to meet real-time monitoring demands and results in delayed responses to potential risks.

Shore-based cameras offer wide-area coverage and are suitable for monitoring macroscopic traffic conditions. However, their deployment is limited by their field of view and camera resolution. In contrast, cameras mounted on buoys, being closer to vessels operating within the waterway, offer flexibility and high-resolution close-range monitoring. This proximity allows them to capture detailed information more effectively.

The current Automatic Identification System (AIS) is primarily designed for macro-level traffic flow analysis [4], making it difficult to achieve refined monitoring and management in localized waterways and critical segments. AIS data is limited by transmission delays and fixed message intervals, resulting in lagged and insufficiently intuitive information. While video surveillance can capture real-time traffic conditions of vessels in waterways, it only provides visual information on vessel appearance and movement trajectories. It lacks the ability to obtain dynamic data such as heading, speed, and precise latitude and longitude coordinates.

Radar, as a traditional monitoring method, offers all-weather operational capability and stable performance in large-scale target detection [5]. However, it also has significant limitations. Radar has low resolution for small targets, making it particularly difficult to distinguish multiple vessels in close proximity under complex waterway conditions. Additionally, radar cannot capture the appearance features of vessels or distinguish target types, which limits its applicability in refined monitoring scenarios.

Therefore, relying on a single monitoring method is insufficient to meet the demands of efficient regulation and rapid acquisition of vessel information in busy waterways. To enhance maritime transport safety, efforts should focus on improving the accuracy of individual sensors or adopting multi-sensor data fusion to address future challenges [6].

In recent years, the rapid development of deep learning technology has provided new approaches for achieving refined perception of waterway conditions, traffic dynamics, and dynamic prediction of vessel behavior. Currently, maritime monitoring methods are mainly divided into three categories: AIS data-based deep learning processing, video target detection, and radar monitoring. These methods have demonstrated their advantages in the shipping industry and are of great significance for improving waterway traffic safety and monitoring accuracy.

AIS data-based processing methods, using time-series models such as LSTM and GRU for trajectory prediction and dynamic analysis [7], can compensate for the limitations of traditional AIS data in real-time monitoring [8]. AIS data provides broad coverage, offering real-time information on vessel positions, speed, and course. However, since AIS data is

inherently discrete in both time and space and is affected by data transmission delays and fixed transmission intervals, significant visualization errors occur in refined monitoring, limiting its application in complex waterway environments [9].

Video monitoring methods employ deep learning-based video target detection algorithms, which efficiently identify and track vessel targets. By using Convolutional Neural Networks (CNN), video surveillance systems can automatically extract vessel appearance features, significantly improving the analysis efficiency and accuracy of video data. Video monitoring allows for real-time capture of vessel appearance and operational trajectories, providing an intuitive display of traffic conditions [10]. However, video data is susceptible to interference from external factors such as lighting, changes in viewing angles, and other environmental conditions, which affect the stability of detection results, limiting its performance in complex environments.

Radar monitoring technology combines deep learning-based target detection, trajectory estimation, and dynamic state monitoring, providing high-resolution vessel detection capabilities in all-weather conditions [11]. Radar technology is unaffected by external factors such as weather and lighting, and can accurately provide distance and speed information of vessels, demonstrating significant advantages in harsh weather and complex waterway environments [12]. However, radar data processing is relatively complex, prone to noise interference, and has high computational costs, thus still requiring optimization in terms of real-time performance and computational efficiency.

Although the above single-sensor methods have achieved significant progress in improving monitoring accuracy, there are still many challenges in waterway regulation under multi-source data fusion. AIS data, video data, and radar data exhibit significant differences in format [13], update frequency, and spatial resolution, complicating data matching and fusion [14]. This makes it difficult to ensure system response speed in complex waterway environments, and deploying a unified fusion system across different waterway segments and environments faces issues such as device compatibility and environmental adaptability. Therefore, further technological breakthroughs and optimizations are needed in multi-source data fusion and unified deployment.

This study addresses these issues by implementing a spatiotemporal positioning method based on AIS data, target recognition based on computer vision, and a spatiotemporal heterogeneous data fusion approach. This enables the effective integration of AIS and video data in waterway monitoring. The proposed fusion method is independent of the camera deployment location, overcoming challenges such as the sparsity of AIS data and the robustness of spatial coordinate transformation in data fusion. It demonstrates strong scalability and adaptability, making it highly applicable to a variety of scenarios.

The innovations of this study include the following:

Design and Implementation of a Multi-Modal Data Fusion Method Based on Regional and Spatiotemporal Matching: The proposed method eliminates the need for data calibration and coordinate system transformations, simplifying the data processing workflow compared to existing approaches.

Integration of Visual Target Detection and Trajectory Prediction Technologies: A data processing method suitable for complex waterway environments is introduced, addressing varying scales and complexities of waterway monitoring. This enhances the recognition and prediction capabilities of existing methods in challenging navigation scenarios.

Provision of a Scalable Data Fusion Method: By improving data fusion accuracy and simplifying system deployment, this method is applicable to real-time monitoring of busy segments and refined management of critical waterways. It reduces system maintenance costs while supporting the development of digital waterways and intelligent shipping.

The remainder of this paper is organized as follows: Section 2 provides a brief review of the latest research progress related to trajectory prediction, target detection and tracking, and data fusion, which are relevant to the objectives of this study. Section 3 presents a detailed introduction to the proposed data processing and data fusion framework. Section 4 validates the effectiveness of the proposed method through experiments. Section 5 discusses the role of the proposed method in data matching as well as the limitations of existing approaches. Finally, Section 6 summarizes the contributions of the proposed method.

2. Literature Review

Current inland waterway monitoring faces challenges such as the pressure of managing high-density traffic and the complexity of vessel trajectories. These issues impact waterway safety and constrain the development of digital management systems. In inland waterway regulation, the Automatic Identification System (AIS) and video monitoring systems are the primary means of data acquisition. AIS can provide real-time information on vessel position, speed, and heading, offering wide coverage without blind spots. However, its data updates occur at fixed intervals and suffer from transmission delays, which result in latency, particularly in high-density traffic areas.

On the other hand, video monitoring can intuitively capture vessel appearance and movement trajectories, making it suitable for real-time monitoring in dynamic environments. However, it cannot provide precise data on vessel heading, speed, or latitude and longitude. Therefore, to achieve effective data fusion, it is essential to integrate AIS data for spatiotemporal positioning based on trajectory prediction with video-based target recognition. By extracting and combining features from both data sources, a comprehensive data fusion approach can be developed.

2.1. Spatiotemporal Positioning Based on Trajectory Prediction

Accurate spatiotemporal positioning plays a crucial role in data fusion. Spatiotemporal positioning methods based on trajectory prediction primarily rely on AIS data to estimate the future positions and movement states of vessels through trajectory forecasting. These methods can be broadly divided into two categories: traditional trajectory prediction methods and deep learning-based trajectory prediction methods.

In traditional research methods, trajectory prediction primarily includes approaches such as Kalman Filtering (KF), Gray Model (GM), and Support Vector Regression (SVR). Perera proposed the use of Extended Kalman Filtering [15] as an adaptive filtering algorithm for predicting vessel trajectories, including position and speed. However, in inland waterway navigation, particularly in crowded multi-vessel environments and under complex hydrological conditions, the adaptability of Kalman Filtering is limited. It struggles to effectively handle nonlinear relationships and the rapid dynamics of vessel motion, leading to reduced prediction accuracy. Yin applied a gray prediction system for online trajectory prediction in vessel navigation [16]. However, as data volumes increase and feature dimensions expand, the accuracy and robustness of gray prediction models become limited, making them inadequate for addressing the complexities of modern inland water transportation. Kawan proposed a vessel trajectory prediction method using Support Vector Regression (SVR) [17], which maps nonlinear functions to higher-dimensional features to develop predictive models for vessel movement. While SVR improves prediction accuracy, it suffers from high computational complexity, making it challenging to achieve real-time predictions, particularly for large-scale datasets. Fossen designed a vessel motion estimation model using AIS data and Extended Kalman Filtering [18], successfully predicting future vessel movements. However, these traditional methods are often constrained by

their inability to handle the nonlinear and dynamic nature of complex inland waterway scenarios effectively.

Overall, traditional prediction algorithms offer strong interpretability and low computational complexity. However, they face significant limitations when dealing with complex nonlinear relationships and large-scale data in inland waterway navigation, restricting their effectiveness in practical applications.

Despite their strengths, traditional prediction methods struggle to address the challenges posed by intricate nonlinear relationships and vast data volumes. With the advancement of neural networks, new vessel trajectory prediction methods have emerged as a research focus in recent years. Neural networks have demonstrated exceptional performance in handling complex data, making them widely applicable to nonlinear and multidimensional datasets. Compared to traditional models, neural network-based approaches exhibit clear advantages in capturing complex nonlinear relationships and multidimensional features, further solidifying their role as a promising solution for trajectory prediction.

Tang [19] and Zhang [20] applied Long Short-Term Memory (LSTM) networks for vessel trajectory prediction by converting AIS data into navigational time series for model training. At the same time, research on attention mechanisms and LSTM variants has been advancing, leading to the proposal of various hybrid network models. For instance, Gao introduced the improved MP-LSTM method based on traditional LSTM [21], while Zhang proposed a novel vessel trajectory prediction model that integrates LSTM autoencoders [22], attention mechanisms, and a bidirectional LSTM (AABiL) structure. Furthermore, Zhao developed the GAT-LSTM method [23], which combines Graph Attention Networks (GAT) with LSTM for trajectory prediction.

Although deep learning-based methods have made significant progress in vessel trajectory prediction, the computational cost and real-time performance of these complex models still require further optimization to meet the stringent demands of inland waterway navigation for real-time processing and limited computational resources.

2.2. Research on Target Recognition Based on Computer Vision

Target detection is a core task in computer vision, and in recent years, significant progress has been made in detection algorithms with the development of deep learning technology. Traditional target detection algorithms mainly relied on feature extraction methods, such as SIFT (Scale-Invariant Feature Transform) proposed by Lowe [24] and HOG (Histogram of Oriented Gradients) proposed by Dalal [25]. These methods require relatively low computational resources and exhibit good stability in specific scenarios or environments. However, these approaches rely heavily on handcrafted feature design, which is dependent on human expertise. Their computational efficiency is relatively low, and they lack the capability to capture high-level semantic information. As a result, they fail to meet the demands of current applications, particularly those involving more complex and diverse scenarios.

With the development of deep learning, traditional methods have gradually been replaced by self-learning models. For instance, Convolutional Neural Networks (CNN), proposed by LeCun [26], have shown powerful feature extraction capabilities and have driven rapid advancements in target detection. Currently, target detection methods based on CNN can be categorized into single-stage and two-stage detection algorithms. Two-stage target detection algorithms, such as RCNN [27], Fast R-CNN [28], Faster R-CNN [29], and Mask R-CNN [30], represent significant milestones in this area. Beyond the development of CNN-based variants, Zhang proposed the Adaptive Training Sample Selection (ATSS) method by introducing an adaptive training sample selection mechanism [31]. This

method excels at handling objects of varying sizes, significantly improving the detection performance for densely packed small objects.

Although two-stage methods excel in accuracy, their slower inference speed makes them unsuitable for real-time applications such as autonomous driving, intelligent surveillance, and security systems, which require fast processing. To improve processing speed while maintaining high accuracy, researchers have focused on single-stage target detection methods. These methods simplify the entire detection pipeline, making them more suitable for current application scenarios.

Among single-stage target detection algorithms, Liu proposed the Single Shot MultiBox Detector (SSD) [32], which utilizes feature maps of multiple scales to enhance small object detection. Lin introduced the RetinaNet model [33], which employs Focal Loss to effectively reduce the impact of negative samples. Zhou proposed the CenterNet algorithm, which simplifies the detection process by predicting the center point of each object [34], eliminating the need for anchor boxes and reducing hyperparameter tuning. Redmon introduced the YOLO (You Only Look Once) algorithm [35], which has since been widely studied and developed further with iterations such as YOLOv5, YOLOv8, and YOLOv11. Many researchers have focused on improving YOLO- and SSD-based algorithms to balance detection accuracy and speed. For example, Ting enhanced detection accuracy by integrating the original feature extraction algorithm with GhostbottleNet [36]. Sun improved detection speed by adding a Bidirectional Feature Fusion Module (Bi-DFFM) [37]. Guo optimized the backbone network and loss function of existing algorithms, achieving excellent detection performance on various datasets [38]. Cai further improved detection accuracy by incorporating channel attention and lightweight GhostConv into the backbone network, alongside a Coordinate Attention Feature Pyramid Network (CA-FPN) and lightweight spatial pyramid pooling [39].

2.3. Research on Spatiotemporal Data Fusion Methods

In recent years, the importance of multi-source data fusion in maritime surveillance and vessel traffic management has become increasingly prominent. AIS data and video surveillance data each provide unique and complementary information—dynamic positional data from AIS and visual appearance features from video. The fusion of these two data types can address the limitations of single data sources, enhancing the comprehensiveness and accuracy of monitoring systems.

Early studies primarily relied on traditional geometric models, focusing on mathematical methods to solve data alignment and matching problems. For instance, Man proposed a fusion method based on Kalman filtering [40], where a marker line was set, and vessels detected by video crossing the recognition line were matched with nearby AIS targets. While this method performed well in static scenarios, its effectiveness was limited in highly dynamic environments. Bloisi employed a Bayesian recursive method [41] to update and estimate vessel positions by comparing the vessel velocity vectors from video data with those from AIS data to find the best match. However, the fusion performance was constrained when vessels engaged in collision-avoidance maneuvers or when waterway traffic density was high. Additionally, their method required manually marked ground-truth data or human intervention to update the rotation parameter θ during camera calibration, making it unsuitable for large-scale maritime surveillance scenarios. Lu calculated the distance and azimuth between vessels and cameras, combining this with the Dead Reckoning algorithm to project the vessel's image coordinates into the world coordinate system [42]. AIS data within a $\pm$60-s timestamp range was used for matching. However, this approach relied heavily on the completeness and reliability of AIS data, and the Dead Reckoning algorithm demonstrated poor generalization, making it unsuitable for refined perception scenarios.

In recent years, with the development of deep learning, an increasing number of researchers have focused on multi-source data fusion by leveraging deep learning networks and object detection technologies. Methods combining coordinate system transformation for data matching have garnered significant attention. Huang proposed a method that transforms image coordinate systems into latitude–longitude coordinates of the river surface using forward perspective projection [43], converting pixel coordinates into georeferenced coordinates on the river surface. Liu employed a pinhole camera model and camera calibration to transform vessel positions in the world coordinate system into pixel coordinates [44]. They achieved data association by leveraging the detected center points of vessels and the distance and angular deviation derived from AIS data. Qu calibrated camera parameters to convert coordinates into the world coordinate system [45]. They further calculated directional components of vessel speed to estimate the time of arrival and matched vessel position coordinates provided by the detection network with the nearest AIS data. Gülsoylu utilized interpolation to estimate azimuth and distance, transforming world coordinates into image coordinates to fuse AIS messages with bounding boxes in images [46]. They assumed vessel speed remained constant since the last observation and used k-d tree-based nearest neighbor search to assign AIS messages to the nearest bounding box in the image. Qu simplified the coordinate transformation problem using relative angle estimation by projecting latitude and longitude coordinates from the geographic coordinate system into the image coordinate system [47]. They measured the similarity of direction, speed, and position between AIS and video-based vessel trajectory points and modeled the AIS–visual data matching problem as a linear assignment problem in a weighted bipartite graph. Guo adopted a pinhole model to project AIS data from the world coordinate system into the pixel coordinate system [48]. They extracted vessel trajectories from video and AIS data, calculated trajectory similarity, and applied the Hungarian algorithm to obtain matching results. Ding proposed a three-step correction method that projects image coordinates into latitude–longitude coordinates [49]. Using the Hungarian algorithm, they identified the optimal match, fusing AIS data with vessel trajectories effectively.

However, most existing methods heavily rely on strict geometric constraints, and the perspective transformation models established under different viewing angles and focal lengths require manual adjustments. The robustness of coordinate transformation often lacks strong generalizability. Moreover, the update frequency and accuracy of AIS data can be affected by environmental factors such as signal interference and weather conditions, potentially resulting in missing or delayed AIS information. Sparse data further degrades the effectiveness of the fusion process.

3. Methodology

To enhance the refined perception and regulation capabilities of waterway states and traffic dynamics, we propose a spatiotemporal heterogeneous data fusion method for waterway monitoring. This method integrates spatiotemporal positioning information derived from AIS data with vessel target recognition from video data. As shown in Figure 1, the spatiotemporal heterogeneous data fusion framework consists of several modules, including time alignment, spatiotemporal positioning based on AIS data, target recognition based on video data, and spatiotemporal data fusion.

First, AIS data obtained from the AIS receiver on the buoys is aligned with video monitoring data from the cameras, using the video timestamps as the temporal reference. After alignment based on the initial timestamps, vessel trajectories are extracted from the AIS data, and spatiotemporal positioning is performed using trajectory prediction. Simultaneously, vessel targets in the video data are identified using deep learning networks. The camera's field of view is then artificially divided into five regions, each assigned a unique

identifier. Using the real-time monitoring timestamps from the video, spatiotemporal positioning information derived from AIS trajectory prediction and target recognition information from the video are synchronized. This allows the determination of the vessel's current position within the waterway, including its specific region and identifier. The association between vessel video data and AIS data is then established through set-based relationships. This framework enables precise monitoring and integration of vessel data from heterogeneous sources.

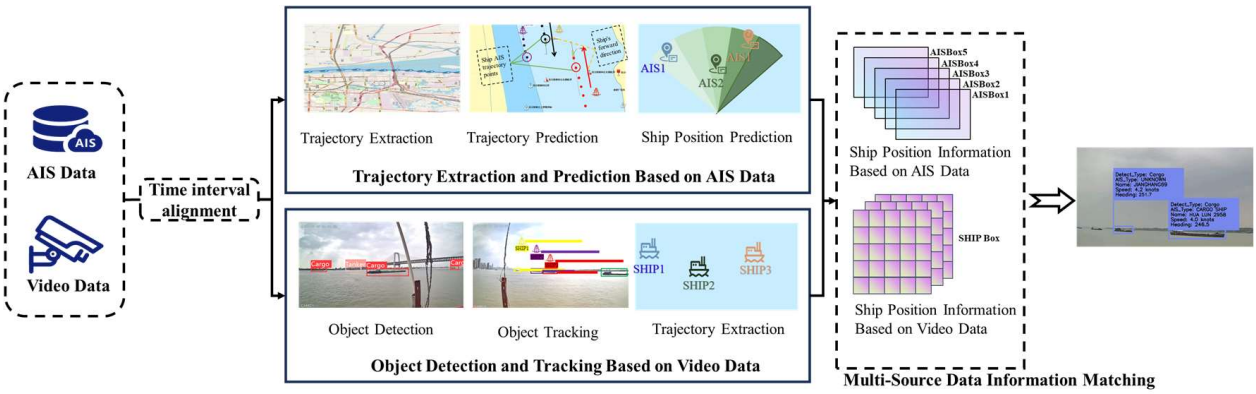


Figure 1. Multi source data fusion framework.

3.1. HATM-Based Spatiotemporal Vessel Positioning Method

For vessel trajectory prediction based on AIS data discussed in this study, a Long Short-Term Memory (LSTM)-based trajectory prediction method is adopted. LSTM is highly effective at capturing long-term dependencies within sequential data and excels at processing multivariate data. The AIS data received is a typical time-series dataset, which primarily includes dynamic information such as vessel latitude and longitude, heading, and speed. Leveraging LSTM enables efficient interpretation of these dynamic vessel characteristics and accurate predictions of vessel information at future timestamps.

To further enhance the accuracy and real-time performance of AIS trajectory prediction, this study proposes an improved Hybrid Attention-based Trajectory Model (HATM). HATM integrates attention mechanisms, Bidirectional LSTM (Bi-LSTM), and Gated Recurrent Units (GRU). Compared to existing models, such as AABiL and GAT-LSTM, HATM demonstrates significant advantages in both prediction accuracy and computational efficiency.

The input to the model consists of historical AIS data of vessels, including features such as latitude, longitude, speed, heading, and message timestamps. This data is represented as a time series of length T, denoted as X, serving as the basis for trajectory prediction.

$$X = \{x_1, x_2, x_3, \ldots, x_T\}$$

where $x_T = [Lat_t, Lon_t, Speed_t, Course_t, Posting_t]$ represents the state of the vessel at time t, including its latitude, longitude, speed, heading, and reporting time.

By incorporating attention mechanisms, HATM enhances its ability to focus on critical time steps and features in the data, while Bi-LSTM improves context understanding by processing sequences bidirectionally. The addition of GRU ensures computational efficiency without sacrificing predictive performance. Together, these components enable HATM to effectively capture the spatiotemporal dependencies in AIS data, providing accurate and efficient predictions of vessel movements and future positions.

The HATM model consists of the following components:

(1) Bidirectional LSTM (Bi-LSTM)

Bidirectional LSTM extends the traditional unidirectional LSTM by adding a reverse propagation network. This reverse network captures both forward and backward dependencies in the time series, as shown in Figure 2. In AIS data, this capability allows for a more comprehensive understanding of how historical vessel trajectories influence future trajectories, thereby enhancing the network's ability to predict vessel movements.

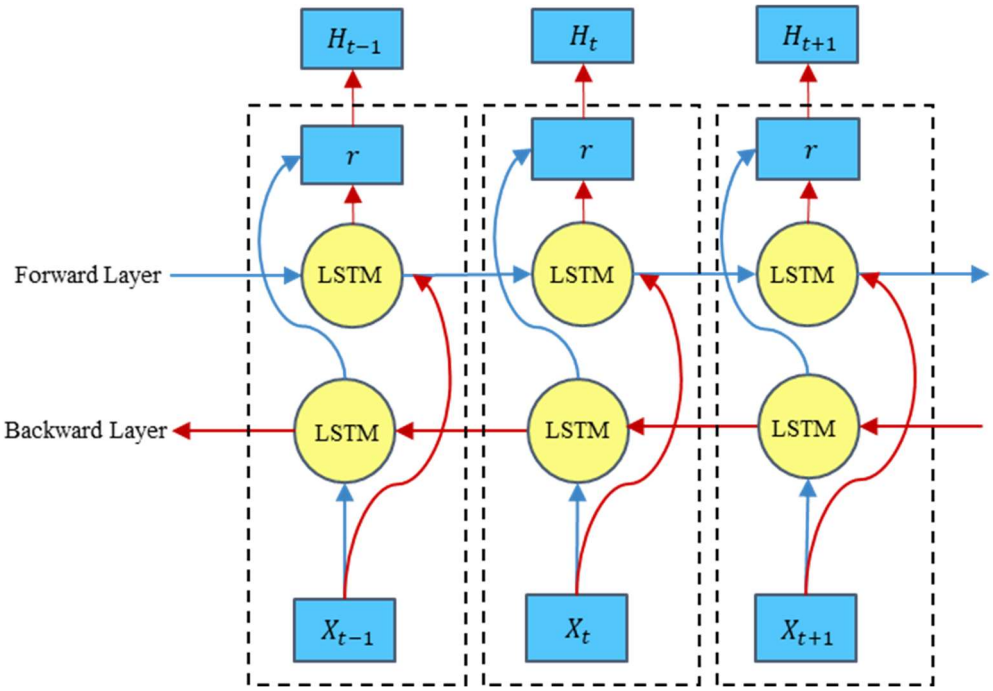


Figure 2. Bi-LSTM Structure Diagram.

When the input time series X is passed to the Bi-LSTM layer, it processes the data and produces a sequence of hidden states for each time step, denoted as $H_{Bi-LSTM}$. The formula for the output is as follows:

$$H_{Bi-LSTM} = \{h_1, h_2, h_3, \ldots, h_T\}$$

where h_T represents the concatenation of the hidden layer states of forward and backward LSTM.

(2) Gated Recurrent Unit (GRU)

In the proposed LSTM-based improvement method, the Gated Recurrent Unit (GRU) is introduced to simplify the LSTM structure. By reducing the number of model parameters, the GRU lowers the computational complexity while maintaining predictive capability, making the model better suited for real-time vessel trajectory prediction tasks.

When the output of the Bi-LSTM layer, $\mathbf{H}_{Bi-LSTM}$, is passed to the GRU layer, the GRU uses update gates and reset gates to control the flow of information. These gates allow the GRU to effectively capture long-term dependencies while preserving relevant historical information. The output sequence of the GRU layer, $\mathbf{H}_{GRU}$, contains both temporal and contextual information. The sequence is expressed as follows:

$$\mathbf{H}_{GRU} = \{\mathbf{h}_{GRU,1}, \mathbf{h}_{GRU,2}, \mathbf{h}_{GRU,3}, \ldots, \mathbf{h}_{GRU,T}\}$$

where $\mathbf{h}_{GRU,T}$ represents the output of the GRU at time step t. T is the length of the input sequence.

By leveraging GRU's efficient gating mechanism, the model achieves a balance between computational efficiency and the ability to capture long-term dependencies, making it well-suited for real-time trajectory prediction.

(3) Attention Mechanism

The attention mechanism in this model mainly works by calculating the time step weight α_t, dynamically adjusting the model's focus on historical data to enhance its ability to capture features from critical time steps. Since key trajectory points, such as turning or speed changes, significantly influence the vessel's subsequent movements, the improvement based on the attention mechanism can capture the impact of these critical trajectory points on future trajectories, effectively improving the model's performance. The calculation formula for α_t is as follows:

$$\alpha_t = \frac{\exp\left(\mathrm{v}^T \tanh(\mathrm{W}\mathrm{h}_{\mathrm{GRU}}, t + \mathrm{b})\right)}{\sum_{t=1}^{T} \exp\left(\mathrm{v}^T \tanh(\mathrm{W}\mathrm{h}_{\mathrm{GRU}}, t + \mathrm{b})\right)}$$

where v, W, and b are the parameters to be learned, and α_t represents the score for the importance of time step t.

By applying the attention mechanism weights α_t, the final weighted context representation $\mathrm{h}^{\mathrm{att}}$ is obtained.

$$\mathrm{h}^{\mathrm{att}} = \sum_{t=1}^{T} \alpha_t \cdot \mathrm{h}_{\mathrm{GRU}}, t$$

Finally, the weighted context representation $\mathrm{h}^{\mathrm{att}}$ is mapped to the predicted vessel position at the next time step $\hat{\mathrm{y}}_{\mathrm{t}}$ through a fully connected layer. The specific formula is as follows:

$$\hat{\mathrm{y}}_{\mathrm{t}} = \mathrm{FC}\left(\mathrm{W}_{\mathrm{out}}\mathrm{h}^{\mathrm{att}} + \mathrm{b}_{\mathrm{out}}\right)$$

As shown in Figure 3, the HATM model proposed in this study integrates attention mechanisms, Bidirectional Long Short-Term Memory (Bi-LSTM) networks, and Gated Recurrent Units (GRU) into a unified architecture for high-precision trajectory prediction of vessel AIS data and multi-source data fusion.

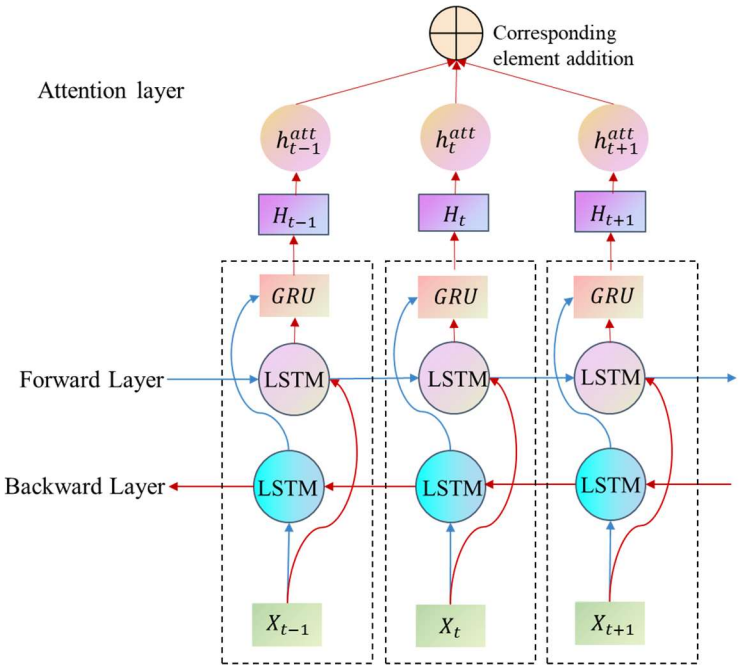


Figure 3. Network Structure Diagram of the HATM Model.

The model first transforms the input historical AIS data of vessels (including latitude, longitude, speed, heading, and timestamp) into a time series and feeds it into the Bi-LSTM network. By leveraging forward and backward propagation, the network captures global spatiotemporal dependencies within the trajectory data. Subsequently, the GRU units further optimize the information flow, utilizing update and reset gates to effectively manage long-term dependencies while reducing computational complexity, thereby enhancing the model's real-time performance and efficiency. On this basis, the attention mechanism dynamically assigns weights to focus on critical nodes in the trajectory, generating a weighted global contextual feature representation. Finally, the extracted features are mapped to the latitude and longitude coordinates of the vessel at the next time step through a fully connected layer, achieving accurate future trajectory prediction.

In this study, to train the vessel position prediction model based on the HATM architecture, we adopt the Mean Squared Error (MSE) as the loss function. MSE measures the difference between the predicted values and the actual values. Specifically, MSE evaluates the prediction accuracy by calculating the squared error between the predicted positions and the true positions and averaging the results. The specific formula is as follows:

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

$$\hat{y}_i = [\hat{Lat}_i, \hat{Lon}_i]$$

$$y_i = [Lat_i, Lon_i]$$

where $\hat{y}_i$ represents the predicted position of the i-th sample, y_i denotes its corresponding true position, and N is the total number of training samples.

During the training process, the model minimizes the MSE loss function using the gradient descent method. By updating the weights W and biases b, the value of the loss function is continuously reduced. The specific formula is as follows:

$$\mathrm{W}_{\mathrm{t+1}} = \mathrm{W}_{\mathrm{t}} - \eta\nabla \mathrm{w}\mathcal{L}_{\mathrm{MSE}}$$

$$b_{t+1} = b_t - \eta\nabla b\mathcal{L}_{MSE}$$

3.2. Multi-Modal Data Fusion

To achieve efficient fusion of AIS data and video data, this study proposes a spatiotemporal matching method that combines region division and an improved Kalman filter. The region division method partitions the camera's monitoring range into equal-angle fan-shaped areas with unique identifiers, improving the efficiency of multi-source data matching through coordinate mapping. The improved Kalman filter method enhances the traditional model by introducing a dynamic weight adjustment mechanism, optimizing for the latency of AIS data and the high frequency of video data, thereby significantly improving the accuracy and real-time performance of spatiotemporal data fusion. The specific implementation is as follows:

1. Region Construction

Region construction in this study refers to dividing the monitoring range of cameras deployed along the waterway into equal-angle fan-shaped regions with a fixed angle $\Delta\varphi$. Each region is defined as $[\varphi_i - \Delta\varphi, \varphi_i + \Delta\varphi]$ and sequentially numbered in a clockwise direction. The regions are named as $S_i, i \in (1,\ 2, \ldots, N)$. The division method is as follows:

$$\Delta\varphi = \frac{\theta_{fov}}{N}, \varphi_i = \varphi_0 + (i-1)\cdot\Delta\varphi, i \in \{1, 2, \ldots, N\}$$

where θ_{fov} represents the field of view angle of the camera, φ_0 is the central direction of the camera's field of view, and N denotes the number of fan-shaped regions.

At the same time, based on the divided angular range, the visible area in the horizontal direction is evenly divided according to the width W and height H of the video frame. The pixel width of each region is Δx.

$$\Delta x = \frac{W}{5}$$

By dividing the area, each horizontal region's boundary can be formed into equally wide rectangular areas, numbered as V_i. This ensures correspondence between the divided regions and the camera's detection range in the real world. The specific division is shown in Figure 4.

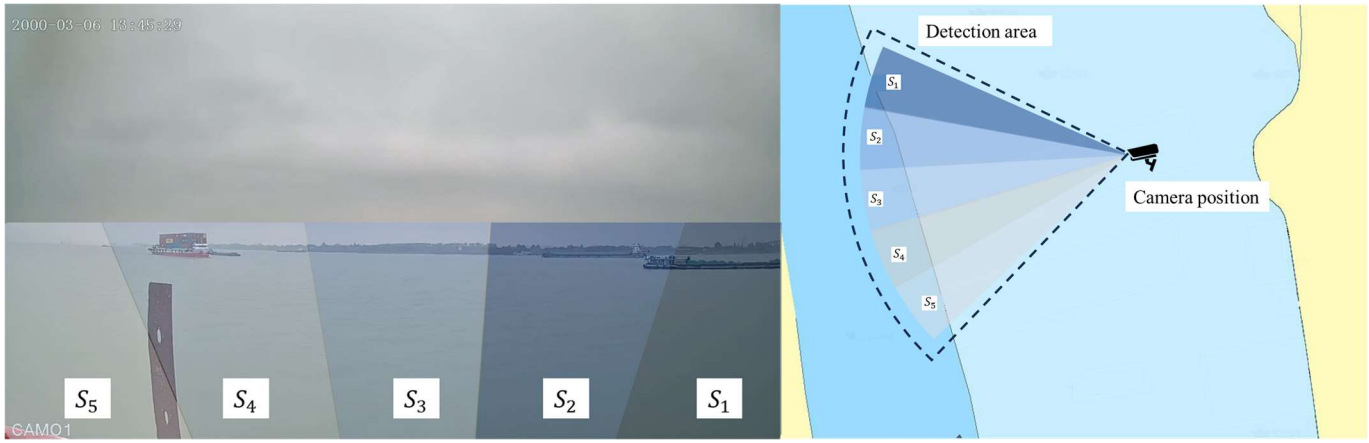


Figure 4. Schematic Diagram of Camera Detection Region Division.

During the data processing, target detection algorithms are used to obtain the real-time vessel position within the camera's detection range, tracking its movement and pixel coordinates. This allows the determination of the vessel's specific subregion $V_{i,j}$ within the fan-shaped region, where i represents the region number and j represents the current vessel number in the subregion.

Based on the LSTM vessel trajectory prediction, the prediction results are used to determine whether the AIS signal corresponds to a vessel entering the camera's actual detection range and its associated fan-shaped region $S_{i,j}$, where i indicates the region number and j represents the current vessel number within the region.

2. Data Transformation and Fusion Method

Based on the region construction mentioned in this study, the pixel coordinates (u, v) of vessel feature points extracted from the target detection process are used for calculations to determine the specific region V_i into which the vessel falls. Additionally, vessels within each region are numbered according to their horizontal pixel coordinates, obtaining the image feature $V_{i,j}$.

$$V_i = \left\lceil \frac{u}{\Delta x} \right\rceil, i \in \{1,\ 2,\ 3,\ 4,\ 5\}$$

Through region construction, the predicted latitude and longitude coordinates of the vessel at a given time, (λ, φ), can be converted into polar coordinates (r_{AIS}, ϕ_{AIS}) based on the camera's deployment location (λ_c, φ_c) and the azimuth angle ϕ. Subsequently, the region number S_i of the corresponding fan-shaped area can be determined using the radial coordinate r. The specific conversion formula is as follows:

$$r_{AIS} = \sqrt{(\lambda - \lambda_c)^2 + (\varphi - \varphi_c)^2}, \phi_{AIS} = \arctan\left(\frac{\varphi - \varphi_c}{\lambda - \lambda_c}\right)$$

After the transformation, the coordinate r_{AIS} can be used to determine the fan-shaped region S_i where the vessel is located. By sorting and numbering all vessels within this region, the actual AIS feature $S_{i,j}$ of the vessel within the region can be obtained.

The fusion of different data for the same vessel can be achieved by comparing the consistency of the two numbered features. Specifically, this involves first calculating whether the region numbers are consistent. The calculation formula is as follows:

$$\Delta S = \left| V_{i,j} - S_{i,j} \right|$$

When $\Delta S = 0$, it indicates that the vessel's position is consistent across the two dimensions, and it can be considered that the video data and AIS data are successfully matched. Once the video data and AIS data are successfully matched, the two data features can be fused. The fused vessel state vector is represented as X_{fused}, and the specific formula is as follows:

$$X_{fused} = \alpha X_{video} + (1 - \alpha) X_{AIS}$$

where X_{video} represents the feature vector from the video data, X_{AIS} represents the feature vector from the AIS data, and α is the weight coefficient.

3. Multi-Source Spatiotemporal Data Fusion Based on Improved Kalman Filtering

First, the vessel feature vector is modeled, and the vessel state vector is defined as x_t.

$$\mathrm{X}_t = \left[\lambda_{AIS}, \varphi_{AIS}, v_x, v_y, x_{video}, y_{video} \right]^T$$

In the formula, λ_{AIS}, φ_{AIS} represent the latitude and longitude information from the vessel's AIS data, $v_{x,t}, v_{y,t}$ denote the velocity components of the vessel based on AIS data at time t, and x_{video}, y_{video} represent the pixel coordinates of the vessel in the video.

Assuming a certain dynamic consistency between the AIS and video data, the state transition model is defined as follows:

$$\mathrm{X}_{AIS,t+1} = \begin{bmatrix} 1 & 0 & \Delta_t & 0 \\ 0 & 1 & 0 & \Delta_t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_{AIS,t} \\ \varphi_{AIS,t} \\ v_{x,t} \\ v_{y,t} \end{bmatrix} + w_{AIS,t}$$

In the formula, Δ_t represents the time step, i.e., the time interval between data points, and $\mathrm{w}_t \sim \mathcal{N}(o, \mathcal{Q})$ denotes the process noise, where the covariance matrix Q represents the internal dynamic disturbance of the system.

$$\mathrm{X}_{VIDEO,t+1} = \mathrm{X}_{VIDEO,t} + w_{VIDEO,t}$$

where $\mathrm{w}_{AIS} \sim \mathcal{N}(o, \mathcal{Q}_{AIS})$ and $\mathrm{w}_{VIDEO} \sim \mathcal{N}(o, \mathcal{Q}_{VIDEO})$ represent the process noise.

Subsequently, the Kalman filter observation model is defined to fuse the observational information from the two data sources. The observation models for AIS data and video data are denoted as Z_{AIS} and Z_{VIDEO}, respectively.

$$\mathrm{Z}_{AIS,t} = \begin{bmatrix} \lambda_t \\ \varphi_t \end{bmatrix} + \boldsymbol{\nu}_{AIS,t}$$

where $\boldsymbol{\nu}_{AIS,t} \sim \mathcal{N}(o, \mathcal{R}_{AIS})$, where the covariance $\mathcal{R}_{AIS}$ represents the observation error of the AIS data.

$$\mathrm{Z}_{VIDEO,t} = \begin{bmatrix} x_{video} \\ y_{video} \end{bmatrix} + \boldsymbol{\nu}_{VIDEO,t}$$

where $\nu_{AIS,t} \sim \mathcal{N}(o, \mathcal{R}_{VIDEO})$, where the covariance $\mathcal{R}_{VIDEO}$ represents the observation error of the video data.

The Kalman filter state prediction is as follows:

$$\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1}$$

$$P_{t|t-1} = AP_{t-1|t-1}A^T + Q$$

If the video and AIS data share the same region number R_i, the observations are considered consistent and can be jointly updated as follows:

$$K_t = P_{t|t-1}H^T\left(HP_{t|t-1}H^T + R\right)^{-1}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t\left(z_t - H\hat{x}_{t|t-1}\right)$$

$$P_{t|t} = (I - K_tH)P_{t|t-1}$$

4. Experiment and Result Analysis

4.1. Experimental Data Foundation

The selected experimental scenario is the Wuhu section of the Yangtze River channel within Wuhu, China, located in the middle reaches of the Yangtze River. The specific location is shown in Figure 5. This channel is characterized by its complex bends, high vessel density, limited visibility, and numerous tributaries, resulting in challenging scenarios, such as crossing and encountering vessels. The curvature of the channel increases the difficulty of vessel navigation and collision avoidance, making it suitable for testing the recognition and prediction capabilities of multi-source data fusion technology in complex environments. The high density of vessels also places stringent demands on the system's real-time performance and accuracy, enabling an assessment of the system's performance under load conditions. Wuhu, as an important transportation hub for Yangtze River shipping, has well-established data collection facilities, ensuring comprehensive and diverse experimental data. The experimental results can be directly applied to improve the safety and efficiency of shipping, providing significant practical application value.

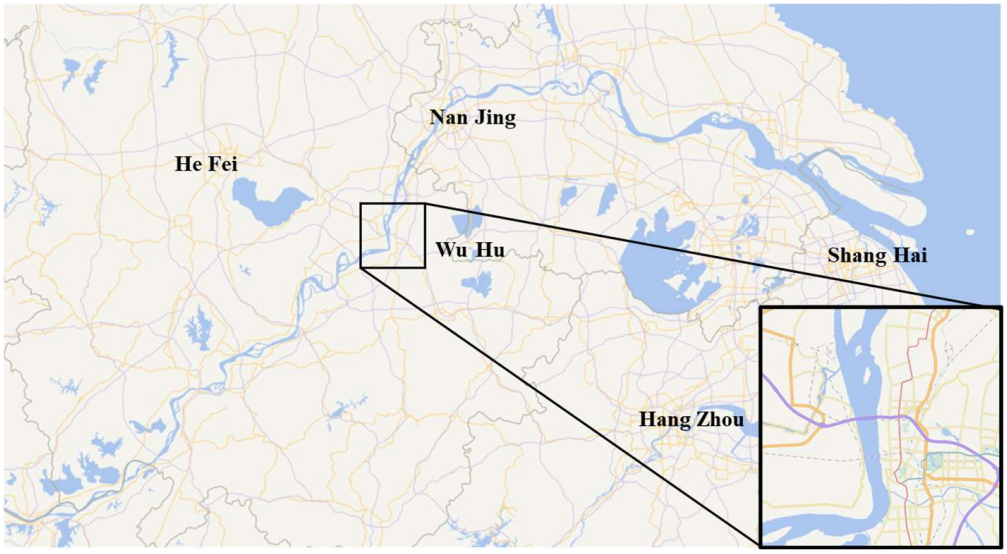


Figure 5. Schematic diagram of the experimental location.

The AIS equipment and network cameras involved in this study are located within the Wuhu channel section near the #4 navigation buoys of the Wuhu Gongtie Bridge. Video and AIS data are remotely transmitted to the local server through an SSH network tunnel.

In this experiment, over 100,000 AIS data entries and synchronized video data were collected using AIS receivers and network cameras through an SSH tunnel from 22 to 23 October 2024, for use in the study. The collected AIS data includes information such as vessel MMSI, vessel dimensions, course, speed, heading, latitude and longitude, vessel name, propagation type, and reporting time, as detailed in Table 1. The primary vessel types include cargo ships, tankers, yachts, fixed navigation aids, and tugboats.

Table 1. AIS Raw Data.

Postime	MMSI	Lon	Lat	Type	Course	Heading	Speed
1729569854024	413879988	118.1832	31.28196	Cargo	224.8	224	4.5
1729564048202	413822215	118.2819	31.30661	Cargo	271.5	271	4
1729570578878	413796194	118.2916	31.41126	Cargo	201.1	201	1.9
1729563796485	413766867	118.2676	31.37651	Cargo	108.5	108	4.7
1729571781557	413769378	118.2844	31.30192	Cargo	89.4	89	4
1729564178513	413232230	118.221	31.29838	Cargo	269.5	267	8.4
1729571094419	413818093	118.2204	31.29995	Cargo	269.1	269	4.6
1729562040699	413979802	118.2437	31.30105	Cargo	256.5	256	4
1729571775865	413803562	118.2897	31.30706	Cargo	273.6	273	4.7

The video data was primarily obtained through network cameras. During the target detection training, the video was processed, saving one image every 60 frames based on the video length. At this stage, 54,856 original images were captured from the perspective of navigation buoys. From these, 35,112 images were selected for annotation and then divided into training, testing, and validation sets in an 8:1:1 ratio, forming the dataset used for this experiment.

4.2. Data Preprocessing

The raw AIS data obtained contained many issues that, if directly input into the neural network for training, would result in large training errors and poor model performance. Thus, data preprocessing was necessary. The preprocessing steps included removing abnormal values, data interpolation, and classification of data using MMSI as the unique identifier for each vessel. The data was also sorted by time after converting the reporting time to Beijing Time.

For significant error data filtering, this study deleted data points outside the longitude range of 118.0°–118.5° and the latitude range of 31.0°–31.5° based on the actual deployment location of the AIS receiver. It also removed data with Maritime Mobile Service Identity (MMSI) values of fewer than nine digits, MMSI values such as 972 (fixed frequency bands for non-vessel equipment), and heading values outside the range of 0°–360°. The specific monitoring range of the camera is shown in Figure 6.

By applying the experimental geographic range (longitude and latitude) and relevant data specifications, AIS data outside the specified range and MMSI values that do not correspond to the vessel types targeted in this experiment were removed during the initial screening process (Algorithm 1).

Algorithm 1: AIS Data Cleaning

1: **Input:** AIS data file path $\text{input}_{\text{file}}$, Output file path $\text{output}_{\text{file}}$
2: Load the dataset from $\text{input}_{\text{file}}$ into memory as data
3: Remove records where:
    lon < 118.0 OR lon > 118.5
    lat < 31.0 OR lat > 31.5
4: Remove records where:
    LENGTH(mmsi) $\neq$ 9 OR mmsi = 972
5: Remove records where:
    heading < 0 OR heading > 360
6: Save the cleaned data to $\text{output}_{\text{file}}$
7: **Output:** Cleaned AIS data saved in $\text{output}_{\text{file}}$

Due to the excessive amount of information transmitted and received in AIS data, spectrum congestion may occur, leading to information conflicts or loss. The received data often exhibits anomalies such as data jumps, as shown in Figure 7. Such data should be removed when extracting vessel trajectories to ensure the accuracy of trajectory extraction and improve the precision of trajectory prediction.

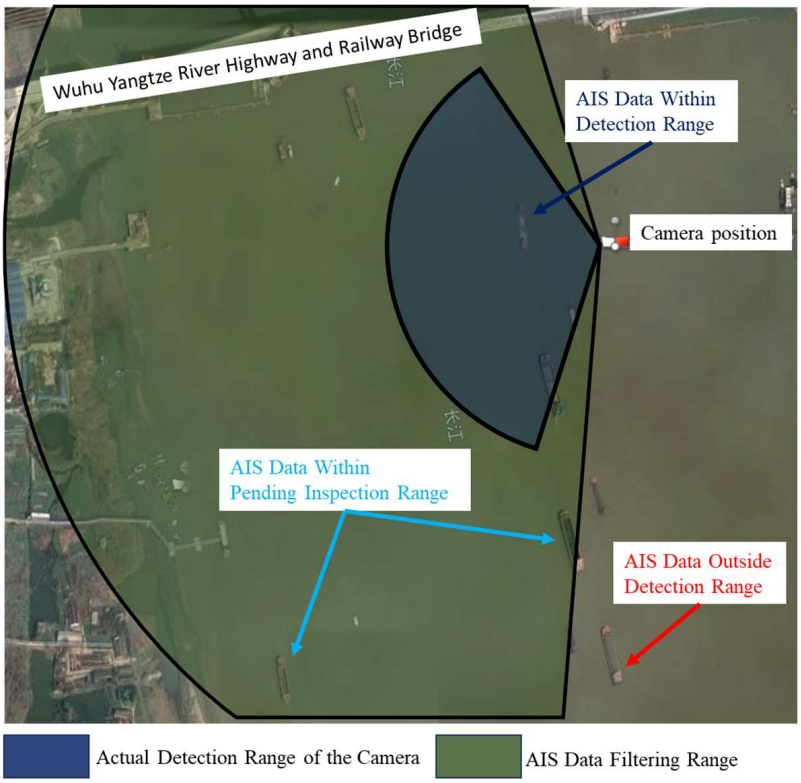


Figure 6. Camera detection range.

To address such abnormal data, this study employs the Z-score method to eliminate outliers. The core principle of the Z-score method is to calculate the deviation distance between a specific value and the mean, identifying outliers based on this distance. These outliers typically represent anomalies within a dataset. For the AIS dataset received, each type of dynamic information for each vessel is grouped as $X = (x_1, x_2 \cdots, x_n)$. First, the mean μ and standard deviation σ for each group are calculated using the following formulas:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_n$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

where x_i represents the i-th data point, and n is the total number of data points in each group.

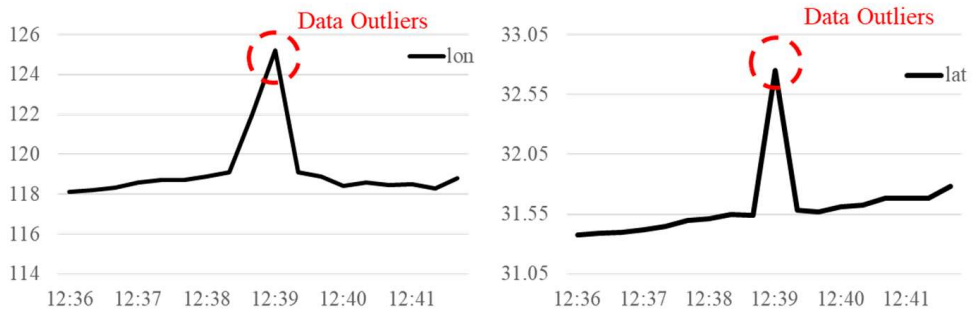


Figure 7. Example of Random Numerical Error.

By calculating the mean μ and standard deviation σ, the degree of dispersion for each data point can be quantified. After obtaining the mean μ and standard deviation σ, the standard score (Z-score) for each data point can be computed using the following formula:

$$z_i = \frac{x_i - \mu}{\sigma}$$

When the calculated result $|z_i| \geq 3$, the corresponding data point x_i is considered an outlier and is removed from the dataset, as shown in Figure 8.

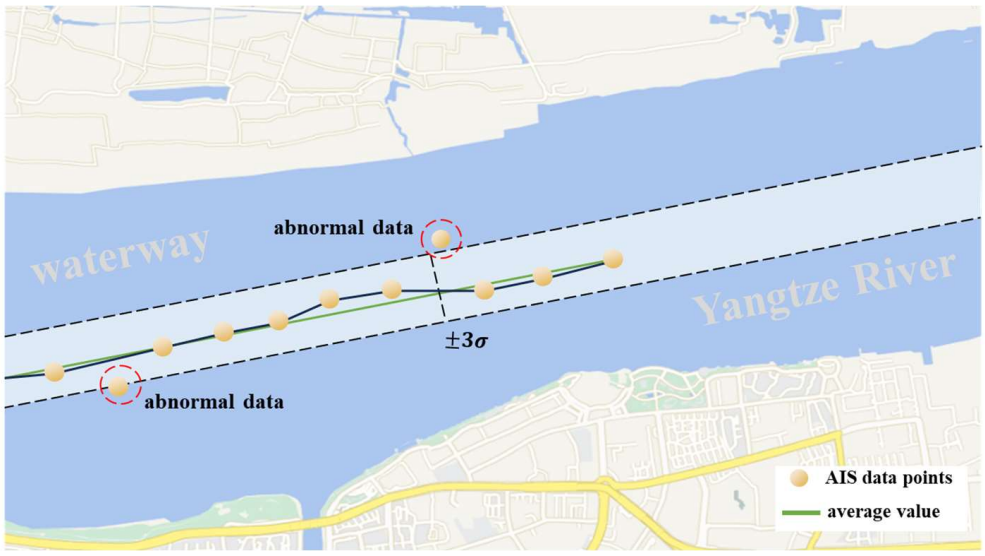


Figure 8. Schematic diagram of AIS abnormal data processing.

After the initial filtering of the output, clustering analysis is performed on the filtered data. Through this analysis, abnormal jump values and erroneous data points present in the dataset are processed using the standard deviation method (Algorithm 2).

Algorithm 2: Anomaly Detection and Removal Using Z-score

1: **Input:** AIS data file path $input_{file}$, Output file path $output_{file}$, Threshold z_threshold = 3
2: Load the dataset from $input_{file}$ into memory as data
3: Group the data by MMSI or dynamic categories
4: For each group g in grouped data do:
5: Calculate mean μ and standard deviation σ for the group
6: For each data point x_i in g do:
7: Calculate Z-score: $z_i = (x_i - \mu)/\sigma$
8: If $|z_i| \geq z_{threshold}$ then:
9: Remove x_i from g
10: Save the cleaned data (after removing outliers) to $output_{file}$
11: **Output:** Cleaned AIS data saved in $output_{file}$

After completing these two steps, AIS data meeting the conditions is obtained. However, due to the limited number of AIS data points and shorter trajectories within the received data range, handling abnormal data results in missing values in the same time slots. To mitigate the impact of missing data on the research results, the K-Nearest Neighbors (KNN) algorithm is employed to fill in the missing values (Algorithm 3).

The KNN interpolation method estimates missing values by calculating the distances between missing data points and known data points and then using a weighted average based on these distances. This method is particularly suitable for irregularly distributed missing values and scenarios where data exhibits local similarity, making it well-suited for the localized traffic environment discussed in this study. The core of the KNN interpolation method involves identifying the K nearest known data points for each missing value and estimating the missing value based on the values of these neighboring points.

The obtained AIS data is structured into a dataset, where the i-th data point is represented as (x_i, y_i), and the missing point is denoted as $x_{missing}$. During computation, the first step is to calculate the distance metric. In this study, Euclidean distance is used as the metric. The distance calculation formula is as follows:

$$d(x_{missing}, x_i) = \sqrt{\sum_{j=1}^{n} \left(x_{missing,j} - x_{i,j}\right)^2}$$

$x_{missing,j}$ and $x_{i,j}$ represent the feature values of the missing point and other known data points in the j-th dimension, where n is the data dimensionality.

Subsequently, for each missing value point $x_{missing}$, the K nearest neighbors (x_{i1}, y_{i1}), $(x_{i2}, y_{i2}) \ldots (x_{in}, y_{in})$ with the smallest distances to the missing point are selected. Finally, the missing value is estimated by weighting these K neighbors inversely proportional to their distances, with the calculation formula as follows:

$$y_{missing} = \frac{\sum_{k=1}^{K} \frac{y_{ik}}{d\left(x_{missing}, x_{ik}\right)}}{\sum_{k=1}^{K} \frac{1}{d\left(x_{missing}, x_{ik}\right)}}$$

In this formula, y_{ik} represents the actual value of the k-th nearest neighbor, and $d\left(x_{missing}, x_{ik}\right)$ denotes the distance between the missing point and the k-th nearest neighbor.

Since this method is highly dependent on the value of K, a K that is too small can make the interpolation results sensitive to noise, while a K that is too large may lead to overly smooth interpolation, failing to effectively capture the local features of the data. Therefore, this paper employs cross-validation to further determine the optimal value of

K. By dividing the dataset and using each *K* value for model training and validation, the mean squared error (MSE) is calculated for each *K*. The *K* value with the smallest MSE is selected as the final interpolation *K*, as shown in Figure 9.

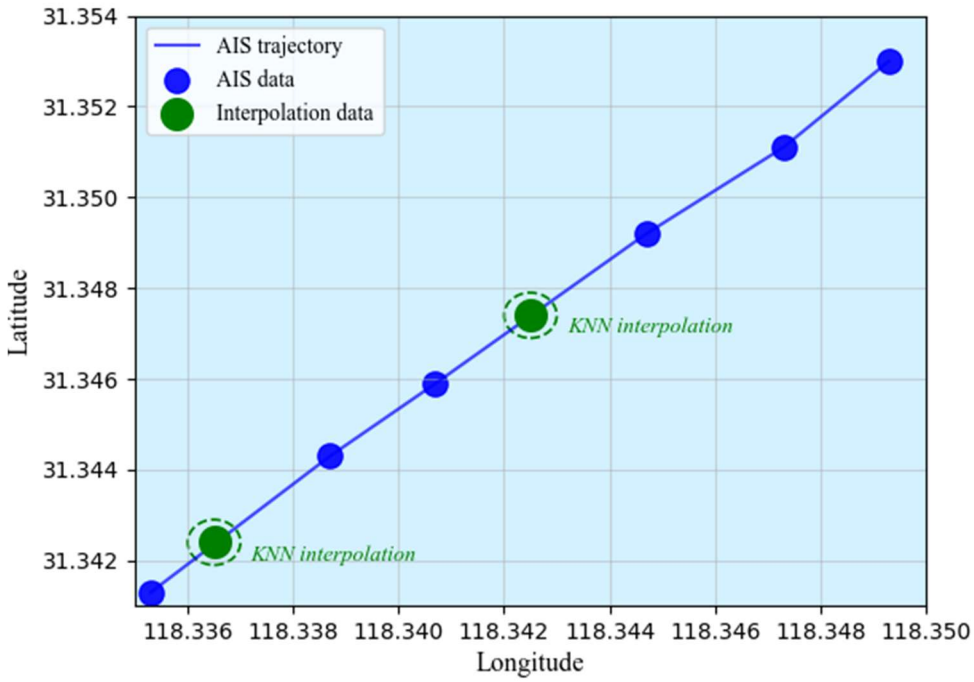


Figure 9. Schematic diagram of KNN interpolation processing.

Algorithm 3: KNN-Based Missing Value Imputation

Input: Dataset data with missing values, Candidate K values K_set, Validation metric MSE
Initialize best_K ← None, min_MSE ← ∞
For each K in K_{set} do:
    Split data into training set $train_{data}$ and validation set val_{data}
    For each missing value $x_{missing}$ in val_{data} do:
        Compute distance $d(x_{missing}, x_i)$ for all x_i in $train_{data}$ using:
            $d(x_{missing}, x_i) = \sqrt{(\sum_{(j=1)\hat{}n} (x_{missing,j} - x_{i,j})^2)}$
        Select the K nearest neighbors $(x_{i1}, y_{i1}), \ldots, (x_{iK}, y_{iK})$
        Estimate the missing value y_missing as:
            $y_{missing} = (\sum_{(k=1)\hat{}K} y_{ik}/d(x_{missing}, x_{ik}))/$
                $(\sum_{(k=1)\hat{}K} 1/d(x_{missing}, x_{ik}))$
    Compute the MSE of the imputed values on val_data
    If MSE < min_{MSE} then:
        $best_K \leftarrow K$, $min_{MSE} \leftarrow$ MSE
End for
Impute all missing values in data using best_K and steps 6–8
Output: Dataset with missing values imputed

After processing, the final processed data can be obtained, as shown in Table 2.

During the preprocessing of video data, the dataset is first manually annotated. Following annotation, the images and label files are divided into training and testing sets in an 8:2 ratio. The annotated dataset has been made publicly available at https://github.com/20221095/AIS-CV (accessed on 28 December 2024).

Table 2. Schematic of Processed Data.

Postime	MMSI	Course	Lat	Lon	Speed
2024-10-22 12:36:44	413776927	193.2	31.38545	118.3371	5.3
2024-10-22 12:37:44	413776927	193.4	31.38399	118.3367	5.4
2024-10-22 12:38:14	413776927	192.8	31.38325	118.3365	5.4
2024-10-22 12:38:45	413776927	193.5	31.38251	118.3363	5.4

The model training process consists of three main steps. First, the spatiotemporal positioning module based on trajectory prediction is used to address delays and missing data in AIS, achieving improved ship trajectory prediction. Second, the target recognition module based on computer vision is employed to quickly detect required targets in waterway videos, enabling the detection and tracking of ship trajectories to obtain the spatiotemporal information of visual data. Finally, the fusion algorithm proposed in this paper, which combines rule-driven and data-driven approaches, integrates AIS information collected by AIS receivers with visual information obtained by cameras. This is achieved by segmenting the visible range of visual detection and the pixel coordinate regions of video monitoring screens, and then performing spatiotemporal consistency checks between pixel coordinates and AIS trajectory geographic positions.

A. Ship trajectory prediction

First, the ship trajectory prediction training is based on the LSTM network, using the HATM model selected in this study. To test the effectiveness of the proposed ABG module improvements, ablation experiments were conducted to compare it with other models. During the training process, the sliding window length was set to 6, with a step size of 1. The dataset was split into training and testing sets at a ratio of 8:2, with the training set used to train the model and the testing set used to verify the model's accuracy.

The training results demonstrate that the proposed model outperforms other comparative models in terms of two evaluation metrics: Mean Absolute Error (MAE) and training loss. As shown in Figure 10, the proposed model exhibits a faster convergence rate and lower loss values, maintaining a high level of stability, especially in the later stages of training. This further validates the effectiveness of the improved modules.

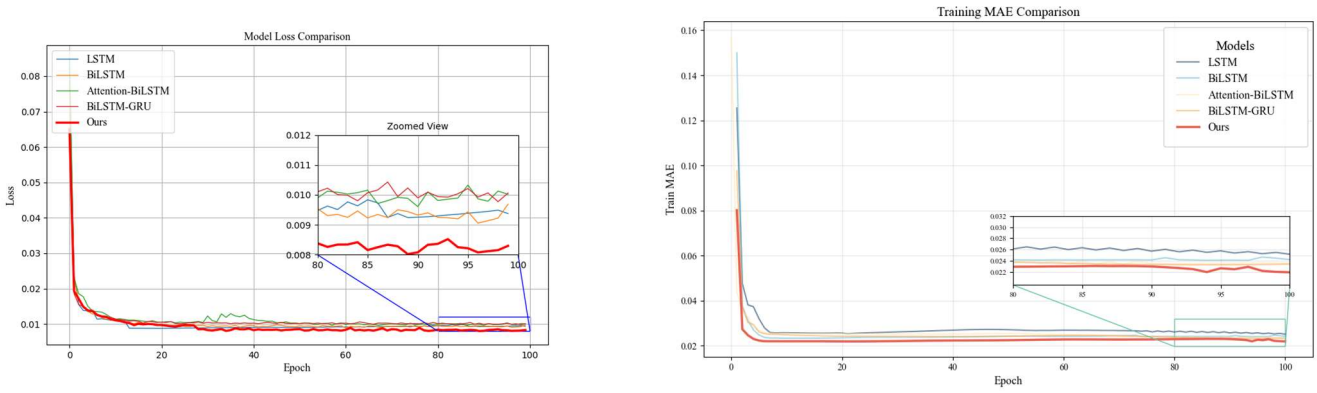


Figure 10. Comparison of Average Absolute Error and Loss in Training Different Models.

The actual training performance of the model in the experiment is shown in Figure 11, which compares the true latitude and longitude values with the predicted values in the dataset. It can be observed that the model accurately fits the ship's trajectory and remains close to the true values in most time steps.

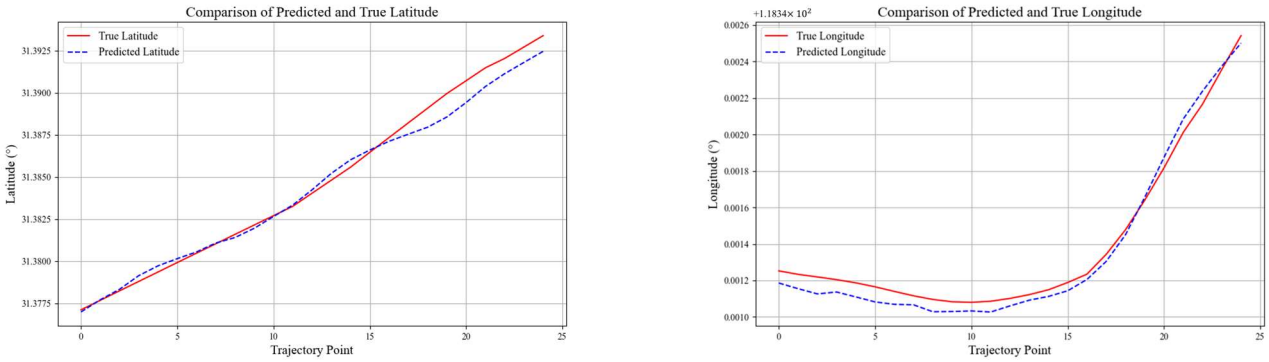


Figure 11. Results of latitude and longitude prediction errors.

To further evaluate the accuracy of the proposed prediction model, the average error was calculated by analyzing the differences between the actual and predicted ship trajectories, as shown in Figure 12. The error distribution ranges approximately from 1 to 5 m, with most errors being below 1.5 m, and the average prediction error is around 1.48 m. From the overall error fluctuation, the model maintains a low error at most time steps, with error peaks primarily occurring at points of sudden motion changes, such as when ships accelerate, decelerate, or turn.

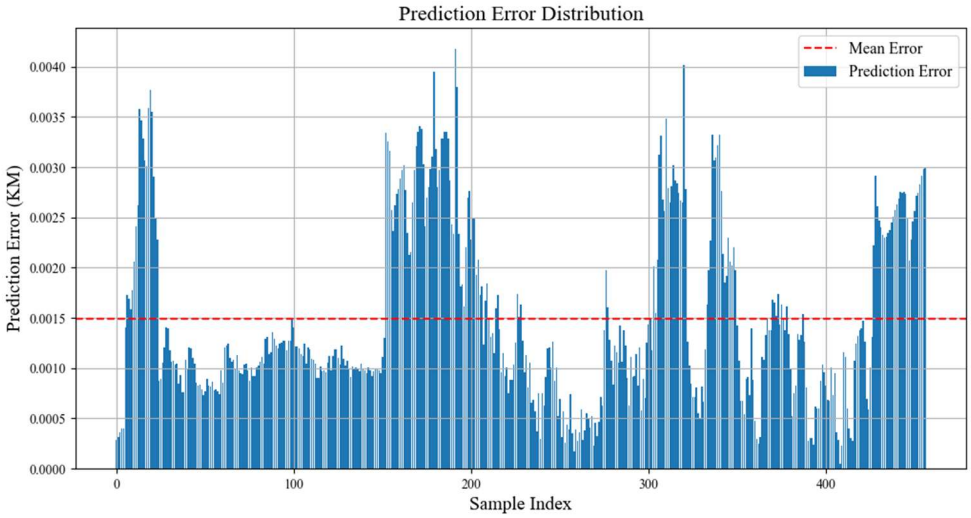


Figure 12. Comparison of Distance Error between Model Predicted Trajectory and Real Trajectory.

To better validate the effectiveness of the improvements made in the proposed HATM model, we conducted comparative experiments on the test dataset under the same experimental conditions. Several mainstream models, including LSTM, Bi-LSTM, Attention-Bi-LSTM, and Bi-LSTM-GRU, were evaluated, using the error between the predicted and actual values as the evaluation criterion. The training comparison results are shown in Figure 13. It can be observed that the proposed model demonstrates a lower range of error fluctuations and provides predictions that are closer to the actual values in complex collision avoidance scenarios. The model exhibits excellent adaptability to various scenarios within this dataset.

From Figure 13, it can be observed that all models are generally capable of predicting ship trajectories; however, their prediction performance exhibits significant differences across various scenarios and dynamic changes. The standard LSTM model performs poorly when ships execute collision avoidance maneuvers, as it is heavily affected by significant changes in heading and speed, resulting in a higher error compared to other improved models. The Bi-LSTM model, by incorporating bidirectional learning, improves predic-

tion accuracy to some extent. However, its error remains relatively large, particularly in highly dynamic and complex scenarios, where it exhibits noticeable error fluctuations. The Attention-Bi-LSTM model, by introducing the attention mechanism, better captures key features, resulting in more stable error performance in scenarios with large dynamic changes. However, its overall accuracy still has certain limitations. The Bi-LSTM-GRU model, by incorporating GRU units, further enhances feature extraction and learning capabilities, reducing the error range compared to Bi-LSTM and Attention-Bi-LSTM. Compared to the above models, the HATM model proposed in this study demonstrates significantly lower mean error throughout the training process, with smaller error fluctuations, showing higher stability and prediction accuracy. Particularly during complex collision avoidance maneuvers, the HATM model effectively handles abrupt changes in speed and heading, with predictions more closely aligning with the true trajectories, showcasing strong adaptability to dynamic scenarios. Its average error is only 1.07 m.

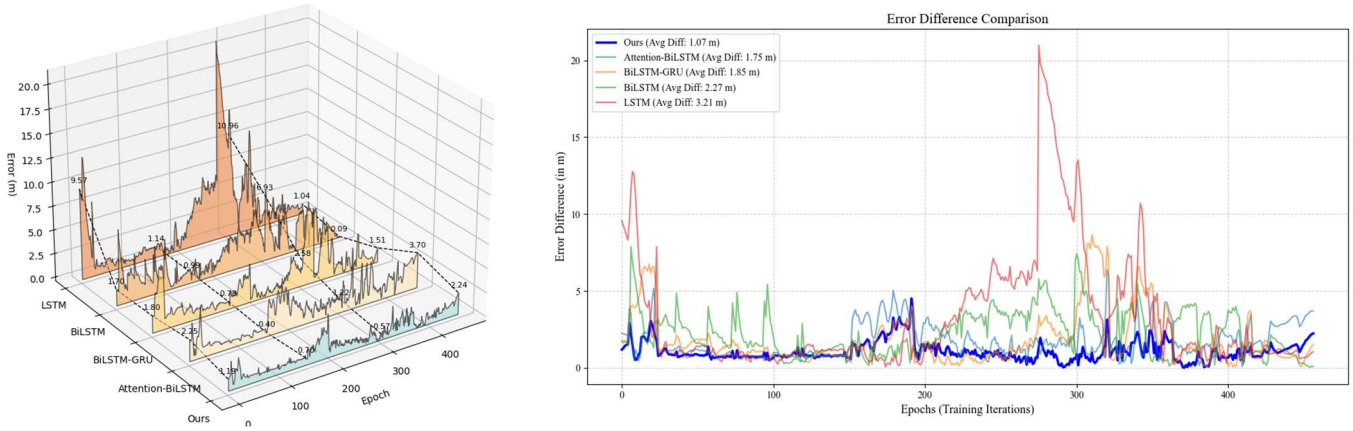


Figure 13. Comparison of latitude and longitude error results for different model training.

B. Model-Driven Ship Target Detection Model

In this experiment, several common and high-performing target detection algorithms were used, including Faster-RCNN, SSD, YOLOv8, and the improved model proposed in this study. The number of iterations was set to 300, the batch size was 16, and the initial learning rate was 0.001. The training results are shown in Table 3.

Table 3. Model Training Results.

Modal	Precision	Recall	mAP@0.5
Faster-RCNN	0.872	0.807	0.891
SSD	0.963	0.927	0.972
YOLOv8	0.952	0.943	0.975
OURS	0.966	0.958	0.981

Experimental results indicate that the proposed improved model outperforms other object detection models in key metrics such as precision, recall, and mean average precision (mAP@0.5). As shown in Figure 14, the comparison of training processes illustrates the trends of precision, recall, and mAP@0.5 across different models. The precision of the proposed model reaches 0.966, showing a notable advantage over SSD (0.963) and YOLOv8 (0.952), while Faster-RCNN falls behind with a precision of only 0.872. In terms of recall, the proposed model achieves a recall of 0.958, surpassing YOLOv8 (0.943), SSD (0.927), and Faster-RCNN (0.807). Regarding the comprehensive performance evaluation of object detection, the mAP@0.5 of the proposed model reaches 0.981, exceeding YOLOv8 (0.975), SSD

(0.972), and Faster-RCNN (0.891). The proposed model demonstrates rapid convergence during the early stages of training and maintains stability in the later stages, consistently outperforming all comparison models across all metrics.

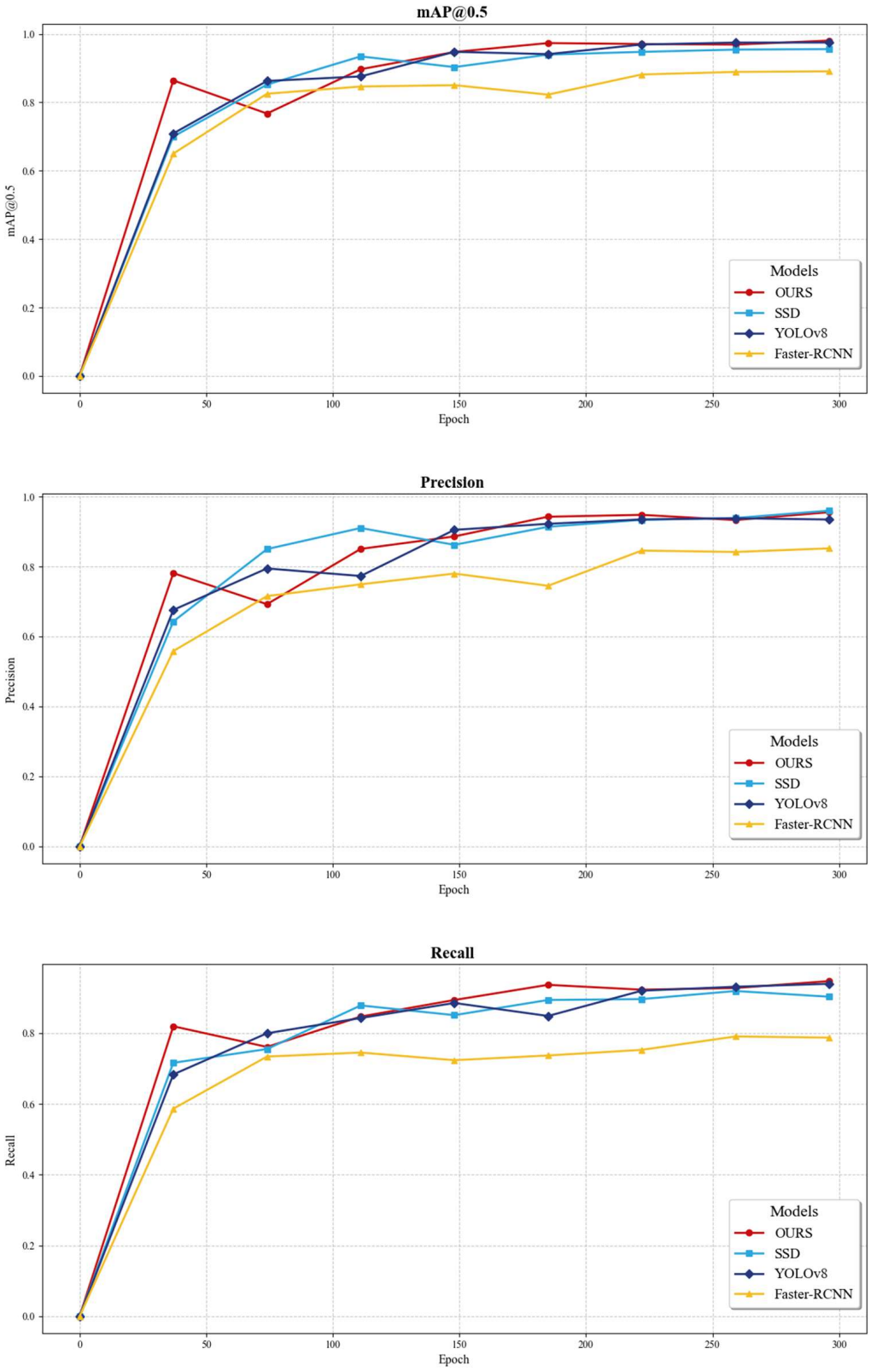


Figure 14. Comparison Chart of Accuracy, Recall Rate and Mean Average Precision in the Training Process of Different Models.

C. Fused Data Generation

In the heterogeneous data fusion method proposed in this paper, real-time video monitoring timestamps are used as the baseline to predict vessel positions from AIS trajectories and determine their future movement dynamics. First, AIS data and video data are aligned based on their timestamps to ensure that both data sources are matched within the same time frame. Next, using a region division method, the pixel coordinates obtained from video monitoring are mapped to actual geographic coordinates, thereby establishing consistency between AIS predicted trajectories and video trajectories, as illustrated in Figure 15. Based on the spatiotemporal location information and target identification numbers shown in Figure 15, logical matching is performed. This matching process yields the effect of spatiotemporal heterogeneous data fusion. Once successful matching is achieved, the dynamic features of AIS and video data are further integrated to generate a vessel dynamic vector containing comprehensive state information.

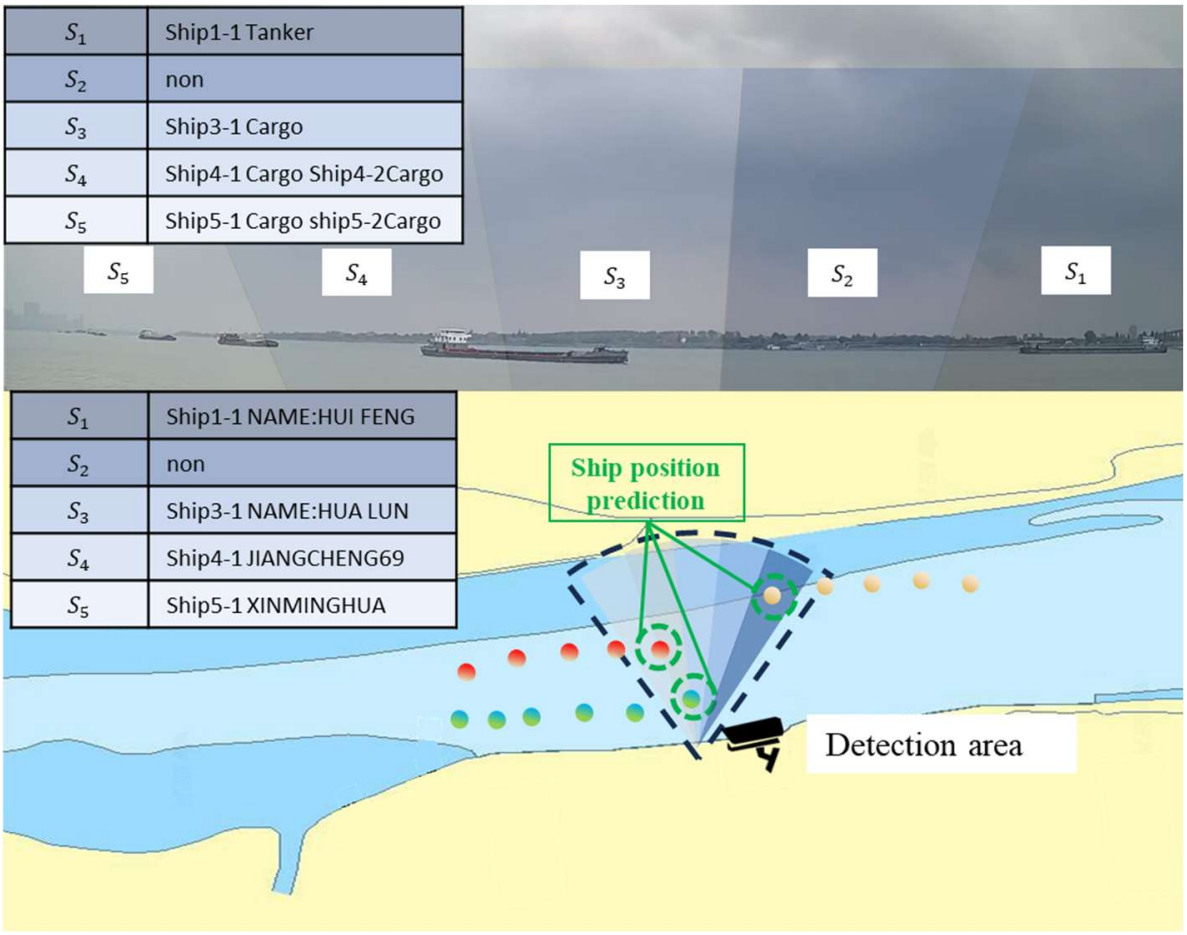


Figure 15. Fusion of Multi-source Spatiotemporal Heterogeneous Data.

In experiments involving data fusion over continuous time periods, this method consistently demonstrated strong performance. As shown in Figure 16, the fusion effects of AIS data and video data across six consecutive time frames are presented. The Figure 16 clearly illustrate the feature integration between AIS predicted trajectories and video detection results, as well as the information tracking outcomes across consecutive time frames.

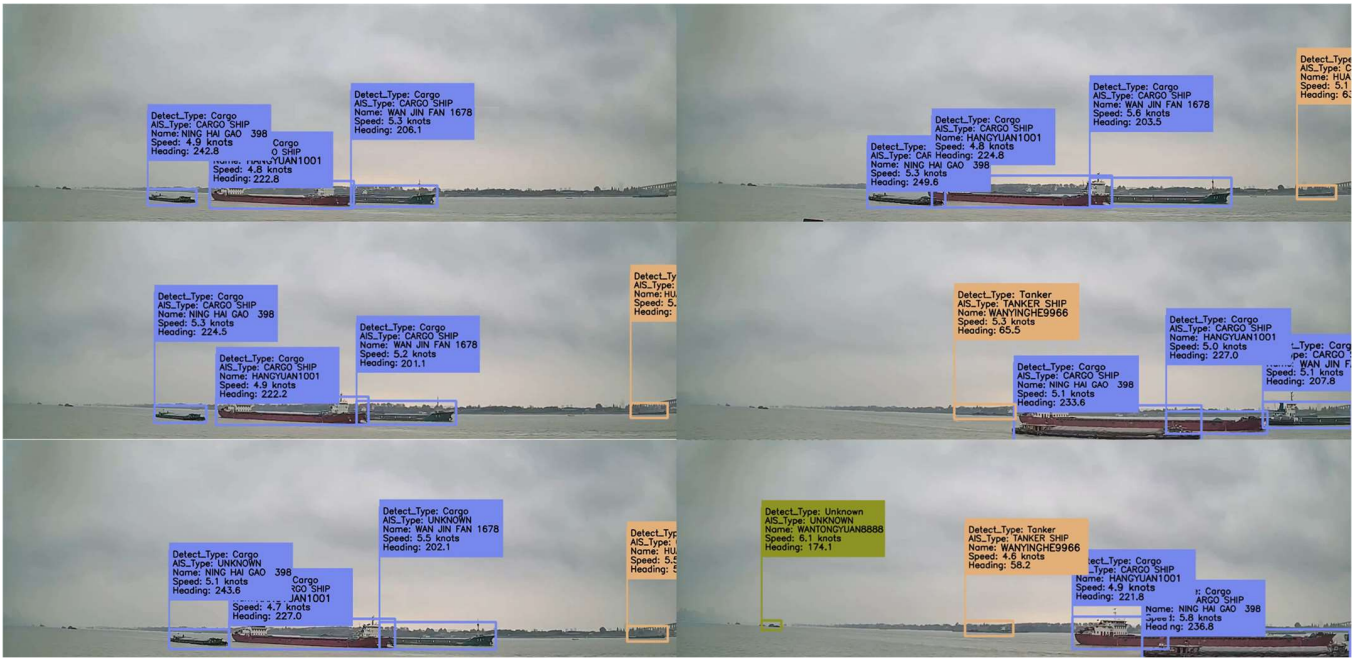


Figure 16. Visualization of the Fusion Effects Between Video Data and AIS Data.

5. Discussion

5.1. Challenges in Heterogeneous Data Fusion

This study addresses the intelligent management needs of inland waterways, focusing on the heterogeneity and real-time challenges of AIS data and video data. It proposes a data fusion method based on the HATM trajectory prediction model and single-stage object detection algorithm. The results show that the proposed method effectively enhances the precision of waterway monitoring and achieves the expected research goals. First, the HATM model significantly improves AIS data accuracy compared to traditional LSTM models. In complex waterway environments with drastic changes in vessel heading and speed, the HATM model captures vessel trajectory dynamics more accurately, reducing the adverse effects of AIS data latency and incompleteness on waterway monitoring. This improvement is critical for the real-time performance and accuracy of waterway management systems. Furthermore, by introducing single-stage object detection, the proposed method achieves real-time detection and tracking of vessel targets in video data, enhancing the accuracy and stability of visual monitoring. This complements the limitations of AIS data in localized waterway monitoring.

Compared to existing research, this study demonstrates significant technological and theoretical advantages and innovations. Currently, most data fusion methods rely on manual calibration or geometric coordinate transformations. Such methods are complex to deploy in engineering applications and have limited generalization capabilities. Changes in camera focal length or placement often disrupt the original geometric relationships. The proposed spatiotemporal matching method, based on region division and improved Kalman filtering, achieves fast and efficient data fusion without the need for coordinate transformation. This is accomplished through dynamic mapping between pixel regions and AIS data trajectories, offering better generalizability. Experimental results confirm the method's exceptional performance in terms of waterway monitoring accuracy and real-time capability, especially in multi-vessel water areas and complex waterways. This contribution not only simplifies the deployment of existing data fusion methods but also provides a scalable technical pathway for intelligent inland waterway management.

The primary innovation of this study lies in achieving refined sensing of localized waterways. By equipping buoys with high-resolution cameras and AIS receivers, a real-time, multi-dimensional data fusion system is constructed, significantly enhancing the monitoring precision and flexibility of localized waterways. Compared to land-based cameras, buoy-mounted sensing devices are closer to the center of the waterway, capable of capturing high-resolution detail and are more flexible to deploy, covering traditional blind spots in monitoring. The integration of AIS and visual data simultaneously provides information on vessel appearance and dynamic behavior, meeting the refined regulatory needs of localized waters. This offers robust technical support for real-time supervision and navigation safety in busy waterway segments.

In conclusion, this study achieves real-time perception of waterway conditions and dynamic prediction of vessel behavior through the efficient fusion of multi-source data. It effectively overcomes the limitations of existing waterway monitoring technologies, providing new theoretical foundations and practical solutions for intelligent inland waterway management.

5.2. Theoretical Significance

A. Theoretical Development of Multi-Source Data Fusion

One of the core contributions of this study is the proposal of a spatiotemporal matching-based multi-source data fusion method, which integrates Automatic Identification System (AIS) data and video surveillance data for intelligent waterway management. This method addresses the monitoring blind spots caused by AIS data delays and incompleteness in high-density traffic areas. Theoretically, the research advances the field of spatiotemporal data fusion, providing a new theoretical framework for waterway management. By introducing deep learning models, the proposed approach not only improves prediction accuracy but also enhances adaptability to complex scenarios, especially in environments with dynamic multi-ship behaviors. Therefore, this study provides an important theoretical foundation for future research in dynamic trajectory prediction and video target detection in the waterway management field.

B. Fusion Application of Deep Learning Models:

This study applies LSTM (Long Short-Term Memory) networks and single-stage object detection models, showcasing how these modern deep learning techniques can be used for spatiotemporal data fusion in waterway monitoring. Theoretically, this study demonstrates an innovative approach to cross-domain fusion, offering a new theoretical perspective for multimodal data joint learning. The research also shows that using the improved LSTM model can better capture dynamic changes in waterway vessels, significantly improving ship trajectory prediction accuracy, especially in complex inland waterways. Theoretically, this study's model and methods provide innovative references for multi-source information fusion and spatiotemporal data analysis in other fields.

C. Refined Management Theory for Intelligent Waterway Surveillance

This study not only proposes a new method for data fusion but also theoretically advances the refined management of intelligent surveillance systems. As the level of automation and intelligence in waterway monitoring continues to grow, it is increasingly essential to rely on multiple data sources to provide more precise information rather than relying on a single data source. This theoretical framework provides theoretical support for improving monitoring efficiency in complex waterway environments and dealing with the increasing density of vessel traffic. It also serves as a theoretical reference for interdisciplinary theoretical applications in future research.

5.3. Managerial Significance

A. Enhancing Waterway Management Efficiency

One of the contributions of this study is the development of a multi-source data fusion method, which can provide waterway authorities with more efficient surveillance tools. In complex waterway environments, traditional monitoring methods often face problems, such as data delays and incomplete information, leading to low surveillance efficiency. By integrating AIS and video surveillance data, management authorities can obtain more accurate and comprehensive real-time information on vessel dynamics, improving the efficiency of management. Waterway managers can make more accurate assessments of vessel navigation conditions and respond more quickly, thus improving the scientific and timely nature of decision making. This method is especially beneficial for monitoring busy waterways or critical segments, as it significantly reduces human error and optimizes the overall waterway management process.

B. Scalability of Intelligent Monitoring Systems

One of the managerial implications of this study is that the proposed method is not only applicable to inland waterways but, theoretically, can be extended to other water environments, including marine environments. With the continuous increase in waterway traffic and the development of intelligent surveillance technologies, authorities can deploy flexible monitoring systems based on the needs of different regions. For instance, in more complex marine environments, the spatiotemporal data fusion method proposed in this study is also adaptable, providing better management solutions for maritime traffic safety. This provides both theoretical justification and practical guidance for future upgrades and cross-regional applications of waterway management systems.

C. Intelligent Decision Support System

By integrating AIS and video data, the proposed multi-source data fusion method can provide more intelligent support for waterway management decisions. Managers can obtain more accurate traffic condition information in real time, enabling dynamic scheduling and incident prediction. Based on real-time feedback from the system, managers can automatically adjust waterway usage strategies, thus reducing the likelihood of accidents and optimizing the flow of traffic. The improved automation and intelligence of the system reduce human intervention, lower labor costs, and reduce the potential for operational errors, further enhancing management efficiency.

D. Formulating Refined Risk Management Strategies

An innovative aspect of this study is its proposed precise monitoring and risk prediction model, which provides a more detailed risk management framework for managers. Based on the results of multi-source data fusion, management authorities can analyze dynamic vessel behaviors, predict potential risk events, and take proactive measures. This is crucial for ensuring waterway safety, particularly in high-density traffic or complex environments. It enhances the ability to avoid potential collisions or accidents and improves the safety and overall management of waterway systems.

5.4. Limitations in Practical Applications

In this study, to validate the effectiveness of the proposed waterway monitoring method based on multi-source heterogeneous data fusion, a one-year field test was conducted on the Yangtze River's Wuhu waterway. Due to its unique traffic conditions, frequent vessel encounters, and high volumes of container and bulk cargo ships, the Wuhu segment is an ideal testing scenario for waterway monitoring technologies.

However, despite achieving significant technical outcomes, certain limitations remain in practical applications, particularly regarding weather conditions, sustained equipment operation, and data fusion accuracy. These specific issues are outlined below:

A. Impact of Weather Conditions on Object Detection Models

In the practical application of the Wuhu intelligent buoy project, weather conditions have proven to be a critical factor affecting visual detection performance. Adverse weather such as low light, backlighting, and heavy fog significantly impacts the clarity and accuracy of video monitoring. For example, during dawn and dusk, poor lighting conditions in the waterway, especially when cameras face light sources, often result in backlighting, making it difficult to clearly identify vessel targets. This affects the ability of object detection algorithms to capture vessel features, particularly for distant vessels. Heavy fog severely limits the camera's field of view, causing distant targets to appear blurry and incomplete in the video data. This reduces the accuracy of object detection and, consequently, impacts the spatiotemporal alignment between AIS data and video data, further diminishing the effectiveness of data fusion.

B. Impact of Weather Conditions on Power Supply for Intelligent Buoy Systems

In practical deployment, intelligent buoys are typically installed on lightships within the Yangtze River waterway. These buoys integrate multiple modules, including hydrometeorological sensors, visual and AIS systems, VHF voice response, and 5G network communication. As a result, their energy supply relies heavily on solar panels. Prolonged cloudy weather or consistently low light conditions can reduce battery charging efficiency, causing power shortages during nighttime or unfavorable weather conditions. This not only affects the normal operation of buoy equipment but also impacts the system's data collection and processing capabilities.

To address these issues in practical engineering applications, the following measures can be implemented:

1. Low-Power Devices and Algorithm Optimization: By introducing additional environmental adaptability modules, object detection algorithms can better adapt to video data under various weather and lighting conditions. For instance, image enhancement techniques tailored to low-light conditions, powered by deep learning models, can enhance targets in dim environments, improving the detection accuracy of vessel targets under weak light. To reduce energy consumption, part of the computational tasks can be offloaded to shore-based servers, thereby alleviating the computational burden on buoy devices. Employing lightweight deep learning algorithms can further lower the computational complexity of object detection and data fusion, extending the operational lifecycle of the intelligent buoy system.
2. Optimizing the Solar Power Management System: By combining solar panels with backup batteries, the system can ensure basic buoy operation even under low sunlight conditions. Additionally, a reinforcement learning-based resource optimization algorithm can be designed to allocate resources efficiently, thereby extending the continuous operational time of the device. When the device's power drops below a certain threshold, the system will automatically issue a maintenance alert and activate backup batteries for emergency power supply.

5.5. Future Work

Based on the limitations and findings of the current research and practical applications, future studies can be conducted in the following directions to further enhance inland waterway monitoring and intelligent management.

A. Collection and Utilization of More Aligned Multi-Source Heterogeneous Data

Current research has demonstrated that multi-source heterogeneous data fusion effectively improves the precision of waterway monitoring, but the scope and types of data coverage remain limited. Future work will focus on expanding data collection to include more waterway segments and a wider range of environmental conditions, gathering additional video data and corresponding AIS data. This will be combined with local knowledge bases to enable deeper integration. By learning vessel operating patterns from the data along the Yangtze River, predictions for vessels without active AIS devices can be optimized, further enhancing comprehensive awareness of waterway dynamics.

B. Improving Multi-Source Data Fusion Algorithms

To address the complexity of heterogeneous data matching and the insufficient optimization of real-time performance in current research, future efforts will focus on developing more efficient fusion algorithms. For instance, dynamic data association methods based on attention mechanisms and self-supervised learning will be explored. These methods can be combined with reinforcement learning to dynamically adjust weight distribution among multiple data sources, enabling the system to achieve adaptive fusion based on actual application scenarios and improving the model's real-time performance and robustness. Additionally, integrating more types of sensor data will be considered to further enhance the system's refined perception capabilities. Efforts will also be made to improve the model's adaptability and robustness to extreme environments such as strong winds and heavy rain.

C. Promoting the Development of a Digital Shipping Ecosystem

Research on multi-modal fusion compensation technologies will be prioritized. In low-visibility environments, the combined use of AIS and local knowledge bases can address the shortcomings of visual detection. To tackle the challenges of insufficient power supply for intelligent buoys during prolonged rainy weather, low-power algorithms and modular designs will be explored to optimize the energy efficiency of buoy hardware. By establishing an integrated digital shipping ecosystem that spans from data collection to decision support, an end-to-end intelligent waterway management solution can be developed. This will provide robust technical support for the construction of "digital waterways".

6. Conclusions

This study proposes an intelligent regulation method based on heterogeneous data fusion to address the challenges of real-time and accurate monitoring in complex waterway environments with dense vessel traffic. By integrating AIS data with video monitoring and employing an improved LSTM network for vessel trajectory prediction, along with a single-stage object detection model for real-time vessel tracking, the method significantly enhances the accuracy of vessel predictions and the real-time performance of video-based detection. A spatiotemporal matching approach, combining regional division and dynamic-weighted Kalman filtering, simplifies data fusion and improves system responsiveness and accuracy, effectively compensating for AIS data latency and visualization errors. Experimental results demonstrate the method's effectiveness in dense inland waterway environments, offering a strong foundation for the development of digital waterways and intelligent shipping.

While the method shows promise in the Wuhu section, limitations remain, such as the impact of weather on video data quality, AIS data transmission delays, and energy shortages in smart buoys under adverse conditions. These challenges highlight the need for further optimization, particularly in data processing and energy management under harsh environments. Future research should focus on enhancing scalability and adaptability, incorporating additional data sources like radar and environmental sensors, exploring deep

learning-based image enhancement for harsh weather conditions, and integrating low-power algorithms to improve the system's robustness and ensure continuous operation.

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

Funding: This work is supported by the Guangxi Key Research and Development Plan (Grant No. GUIKE AB22080107), this work is also supported by the International Association of Maritime Universities research project(Grant No.20240201).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Due to confidentiality agreements with the data provider, this dataset is partially published on GitHub.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIS	Automatic Identification System
CNN	Convolutional Neural Networks
KF	Kalman Filtering
GM	Gray Model
SVR	Support Vector Regression
LSTM	Long Short-Term Memory
SIFT	Scale-Invariant Feature Transform
HOG	Histogram of Oriented Gradients
RCNN	Region-based Convolutional Neural Networks
Fast-RCNN	Fast Region-based Convolutional Network
Faster-RCNN	Faster Region-based Convolutional Network
SSD	Single Shot MultiBox Detector
GRU	Gated Recurrent Units

References

1. UN Trade and Development. *Review of Maritime Transport 2024—Navigating Maritime Chokepoints (Overview)*; United Nations: New York, NY, USA, 2024.
2. Fang, X.; Zhang, Y.; Yang, J.; Zhan, G. An Evaluation of Marine Economy Sustainable Development and the Ramifications of Digital Technologies in China Coastal Regions. *Econ. Anal. Policy* **2024**, *82*, 554–570. [CrossRef]
3. An, J. Maritime Logistics and Digital Transformation with Big Data: Review and Research Trend. *MABR* **2024**, *9*, 229–242. [CrossRef]
4. Luin, B.; Al-Mansour, F.; Perkovič, M. Optimization of Shipping Routes with AIS Data. *Therm. Sci. Eng. Prog.* **2024**, *56*, 103042. [CrossRef]
5. Nešković, Đ.; Marković, M.; Kašca, L.; Drajić, D. Analysis of Maritime Positioning Techniques for Use in Maritime Internet of Things Systems. In Proceedings of the 2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 20–22 March 2024; pp. 1–5.
6. Briguglio, G.; Crupi, V. Review on Sensors for Sustainable and Safe Maritime Mobility. *J. Mar. Sci. Eng.* **2024**, *12*, 353. [CrossRef]
7. Yang, C.-H.; Wu, C.-H.; Shao, J.-C.; Wang, Y.-C.; Hsieh, C.-M. AIS-Based Intelligent Vessel Trajectory Prediction Using Bi-LSTM. *IEEE Access* **2022**, *10*, 24302–24315. [CrossRef]
8. Huang, F.; Liu, Z.; Li, X.; Mou, F.; Li, P.; Fan, Z. Vessel Trajectory Prediction Based on AIS Data: Dual-Path Spatial–Temporal Attention Network with Multi-Attribute Information. *J. Mar. Sci. Eng.* **2024**, *12*, 2031. [CrossRef]

9. Zhang, X.; Fu, X.; Xiao, Z.; Xu, H.; Qin, Z. Vessel Trajectory Prediction in Maritime Transportation: Current Approaches and Beyond. *IEEE Trans. Intell. Transport. Syst.* **2022**, *23*, 19980–19998. [CrossRef]
10. Jiang, B.; Wu, X.; Tian, X.; Jin, Y.; Wang, S. Proposal of Innovative Methods for Computer Vision Techniques in Maritime Sector. *Appl. Sci.* **2024**, *14*, 7126. [CrossRef]
11. Gawai, N.S.; Rojatkar, D.D.V. Literature Review on Ship Detection Methods Using Satellite-Borne Synthetic Aperture Radar for Maritime Surveillance. *Grenze Int. J. Eng. Technol.* **2024**, *10 Pt 2*, 1774.
12. Yasir, M.; Shanwei, L.; Mingming, X.; Jianhua, W.; Nazir, S.; Islam, Q.U.; Dang, K.B. SwinYOLOv7: Robust Ship Detection in Complex Synthetic Aperture Radar Images. *Appl. Soft Comput.* **2024**, *160*, 111704. [CrossRef]
13. Zhang, Z.; Zhang, L.; Wu, J.; Guo, W. Optical and Synthetic Aperture Radar Image Fusion for Ship Detection and Recognition: Current State, Challenges, and Future Prospects. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 132–168. [CrossRef]
14. Fei, P.; Geng, X.; Chu, X.; Luo, J. A Inland Waterway Monitoring Virtual-GIS System Based on Multi Heterogeneous Navigation Data Fusion. In Proceedings of the 2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST), Guangzhou, China, 10–12 December 2021.
15. Perera, L.P.; Soares, C.G. Ocean Vessel Trajectory Estimation and Prediction Based on Extended Kalman Filter. In Proceedings of the ADAPTIVE 2010: The Second International Conference on Adaptive and Self-Adaptive Systems and Applications, Lisbon, Portugal, 21–26 November 2010.
16. Yin, J.-C.; Zou, Z.-J.; Xu, F.; Wang, N.-N. Online Ship Roll Motion Prediction Based on Grey Sequential Extreme Learning Machine. *Neurocomputing* **2014**, *129*, 168–174. [CrossRef]
17. Kawan, B.; Wang, H.; Li, G.; Chhantyal, K. Data-Driven Modeling of Ship Motion Prediction Based on Support Vector Regression. In Proceedings of the 58th Conference on Simulation and Modelling (SIMS 58), Reykjavik, Iceland, 25–27 September 2017; pp. 350–354.
18. Fossen, S.; Fossen, T.I. Extended Kalman Filter Design and Motion Prediction of Ships Using Live Automatic Identification System (AIS) Data. In Proceedings of the 2018 2nd European Conference on Electrical Engineering and Computer Science (EECS), Bern, Switzerland, 20–22 December 2018; pp. 464–470.
19. Tang, H.; Yin, Y.; Shen, H. A Model for Vessel Trajectory Prediction Based on Long Short-Term Memory Neural Network. *J. Mar. Eng. Technol.* **2022**, *21*, 136–145. [CrossRef]
20. Zhang, Z.; Ni, G.; Xu, Y. Ship Trajectory Prediction Based on LSTM Neural Network. In Proceedings of the 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 12–14 June 2020; pp. 1356–1364.
21. Gao, D.; Zhu, Y.; Zhang, J.; He, Y.; Yan, K.; Yan, B. A Novel MP-LSTM Method for Ship Trajectory Prediction Based on AIS Data. *Ocean Eng.* **2021**, *228*, 108956. [CrossRef]
22. Zhang, S.; Wang, L.; Zhu, M.; Chen, S.; Zhang, H.; Zeng, Z. A Bi-Directional LSTM Ship Trajectory Prediction Method Based on Attention Mechanism. In Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 March 2021; pp. 1987–1993.
23. Zhao, J.; Yan, Z.; Zhou, Z.; Chen, X.; Wu, B.; Wang, S. A Ship Trajectory Prediction Method Based on GAT and LSTM. *Ocean Eng.* **2023**, *289*, 116159. [CrossRef]
24. Lowe, D.G. Object Recognition from Local Scale-Invariant Features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 1150–1157.
25. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
26. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
27. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
28. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
29. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
30. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
31. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the Gap Between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 9756–9765.

32. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9905, pp. 21–37.
33. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *arXiv* **2018**, arXiv:1708.02002.
34. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as Points. *arXiv* **2019**, arXiv:1904.07850.
35. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
36. Ting, L.; Baijun, Z.; Yongsheng, Z.; Shun, Y. Ship Detection Algorithm Based on Improved YOLO V5. In Proceedings of the 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE), Dalian, China, 15–17 July 2021; pp. 483–487.
37. Sun, Z.; Leng, X.; Lei, Y.; Xiong, B.; Ji, K.; Kuang, G. BiFA-YOLO: A Novel YOLO-Based Method for Arbitrary-Oriented Ship Detection in High-Resolution SAR Images. *Remote Sens.* **2021**, *13*, 4209. [CrossRef]
38. Guo, Y.; Chen, S.; Zhan, R.; Wang, W.; Zhang, J. LMSD-YOLO: A Lightweight YOLO Algorithm for Multi-Scale SAR Ship Detection. *Remote Sens.* **2022**, *14*, 4801. [CrossRef]
39. Cai, S.; Meng, H.; Wu, J. FE-YOLO: YOLO Ship Detection Algorithm Based on Feature Fusion and Feature Enhancement. *J. Real-Time Image Process.* **2024**, *21*, 61. [CrossRef]
40. Man, W.; Zhiyong, L. The Information Fusion Based on AIS and Video Data. In Proceedings of the 2016 5th International Conference on Computer Science and Network Technology (ICCSNT), Changchun, China, 10–11 December 2016; pp. 336–339.
41. Bloisi, D.D.; Previtali, F.; Pennisi, A.; Nardi, D.; Fiorini, M. Enhancing Automatic Maritime Surveillance Systems with Visual Information. *IEEE Trans. Intell. Transport. Syst.* **2017**, *18*, 824–833. [CrossRef]
42. Lu, Y.; Ma, H.; Smart, E.; Vuksanovic, B.; Chiverton, J.; Prabhu, S.R.; Glaister, M.; Dunston, E.; Hancock, C. Fusion of Camera-Based Vessel Detection and AIS for Maritime Surveillance. In Proceedings of the 2021 26th International Conference on Automation and Computing (ICAC), Portsmouth, UK, 2–4 September 2021; pp. 1–6.
43. Huang, Z.; Hu, Q.; Mei, Q.; Yang, C.; Wu, Z. Identity Recognition on Waterways: A Novel Ship Information Tracking Method Based on Multimodal Data. *J. Navig.* **2021**, *74*, 1336–1352. [CrossRef]
44. Liu, R.W.; Guo, Y.; Nie, J.; Hu, Q.; Xiong, Z.; Yu, H.; Guizani, M. Intelligent Edge-Enabled Efficient Multi-Source Data Fusion for Autonomous Surface Vehicles in Maritime Internet of Things. *IEEE Trans. Green Commun. Netw.* **2022**, *6*, 1574–1587. [CrossRef]
45. Qu, J.; Guo, Y.; Lu, Y.; Zhu, F.; Huan, Y.; Liu, R.W. Intelligent Maritime Surveillance Framework Driven by Fusion of Camera-Based Vessel Detection and AIS Data. In Proceedings of the 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), Macau, China, 8–12 October 2022; pp. 2280–2285.
46. Gülsoylu, E.; Koch, P.; Yıldız, M.; Constapel, M.; Kelm, A.P. Image and AIS Data Fusion Technique for Maritime Computer Vision Applications. In Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa, HI, USA, 1–6 January 2024; pp. 859–868.
47. Qu, J.; Liu, R.W.; Guo, Y.; Lu, Y.; Su, J.; Li, P. Improving Maritime Traffic Surveillance in Inland Waterways Using the Robust Fusion of AIS and Visual Data. *Ocean Eng.* **2023**, *275*, 114198. [CrossRef]
48. Guo, Y.; Liu, R.W.; Qu, J.; Lu, Y.; Zhu, F.; Lv, Y. Asynchronous Trajectory Matching-Based Multimodal Maritime Data Fusion for Vessel Traffic Surveillance in Inland Waterways. *IEEE Trans. Intell. Transport. Syst.* **2023**, *24*, 12779–12792. [CrossRef]
49. Ding, H.; Weng, J. A Robust Assessment of Inland Waterway Collision Risk Based on AIS and Visual Data Fusion. *Ocean Eng.* **2024**, *307*, 118242. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.