Multi-Ship Future Interaction Trajectory Prediction via Pre-Initializer Diffusion Model

Kun Ma
Harbin Engineering University
Harbin, China
makun@hrbeu.edu.cn

Qilong Han
Harbin Engineering University
Harbin, China
hanqilong@hrbeu.edu.cn

Jingzheng Yao*
Harbin Engineering University
Harbin, China
yaojingzheng@hrbeu.edu.cn

ABSTRACT

Real-time stochastic multi-ship trajectory modeling is crucial for maritime safety. However, it remains challenging due to the uncertainty of dynamic vessel intentions and their complex interactions. Most existing studies rely on deterministic social data from historical time steps for modeling, which often fail to capture the future states of interacting ships, leading to unrealistic trajectory overlaps. Recent research has demonstrated that diffusion models excel in trajectory prediction due to their high generation quality, training stability, and diversity. However, their slow sampling speed limits real-time perception in maritime environments, as generating high-quality trajectories typically requires hundreds of denoising steps. To address these challenges, we propose a Multi-Ship Future interaction trajectory prediction approach based on a Pre-initializer Diffusion model (MFPD). By training a parameterized pre-initializer to directly learn the joint distribution of multiple denoising steps in the reverse diffusion process, our method significantly reduces the time cost of denoising while retaining only a few steps for fine-tuning the distribution. Specifically, in addition to encoding historical trajectory information and social interactions as state embeddings, we also incorporate future trajectory and multimodal maritime environmental information as input condition embeddings to fully capture potential future interactions and environmental features. Experimental results demonstrate that the proposed model significantly improves performance on two real-world datasets while greatly accelerating the sampling speed, demonstrating the superiority in real-world maritime environments.

KEYWORDS

Ship Automatic Identification System; Multi-ship; Trajectory Prediction; Diffusion Models; Parameterized Pre-initializer

ACM Reference Format:
Kun Ma, Qilong Han, and Jingzheng Yao*. 2025. Multi-Ship Future Interaction Trajectory Prediction via Pre-Initializer Diffusion Model. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 10 pages.

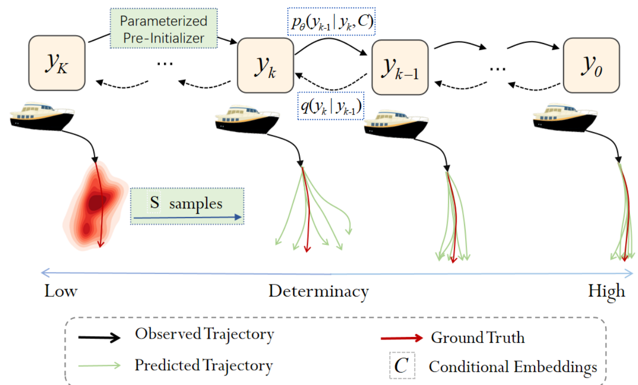

Figure 1: Illustration of the MFPD diffusion and reverse process with a pre-initializer. we first use a parameterized pre-initializer to estimate the distribution and generate samples, and then perform few denoising process for fine-tuning.

1 INTRODUCTION

Maritime transportation is a cornerstone of global trade, with over 90% of goods transported by sea. As global trade continues to grow, so does the density of maritime traffic, leading to increased risks, especially in congested waterways and adverse weather conditions. Ensuring maritime safety is crucial, particularly in situations where interactions between vessels are inevitable. Accurate ship trajectory prediction is a vital tool for avoiding collisions, improving traffic flow, and enhancing operational efficiency.The Automatic Identification System (AIS)[42, 48] plays a pivotal role in maritime traffic monitoring by providing real-time data on ship positions, speeds, and courses. Leveraging AIS data for ship motion analysis and trajectory prediction is essential for improving safety and assisting ship crews in making timely decisions to reduce the risk of collisions. Although human operators are experienced, they are susceptible to errors under stress, fatigue, or marine environments with low visibility, such as rainy and foggy conditions. Automated systems, by contrast, can process large volumes of trajectory data rapidly and consistently, offering more accurate and timely predictions, which are crucial for modern maritime navigation and safety.

However, predicting future ship trajectories remains challenging due to the inherent uncertainty in ship behavior, such as the ability to dynamically adjust speed and course. Previous research on generative models has attempted to address this uncertainty. Some approaches utilize Generative Adversarial Networks (GANs) [14, 27] or Conditional Variational Autoencoders (CVAE) [10, 19, 47] to model multimodal future trajectories. GANs, while useful, often suffer from instability during adversarial training and produce limited diversity in generated trajectories. Although CVAEs can

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

generate diverse trajectories, it often lacks accuracy. A key limitation of diffusion models is the slow sampling speed during the reverse diffusion process, which is inherently reliant on a Markov process[16, 38]. This reliance results in prolonged inference times in real-time prediction environments, as generating high-quality samples typically requires hundreds of denoising steps. For instance, experiments on the Weihai Port dataset show that a standard diffusion model requires approximately 200 denoising steps to achieve satisfactory prediction performance based on historical social trajectory data and multimodal future trajectory and environmental information, with each prediction taking around 3110 milliseconds. This speed bottleneck presents significant challenges, especially in real-time prediction scenarios and complex maritime environments.

Furthermore, many related studies[4, 7, 22] have overlooked several critical issues in modeling maritime environments. First, they often focus solely on modeling historical trajectory data, neglecting the benefits of multimodal modeling. Second, they primarily concentrate on predicting individual or multiple ship trajectories based on temporal and social relationships in the historical data, overlooking potential interactions between future trajectories of multiple ships. As a result, these approaches fail to capture the true dynamics of the maritime environment. Finally, previous trajectory prediction models have not fully leveraged the geometric properties of trajectories. As sequences of two-dimensional positions, ship trajectories exhibit geometric invariance. In this study, we consider the invariance and equivariance properties of SO(2)[8, 46] with the aim of enhancing the model's generalization ability under a limited ship trajectory dataset. Invariance implies that the model's output remains unaffected when the input data is rotated by a certain angle, meaning the model's predictions should remain consistent under any rotation. Equivariance, on the other hand, refers to the property where the model's output transforms according to the same rules when the input data undergoes a specific transformation.

To address the above challenges, our proposed Multi-Ship Future Interaction Trajectory Prediction (MFPD) model follows the same forward diffusion process as standard diffusion models but introduces a novel approach to the reverse denoising process. By training a parameterized pre-initializer, we directly learn the joint distribution of multiple denoising steps in the reverse diffusion process using state embeddings and conditional embeddings C derived from temporal-social information. This approach significantly reduces the time cost of denoising while retaining only a few denoising steps for fine-tuning the distribution, as shown in Figure 1. To ensure sufficient prediction diversity, we sample multiple correlated prediction trajectories to distribute path diversity, rather than relying on independent and identically distributed samples.The proposed method does not strictly depend on Markov process-based sampling strategies, providing a novel solution to overcome the inference speed bottleneck in ship trajectory prediction.

More specifically, MFPD introduces a novel encoder-decoder architecture. In addition to encoding historical trajectory information and social interactions into state embeddings, the encoder also incorporates future trajectories and multimodal maritime environment data as conditional inputs. This allows the model to fully capture potential future interactions and environmental features. The parameterized pre-initializer is decomposed into three trainable components: mean trajectory, variance, and multiple correlated prediction samples. To train these components, the multimodal encoder generates state and conditional embeddings, which are used to produce accurate estimates. The decoder follows a standard diffusion model structure to fine-tune the distribution generated by the pre-initializer. Furthermore, it employs a Transformer-based[43] architecture to effectively capture temporal dependencies within the trajectories.

The primary contributions of this paper are as follows:

- We propose a novel stochastic trajectory prediction framework (MFPD) as shown in Figure 2, which accelerates the denoising diffusion process by training a parameterized pre-initializer to sample initial correlated trajectories. It achieves precise and diverse predictions with fast inference speed, making it suitable for real-time multi-ship trajectory modeling.
- In addition to encoding historical trajectory temporal information and social interactions as state embeddings, we incorporate future trajectory data and multimodal maritime environmental information as conditional embeddings. This allows the model to fully capture potential future interactions between multiple vessels and the associated environmental features.
- We conduct extensive experiments on two real-world ship trajectory datasets. The results demonstrate that our method significantly improves prediction performance while reducing inference time by 20 times, showcasing its advantage in real-time predictions within maritime scenarios.

2 RELATED WORK

2.1 Discriminative Models

Discriminative models are a class of machine learning models that characterize conditional distributions and learn trajectory motion patterns from historical trajectories. Previous approaches, such as nonlinear filtering, Support Vector Machines (SVM)[44], Recurrent Neural Networks (RNN)[4], and Long Short-Term Memory (LSTM) networks[12, 23, 40], have primarily focused on modeling historical trajectories. Additionally, social-aware models like Social-LSTM[1, 22] and Spatio-Temporal Graph convolutional neural networks (STGCN)[31] have been widely applied to ship trajectory prediction. While these methods effectively capture temporal dependencies, they struggle to fully represent the inherent uncertainty and multimodal nature of ship trajectories. Specifically, Support Vector Regression (SVR) models [44] have been utilized to map input historical trajectory data to outputs, establishing a relationship between observed values and future trajectories. In the BP method [50], longitude, latitude, vessel heading, and speed serve as inputs to Backpropagation (BP) neural networks for vessel trajectory prediction. Long Short-Term Memory (LSTM) networks [40] have also been employed to predict vessel trajectories, with results validated using real AIS data.

Recent trajectory prediction approaches have focused on modeling complex social interactions and risks. Social-LSTM [1] introduces a social pooling layer to aggregate neighborhood interaction information, using sophisticated networks to model social interactions. Spatio-temporal graph models [9, 21] adopt spatio-temporal graph convolutional neural networks to jointly model

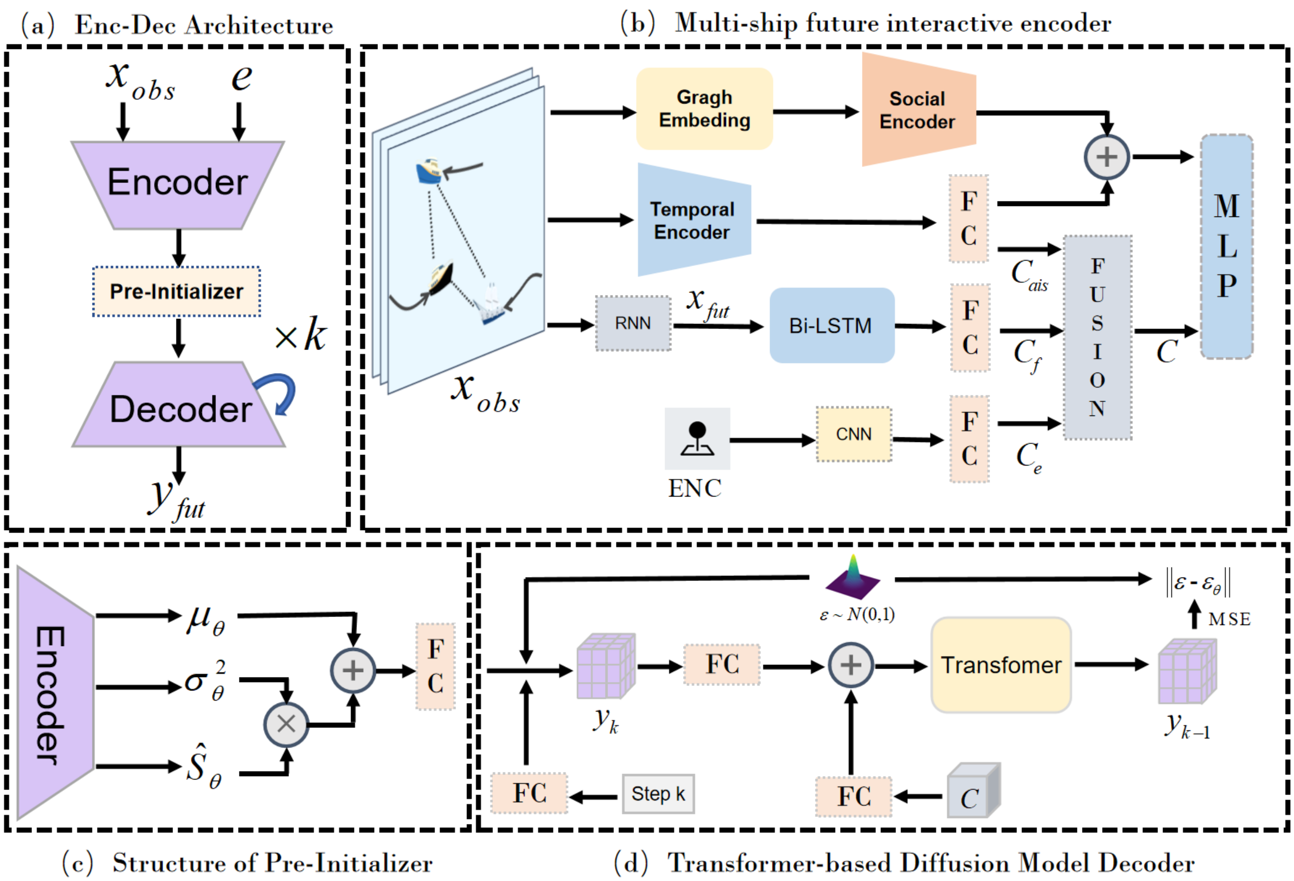

Figure 2: The proposed MFPD model consists of an encoder and a decoder (a). The encoder integrates historical trajectory data and maritime environment information to generate trajectory features, as shown in (b). In addition to encoding temporal information and social interactions, the model incorporates future trajectory data and multimodal marine environment information as conditional embeddings to enhance representation. (c) illustrates the structure of the pre-initializer, which reparameterizes to generate relevant samples. Finally, (d) presents the Transformer-based diffusion model decoder, which refines the pre-initialized samples and iteratively generates predicted trajectories over k steps.

temporal clues and social interaction behaviors. However, some studies [5, 29] suggest that methods analyzing social interactions may exhibit biases, as evidenced in empirical results.

2.2 Generative Models

2.2.1 General Generative Models. Generative models characterize joint distributions to capture the distribution of trajectory data, enabling the generation of new trajectories. Addressing the inherent uncertainty in vessel motion, a model proposed by [34] uses Gaussian Processes (GP) to represent trajectory uncertainty as continuous probability distributions. The rapid development of Social-GAN in autonomous driving has inspired trajectory prediction in vessel navigation. Recent studies [17, 35] explore the use of Generative Adversarial Networks (GAN) and Social-GAN[15] to model multimodalities with variable noise. GANs hold promise for overcoming the challenges of complex multimodal data by learning the underlying distribution of training data. For instance, Col-GAN [24] can generate multiple plausible trajectories while minimizing potential collisions through the use of a collision discriminator.

Variational Autoencoder (VAE) structures [6, 37] have also been employed, utilizing variational inference to learn distributions and introducing conditional information during trajectory generation. This results in more flexible and adaptable trajectories under specific conditions. Despite significant progress made by these stochastic prediction methods[25, 45], they exhibit limitations such as unstable training processes or anomalous trajectory generation.

2.2.2 Denoising Diffusion Models. Diffusion models have attracted widespread attention due to their powerful representation capabilities, diversity, and stable training process in generative modeling[32, 33, 41]. To capture multimodal trajectory distributions, recent efforts have turned towards stochastic trajectory prediction using Denoising Diffusion Probabilistic Models (DDPM) [16], inspired by non-equilibrium thermodynamics[38]. In the trajectory prediction domain [13, 28], diffusion models learn a parameterized Markov

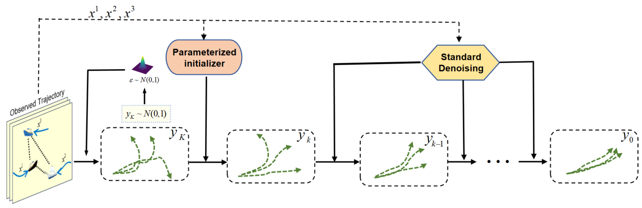

Figure 3: Inference phase of the proposed model MFPD.

chain through a denoising process, gradually transitioning from an initial noisy distribution to specific data distributions.

To expedite the sampling process, DDIM [39] and LED [30] predict on raw data and estimate the position of the next expected step based on non-Markovian processes. Progressive distillation (PD) [36] has been applied to the denoising step of deterministic diffusion sampling, accelerating sampling by iteratively refining the process. Most recent studies have focused on designing efficient samplers to accelerate the inference process of diffusion models, such as DPM-Solver++[26] and karras[18]. However, these samplers often compromise the model's expressive power or require complex tuning, making them difficult to generalize across diverse scenarios. Furthermore, many of these samplers rely on specific diffusion process structures, limiting their flexibility and providing only modest improvements in reducing time consumption. In our work, we propose training a parameterized pre-initializer to directly learn the multimodal distribution of trajectories for sampling, followed by standard denoising steps to achieve faster inference speeds, as shown in Figure 3.

3 METHOD

3.1 Problem Formulation

Supposing an observed vessel can receive AIS broadcast signals from n surrounding vessels within the effective interaction range, the objective of vessel trajectory prediction is to generate credible trajectories for a future time period based on the observed vessel's own motion trajectory and the past motion trajectories of the n surrounding vessels. First, lat, lon, SOG and COG represent the latitude, longitude, speed over ground, and course over ground at timestamp t, respectively. Additional features such as heading, type, status, and draught serve as independent ais conditions C_{ais} for the model. These features, indicating the ship's current heading direction, ship type (e.g., container ship, LNG), operational status (e.g., at anchor, under way), and draught (ship's depth in water), help the model better understand the ship's behavior in varying scenarios, facilitating more precise trajectory predictions. All features are encoded into real numbers during preprocessing. That is, input $x^i = \{s^i{}_t \in \mathbb{R}^8 | t = 0, 1, \cdots, m\}, i \in \{0, 1, \cdots, n\}$, Where $s^i{}_t$ is represented by an 8-dimensional feature vector $s^i{}_t = [lat, lon, SOG, COG, heading, type, status, draught]$ at timestamp t, the position of the vessel on the water surface at time t, m is the observation duration of the trajectory. The predicted future trajectories are written as $y^i = \{s^i{}_t \in \mathbb{R}^4 | t = m+1, m+2, \cdots m+l\}, i \in \{0, 1, \cdots, n\}$, where l refers to the time length of the predicted. Each predicted point $s^i{}_t$ is a 4-dimensional feature vector: $s^i{}_t = [lat, lon, SOG, COG]$.

Given the high uncertainty associated with future trajectories, practical applications often involve predicting multiple trajectories to strike a balance between the determinacy and diversity of generated trajectories. This paper delves into stochastic trajectory prediction, where the aim is to forecast the distribution of future trajectories rather than a single trajectory. The objective of stochastic trajectory prediction is to train a predictive model $M_\theta()$ with parameters θ to generate a distribution $\text{Pred}_\theta = M_\theta(x^0, x^1 \cdots x^n)$. The result of S sampling based on the distribution of the prediction model is $\hat{y} = \{y^1, y^2, \cdots, y^S\}$. In order to ensure that at least one sampled trajectory approximates the ground-truth trajectory y_{gt}, the training objective is:

$$\text{Pred}_\theta = \min_\theta \min_{y^i \in \hat{y}} Distance(y^i, y_{gt}), i \in \{1, 2, \cdots, S\}. \tag{1}$$

3.2 Environment Embedding and Encoding Future Interactive Motion

Modern ship sensor suites provide far more information than just the observation of surrounding ship trajectories. Notably, various data sources such as radar, sonar, weather sensors, and electronic nautical charts assist in positioning and navigation. Depending on sensor availability and task relevance, we utilize Electronic Nautical Charts (ENC) as a key reference for environmental perception. These charts are encoded with semantic region types (e.g., "navigable area," "land-water boundary," "waterway intersection," "offshore platform") based on the environmental information they provide. To exploit this data, we use Convolutional Neural Networks (CNN) to encode the local chart and rotate it to align with the ship's heading. The environment-based conditional embedding is denoted as C_e.

Encoding the future motion plans of multiple ships as conditional embeddings is crucial to preventing trajectory overlap in multi-ship environments. Specifically, we use a bidirectional LSTM with 256 hidden dimensions to encode future multi-ship motion plans, denoted as C_f, over l time steps. The model takes the historical single-dimensional trajectories of both the ego ship and surrounding ships as input. A bidirectional LSTM is chosen due to its strong performance in simple sequence prediction tasks [3]. The final conditional embeddings are then concatenated into the main representation vector. The final overall model condition C is the fusion of future interaction, environmental embedding, and independent AIS input conditions:

$$C = Fusion(C_e, C_f, C_{ais}). \tag{2}$$

3.3 Denoising Diffusion Process

First we define the diffusion process of adding noise as $(y_0, y_1, \cdots, y_K)$, where y_0 is given initial ground truth trajectory, K is the total number of diffusion steps, and y_K is random noise that follows a Gaussian distribution after K times of noise addition. This process aims to gradually increase the uncertainty of the ship navigation trajectory and destination until the ground truth trajectory becomes a completely random fuzzy motion region. Conversely, we have learned the reverse process of adding noise $(y_K, y_{K-1}, \cdots, y_0)$, to gradually denoise completely random uncertain trajectories into deterministic ones. In the standard denoising module, both the diffusion process and the reverse diffusion process are represented using Markov chains with Gaussian transitions.

The diffusion model is different from the traditional latent variable model in the form of $p_\theta(y_0|C) := \int p_\theta(y_{0:K}|C)dy_{1:K}$. The joint distribution $p_\theta(y_{0:K}|C)$ is called the reverse process. The reverse process is defined as a Markov chain with learned Gaussian transitions starting at $p(y_K) = N(y_k; 0, I)$,We formulate the reverse process as follows:

$$p_\theta(y_{0:K}|C) := p(y_K)\prod_{k}^{K} p_\theta(y_{k-1}|y_k, C), \tag{3}$$

$$p_\theta(y_{k-1}|y_k, C) := N(y_{k-1}; \mu_\theta(y_k, k, C), \textstyle\sum_\theta (y_k, k)), \tag{4}$$

Diffusion model of approximate posterior $q(y_{1:K}|y_0)$, called the forward process or diffusion process. It is fixed in a Markov chain that gradually adds Gaussian noise to the data according to the variance schedule $\beta_1, \beta_2, \cdots, \beta_K$.The posterior distribution of the diffusion process from y_0 to y_K is given by:

$$q(y_{1:K}|y_0) := \prod_{k=1}^{K} q(y_k|y_{k-1}), \tag{5}$$

$$q(y_k|y_{k-1}) := N(y_k; \sqrt{1-\beta_k}y_{k-1}, \beta_k I), \tag{6}$$

Due to the notable property of the Gaussian transitions, we calculate the diffusion process at any step k in a closed form as:

$$q(y_k|y_0) := N(y_k; \sqrt{\bar{\alpha}_k}y_0, (1-\bar{\alpha}_k)I). \tag{7}$$

where $\bar{\alpha}_k = \prod_{s=1}^{k} \alpha_s$ and $\alpha_k = 1 - \beta_k$. Thus, as K becomes sufficiently large, we approximate $y_K \sim N(0, I)$ [16]. This suggests that deterministic trajectories gradually converge to a Gaussian noise distribution as noise is progressively added, aligning with the non-equilibrium thermodynamic phenomena observed in the diffusion process.

3.4 Parameterized Pre-Initializer

Assuming the number of standard denoising steps is K, train a parameterized pre-initializer to replace the initial $K-k$ denoising steps, then jump directly to the kth denoising step through the pre-initializer. Since the pre-initializer fits and replaces the distribution $P(\hat{Y}^k)$ of the first $K-k$ steps, it then follows the standard denoising procedure for k steps to achieve the effect of fine-tuning and refinement. Here, the number of trajectories sampled according to the denoised distribution $P(\hat{Y}^k)$ is S, and the sampled trajectories obtained after the pre-initializer are denoted as: $\hat{y}^k = \{\hat{Y}_1^k, \hat{Y}_2^k, \cdots, \hat{Y}_S^k\}$.

We now illustrate how the denoised distribution $P(\hat{Y}^k)$ of the parameterized pre-initializer is represented. However, it is not easy for a learned model to directly capture complex distributions and often leads to large biases and unstable training. The denoised distribution $P(\hat{Y}^k)$ of the parameterized pre-initializer contains three trainable modules, which are μ_θ, σ_θ, $\hat{S}_\theta$. The three modules are:

$$\begin{aligned} &\mu_\theta = f_\mu(x^0, x^1, \cdots, x^n, C) \\ &\sigma_\theta^2 = f_\sigma(x^0, x^1, \cdots, x^n, C) \\ &\hat{S}_\theta = f_{\hat{S}}(\sigma_\theta, x^0, x^1, \cdots, x^n, C) = (\hat{S}_{\theta,1}, \cdots, \hat{S}_{\theta,s}, \cdots, \hat{S}_{\theta,S}), \end{aligned} \tag{8}$$

Here μ_θ,σ_θ^2 are the mean and variance of the distribution $P(\hat{Y}^k)$, and $\hat{S}_\theta$ is the normalized trajectory position of S samples.

$$\begin{aligned} \hat{Y}_s^k &= \mu_\theta + \sigma_\theta \hat{S}_{\theta,s} \\ \hat{Y}^k &= \mu_\theta + \sigma_\theta \hat{S}_\theta. \end{aligned} \tag{9}$$

Where the mean estimate μ_θ is the average trajectory inferred from past trajectories with respect to the denoising distribution $P(\hat{Y}^k)$ at step k, and the variance estimate σ_θ^2 is the variance inferred from past trajectories with respect to the denoising distribution $P(\hat{Y}^k)$ at step k. These two estimates are the same across S samples. The sample trajectory prediction $\hat{S}_\theta$ is the normalized trajectory position of S samples predicted from the past trajectory and the variance estimate σ_θ^2. It is worth noting that the S sampled trajectories are not independent and identically distributed, the purpose is to ensure the diversity of future trajectories.

3.5 Training Objective

Training our model containing a parameterized pre-initializer is different from training a standard diffusion model. Here, the method of separate training of the parameterized pre-initializer for two stages of pre-initializer and standard diffusion process is used. In the first stage, we only focus on training the standard denoising model. In the second stage, we train a parameterized pre-initializer that is continuously optimized to skip and replace a large number of denoising steps, and then perform a small number of denoising steps to refine the sampled trajectory. This makes the training more stable and speeds up convergence somewhat.

In the first stage of training the standard denoising module, in order to predict the true trajectory, the expected training should optimize the log-likelihood in the reverse process, we maximize the variational lower bound to do the optimization:

$$\begin{aligned} \mathbb{E}[\log p_\theta(y_0)] &\geq \mathbb{E}_q[\log \frac{p_\theta(y_{0:K}, C)}{q(y_{1:K}|y_0)}] \\ &= \mathbb{E}_q[\log p(y_K) + \sum_{k=1}^{K} \log \frac{p_\theta(y_{k-1:k}, C)}{q(y_k|y_{k-1})}], \end{aligned} \tag{10}$$

In the loss function, we ignore the term $\mathbb{E}_q \log p(y_K)$. Since $p(y_K)$ is a standard Gaussian, it has no parameters to learn, so it doesn't make any sense to include it in the loss function.

Here we refer to [16] and omit the calculation of D_{KL}. Finally, we eliminate D_{KL} and obtain the following loss function:

$$L_1 = \mathbb{E}_{\varepsilon, y_0, k}||\varepsilon - \varepsilon_\theta(y_k, k, C)||, \tag{11}$$

Where x is the observed trajectory, $\varepsilon \sim N(0, I)$, $y_k = \sqrt{\bar{\alpha}_k}y_0 + \sqrt{1-\bar{\alpha}_k}\varepsilon$ and the training is performed at each step $k \in 1, 2, \cdots, K$.

In the second stage, we focus on training the parameterized pre-initializer module, and the optimization method uses the frozen denoising module. For each sample, the loss function is given by:

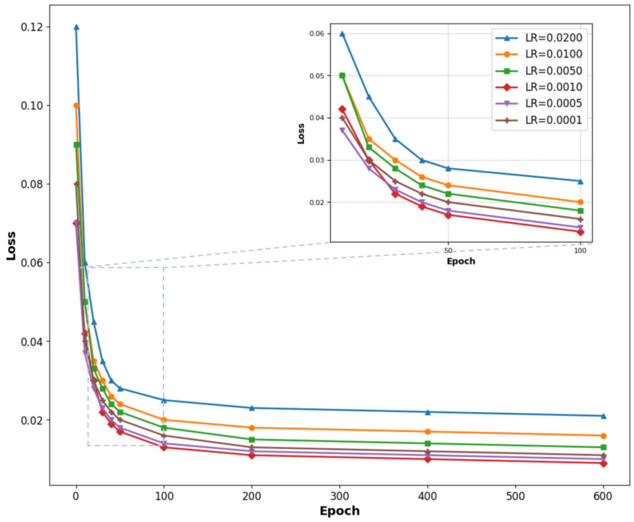

Figure 4: Experiment on the impact of different learning rates on prediction performance.

$$L_2 = \sqrt{m} \cdot \min_s \parallel y_{gt} - \hat{y}^s \|_2 + \sqrt{1-m}\left(\frac{\sum_{s=1}^{S} \parallel y_{gt} - \hat{y}^s \|_2}{\sigma_\theta^2 S} + \log \sigma_\theta^2\right). \tag{12}$$

Where y_{gt} is the ground truth trajectory, $\hat{y}^s$ is the s-th predicted trajectory, and $m \in (0,1)$ is a hyperparameter. $\min_s \parallel y_{gt} - \hat{y}^s \|_2$ means that the generated estimate is reasonable when one of the S generated trajectories is close to the true trajectory.The latter term is the normalization of the variance estimate by the uncertainty loss, which balances the accuracy and diversity of the trajectory prediction. $\frac{\sum_{s=1}^{S} \parallel y_{gt} - \hat{y}^s \|_2}{\sigma_\theta^2 S}$ makes the value of σ_θ^2 proportional to the complexity of the prediction. The second part $\log \sigma_\theta^2$ is a regularization term used to avoid trivial solutions to σ_θ^2.

3.6 Inference

With the parameterized pre-initializer described above, $K-k$ denoising steps of the reverse diffusion process are skipped, followed by standard denoising steps k to 0, as shown in Figure 3. We generate progressively deterministic trajectories, that is, trajectories from y_k to y_0 as follows:

$$y_{k-1} = \frac{1}{\sqrt{\alpha_k}}(y_k - \frac{\beta_k}{\sqrt{1-\bar{\alpha}_k}}\varepsilon_\theta(y_k, k, C)) + \sqrt{\beta_k}\mathbf{z}. \tag{13}$$

Where $\varepsilon_\theta(y_k, k, C)$ is the network trained as described above, whose inputs are the prediction y_k from the previous step, the condition embedding C, and step k, respectively. $\mathbf{z}$ is a random variable with standard Gaussian distribution.

4 EXPERIMENTS

4.1 Datasets

To enhance the model's adaptability, we utilized two real-world vessel trajectory datasets collected from Weihai Port and the Nantong waterway at the Yangtze Estuary. These datasets contain various AIS (Automatic Identification System) attributes, including MMSI, latitude, longitude, speed over ground, course over ground, heading, ship type, ship status, draught, and recording time. The geographical range of the Weihai Port AIS data spans from 121°59'E to 122°40'E longitude and 37°16'N to 37°41'N latitude, while the Nantong waterway data covers 120°21'E to 121°38'E longitude and 31°16'N to 32°5'N latitude.

Due to AIS communication delays, signal interference, and other environmental factors, the raw AIS data contains numerous missing values and outliers. To ensure accurate vessel trajectory extraction, extensive data preprocessing, including completion and cleaning, was performed. Scenes with more than 5% missing values were filtered to minimize noise interference. After this preprocessing step, we obtained 3251 interactive scenarios, each involving at least two vessels. In total, the number of processed timestamp points exceeded 9.5 million, which provided ample data for robust model training and evaluation. The dataset was split into training (2601 scenarios), validation (325 scenarios), and test sets (325 scenarios) for model development and performance assessment.

Additionally, because the AIS data broadcasting interval changes depending on the vessel's operational status, we applied cubic spline interpolation to normalize the AIS data intervals. The time gap between any two adjacent timestamp points was set to 20 seconds to ensure uniformity. Furthermore, using the projection method[20], the water areas were divided into finite grids, with the vessel trajectories' latitude and longitude coordinates approximated by the corresponding grid rows and columns. For consistency across the experiments, the grid size was uniformly set to a side length of 0.0001°.

4.2 Experiment metrics and Implementation Details

To assess the effectiveness of the trajectory prediction model, we adopt two commonly used metrics: the minimum average displacement error (minADE) and the minimum final displacement error (minFDE). These metrics are computed as follows.

The **minADE** is defined as the minimum of the average displacement error over S sampled trajectories between the predicted trajectory $\hat{Y}$ and the ground truth trajectory Y, where each trajectory consists of T time steps:

$$\text{minADE} = \min_{s \in S} \frac{1}{T} \sum_{t=1}^{T} \left\| Y_t - \hat{Y}_t^s \right\|, \tag{14}$$

Here, Y_t is the true position at time step t, and $\hat{Y}_t^s$ is the predicted position at time step t for the s-th sampled trajectory. The **minFDE** focuses on the final displacement error between the true endpoint Y_T and the predicted endpoint $\hat{Y}_T^s$:

$$\text{minFDE} = \min_{s \in S} \left\| Y_T - \hat{Y}_T^s \right\|. \tag{15}$$

Table 1: Comparison with the baseline models on two datasets, minADE/minFDE (meters) are reported. Bold indicates the best results, and lower is better.

Model	Weihai Port			Nantong waterway at the Yangtze Estuary		
	5min	10min	15min	5min	10min	15min
LSTM[23]	15.54/23.49	33.59/45.19	48.42/68.59	13.56/23.49	29.64/39.48	35.36/55.94
Seq2seq[11]	15.92/25.01	32.62/42.57	46.15/67.18	14.87/26.93	28.81/38.16	34.43/45.86
TCNN[2]	15.55/24.14	31.51/37.42	50.48/71.05	13.54/24.22	28.24/37.34	49.86/46.77
Social-STGCNN[9]	8.77/16.17	22.12/29.12	26.45/37.83	7.62/14.17	18.35/29.60	24.52/34.43
Social-LSTM[1]	9.24/21.21	23.51/31.24	31.92/45.29	9.25/20.21	23.19/33.86	27.45/38.65
Social-GAN[14]	8.85/19.37	22.45/29.68	29.47/42.29	8.77/18.37	20.49/31.43	25.71/35.94
Social-VAE[49]	8.93/16.51	20.42/29.25	29.71/45.19	7.92/14.51	19.46/29.15	26.01/38.53
Trajectron++[37]	6.48/14.81	16.84/26.59	28.89/41.42	6.47/**9.76**	14.42/25.26	26.49/40.77
MID[13]	6.22/**12.61**	17.21/25.51	24.16/36.43	6.23/11.61	16.86/27.59	20.42/32.79
LED[30]	6.15/12.98	16.82/23.84	22.58/34.16	**6.02**/10.86	15.81/25.67	22.58/34.19
Ours (MFPD)	**5.98**/12.91	**15.68**/**22.87**	**22.34**/**34.08**	6.15/10.85	**13.27**/**23.73**	**20.24**/**31.92**

Both minADE and minFDE are computed over S sampled trajectories, and lower values indicate better prediction accuracy. The prediction is evaluated for time intervals of 5, 10 and 15 minutes, with samples taken every 20 seconds.

4.3 Network Architecture and Parameters Setting

The model proposed in this paper, as well as the baseline models used for comparison, were trained and fine-tuned using the training and validation sets. We constructed a Transformer-based social encoder with a dimension of 256, 3 heads, and 3 encoder layers. To determine the optimal number of layers and residual blocks, we employed cross-validation to test prediction accuracy, ultimately confirming the use of these hyperparameters as the best configuration.

For encoding temporal information, we utilized a temporal encoder with a 1D convolution kernel of size 4 and 32 output channels. Additionally, an LSTM with a hidden size of 256 was used to capture temporal dependencies in the trajectory data. Following this, the decoder module employs a Transformer with a hidden size of 256, which is responsible for extracting contextual information and constructing the core denoising module necessary for prediction. Moreover, after the social encoding, we employed a bidirectional LSTM with 256 hidden dimensions to encode the future states of neighboring vessels, which were pre-generated using the same dimension RNN. Simultaneously, we encoded local Electronic Nautical Charts (ENC) using a four-layer Convolutional Neural Network (CNN), aligning them with the vessel's heading direction by applying rotation, thus ensuring spatial consistency in the model's perception of the environment.

The learning rate (LR), a crucial hyperparameter in deep neural networks, significantly influences both the prediction accuracy and the robustness of the model. As LR changes, the model's performance can vary dramatically. Selecting an optimal LR remains a challenging task. As shown in Figure 4, we conducted extensive experiments to explore the impact of different LR values. Based on these experiments, we selected 0.001 as it allowed for stable and rapid convergence of the training process. The activation functions in the model employ ReLU to capture non-linearities in the learned features. The entire framework was trained on an NVIDIA RTX 4080 GPU using the Adam optimizer, which was set with default parameters for efficient gradient-based optimization.

Table 2: Ablation experiments for 10-minutes future trajectory prediction on the Weihai Port dataset.

Method	Architecture	Steps	Sampling	minADE	minFDE	Inference time(s)
MFPD (w/o pre-initializer)	Transformer	200	20	16.51	23.59	~3.11
MFPD (w/o pre-initializer)	Transformer	20	20	21.34	28.61	~0.29
MFPD (w/o future and ENC embedding)	Transformer	10	20	17.53	26.41	~0.15
MFPD (replace MLP decoder)	MLP	10	20	16.31	25.21	~0.13
MFPD (replace GRU decoder)	GRU	10	20	16.14	24.94	~0.15
MID[13]	Transformer	200	20	17.21	25.51	~3.02
MFPD	Transformer	10	20	15.68	22.87	~0.15
MFPD (with more denoising steps)	Transformer	20	20	15.59	22.61	~0.31

4.4 Comparison with other methods

We compare the proposed method, MFPD, with state-of-the-art baseline methods in the field of vessel trajectory prediction using two real-world datasets from Weihai Port and Nantong waterway at the Yangtze Estuary. The comparison results are presented in Table 1. To evaluate the performance of different trajectory prediction methods, we adopt two two widely used metrics, minADE (Minimum Average Displacement Error) and minFDE (Minimum Final Displacement Error), instead of the commonly used ADE and FDE, as they better capture the performance of multimodal trajectory predictions by focusing on the most accurate predicted trajectory.

Traditional models such as LSTM, Seq2Seq, and TCNN fail to generate satisfactory prediction results. This is primarily because these methods do not consider the latent interactions between surrounding vessels. While models like Transformer, Social-LSTM, and Social-STGCNN do incorporate social interactions, they lack sufficient representation of trajectory uncertainty and diversity, resulting in suboptimal performance in complex maritime environments. Compared to other generative models, diffusion models exhibit stronger representational power. Both MID and the proposed MFPD method show improved performance in terms of representation and diversity. However, MID suffers from longer inference times, which limits its utility for real-time maritime prediction tasks. In contrast, MFPD addresses this issue by incorporating a pre-initializer, reducing inference time by approximately 20 times without compromising prediction accuracy (as confirmed by comparisons with MFPD without the pre-initializer).

Additionally, MFPD achieves superior performance in real-world vessel trajectory prediction tasks because it encodes future interaction trajectories and the maritime environment, rather than solely relying on historical trajectory state encoding. The proposed method achieves an average minADE/minFDE of 13.94/22.73 on the two datasets from Weihai Port and the Nantong waterway at the Yangtze Estuary, outperforming all baseline methods. Moreover, we observe that MFPD demonstrates an advantage on larger vessel trajectory datasets and longer prediction horizons. This indicates that the proposed MFPD method is better suited for handling larger-scale datasets and extended trajectory predictions, making it a promising approach for future complex prediction tasks.

4.5 Ablation Studies

We conducted ablation studies to evaluate the contribution of key components in the proposed method, including the parameterized

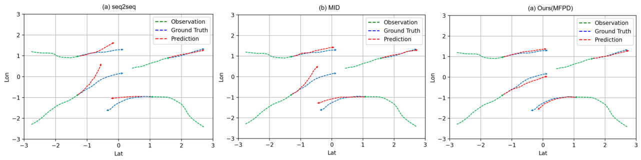

Figure 5: Visualization of the comparison with the baselines at 5 and 10 minute prediction times

Figure 6: Visualization of 5-minute and 10-minute trajectory predictions for a scenario in the Nantong waterway at the Yangtze Estuary dataset.

pre-initializer, Transformer-based architecture, and the encoding of future interaction and environmental information. Additionally, we examined the effect of different denoising steps on predictive performance and inference time. The comparison results and visualization are shown in Table 2.

First, we evaluated the impact of removing the parameterized pre-initializer by comparing the performance of the full MFPD model and MFPD without pre-initializer, which uses the standard diffusion model instead of our parameterized pre-initializer. The results indicate that the parameterized pre-initializer significantly improves prediction accuracy and reduces inference time by nearly 20-fold. Next, we analyzed the importance of future interaction and environmental information encoding by comparing MFPD with and without these components. The results show that removing this encoding leads to a notable decrease in performance, highlighting its crucial role in modeling complex maritime environment.

We also tested the effect of different decoder architectures by replacing the Transformer in MFPD with MLP and GRU, respectively. The comparison demonstrates that the Transformer-based architecture achieves the best prediction performance. Lastly, we assessed the influence of different denoising steps. Increasing the number of denoising steps improves predictive accuracy but also increases inference time and reduces the diversity of the generated trajectory samples. This suggests that tuning the number of denoising steps is essential to balance accuracy and diversity.

4.6 Qualitative Evaluation

The visual comparison of the predicted trajectories is shown in Fig. 5. These figures depict the future trajectory predictions for 5-minute and 10-minute intervals on the Weihai port dataset, comparing the predicted trajectories of the baselines seq2seq and MID with our MFPD model. It can be observed that the predictions made by MFPD are more accurate, and the diversity of the trajectory samples obtained through sampling enables the model to handle complex and dynamic maritime conditions effectively. The visual comparison of different navigation scenarios shown in Figure 5 shows that it is apparent that in simple scenarios with minimal ship interaction and straight-line movement, seq2seq, MID, and MFPD can all accurately and robustly predict ship trajectories. However, in more complex scenarios involving multiple interacting ships, turning maneuvers, or potential overlap of future trajectories, MFPD demonstrates superior performance. This is because future changes in a ship's position and orientation tend to be nonlinear, especially in scenarios where ships need to maintain a minimum safe distance from one another, which is a key consideration in practice.

Furthermore, the predictive performance is closely tied to the prediction horizon (i.e., the future time range). In theory, it is feasible to generate satisfactory predictions over short time horizons. However, as the prediction horizon increases, accurately forecasting ship trajectories becomes more challenging. Fig. 6 illustrates the visualization of 5-minute and 10-minute trajectory predictions on the Weihai Port dataset. The results show that the MFPD model not only achieves excellent prediction accuracy for short-term forecasts but also maintains strong performance over longer time horizons. In particular, it excels in predicting trajectory endpoints, and course over ground. This success is attributed to the inclusion of multimodal data during input encoding, rather than relying solely on latitude and longitude, which enhances the model's representational capability. These findings highlight the robustness of MFPD in complex maritime environments, particularly under conditions of non-linear ship interactions and long-range trajectory prediction.

5 CONCLUSION

In this paper, we presented a novel Multi-Ship Future interaction trajectory prediction model (MFPD) to address the challenges of real-time multi-ship trajectory modeling in complex maritime environments. Our approach integrates a parameterized pre-initializer with a diffusion-based framework, significantly reducing the denoising steps required for trajectory prediction while maintaining high accuracy and diversity. By incorporating both historical trajectory data, future interactions, and multimodal maritime environmental information as conditional embeddings, our model is able to better capture dynamic vessel behaviors and potential future interactions. Experimental results on two real datasets confirm the superior performance of MFPD, which not only achieves state-of-the-art performance in the task of ship trajectory prediction, but also has an inference speed 20 times faster than general diffusion models. These advantages make the MFPD model a valuable tool for real-time maritime safety applications, offering both efficiency and robustness in uncertain, dynamic environments. Future work could further explore the integration of additional multimodal data sources and extend the model's applicability to other real-time maritime decision-making tasks.

ACKNOWLEDGMENTS

This research was supported by the Key R&D Program of Shandong Province (2022CXGC020403) and Major Strategic Product Innovation (Class A) Project (BA2023008).

REFERENCES

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. 2016. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 961–971.

[2] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271 [cs.LG] https://arxiv.org/abs/1803.01271

[3] Denny Britz, Anna Goldie, Minh-Thang Luong, and Quoc Le. 2017. Massive Exploration of Neural Machine Translation Architectures. arXiv:1703.03906 [cs.CL] https://arxiv.org/abs/1703.03906

[4] Samuele Capobianco, Leonardo M. Millefiori, Nicola Forti, Paolo Braca, and Peter Willett. 2021. Deep Learning Methods for Vessel Trajectory Prediction Based on Recurrent Neural Networks. *IEEE Trans. Aerospace Electron. Systems* 57, 6 (2021), 4329–4346. https://doi.org/10.1109/TAES.2021.3096873

[5] Guangyi Chen, Junlong Li, Jiwen Lu, and Jie Zhou. 2021. Human trajectory prediction via counterfactual analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9824–9833.

[6] Guangyi Chen, Junlong Li, Nuoxing Zhou, Liangliang Ren, and Jiwen Lu. 2021. Personalized trajectory prediction via distribution discrimination. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 15580–15589.

[7] Kehua Chen, Xianda Chen, Zihan Yu, Meixin Zhu, and Hai Yang. 2023. EquiDiff: A Conditional Equivariant Diffusion Model For Trajectory Prediction. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. 746–751. https://doi.org/10.1109/ITSC57777.2023.10421892

[8] Kehua Chen, Xianda Chen, Zihan Yu, Meixin Zhu, and Hai Yang. 2023. EquiDiff: A Conditional Equivariant Diffusion Model For Trajectory Prediction. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. 746–751. https://doi.org/10.1109/ITSC57777.2023.10421892

[9] Hui Feng, Guilin Cao, Haixiang Xu, and Shuzhi Sam Ge. 2022. IS-STGCNN: An Improved Social spatial-temporal graph convolutional neural network for ship trajectory prediction. *Ocean Engineering* 266 (2022), 112960.

[10] Xidong Feng, Zhepeng Cen, Jianming Hu, and Yi Zhang. 2019. Vehicle trajectory prediction using intention-based conditional variational autoencoder. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 3514–3519.

[11] Nicola Forti, Leonardo M. Millefiori, Paolo Braca, and Peter Willett. 2020. Prediction oof Vessel Trajectories From AIS Data Via Sequence-To-Sequence Recurrent Neural Networks. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 8936–8940. https://doi.org/10.1109/ICASSP40776.2020.9054421

[12] Klaus Greff, Rupesh K. Srivastava, Jan Koutnik, Bas R. Steunebrink, and Jurgen Schmidhuber. 2017. LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems* 28, 10 (Oct. 2017), 2222–2232. https://doi.org/10.1109/tnnls.2016.2582924

[13] Tianpei Gu, Guangyi Chen, Junlong Li, Chunze Lin, Yongming Rao, Jie Zhou, and Jiwen Lu. 2022. Stochastic trajectory prediction via motion indeterminacy diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17113–17122.

[14] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. 2018. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2255–2264.

[15] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. 2018. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2255–2264.

[16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.

[17] Chengfeng Jia and Jie Ma. 2023. Conditional temporal GAN for intent-aware vessel trajectory prediction in the precautionary area. *Engineering Applications of Artificial Intelligence* 126 (2023), 106776.

[18] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. 2022. Elucidating the Design Space of Diffusion-Based Generative Models. arXiv:2206.00364 [cs.CV] https://arxiv.org/abs/2206.00364

[19] Mihee Lee, Samuel S Sohn, Seonghyeon Moon, Sejong Yoon, Mubbasir Kapadia, and Vladimir Pavlovic. 2022. Muse-VAE: multi-scale VAE for environment-aware long term trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2221–2230.

[20] Maohan Liang, Ryan Wen Liu, Shichen Li, Zhe Xiao, Xin Liu, and Feng Lu. 2021. An unsupervised learning method with convolutional auto-encoder for vessel trajectory similarity computation. *Ocean Engineering* 225 (2021), 108803. https://doi.org/10.1016/j.oceaneng.2021.108803

[21] Maohan Liang, Ryan Wen Liu, Yang Zhan, Huanhuan Li, Fenghua Zhu, and Fei-Yue Wang. 2022. Fine-grained vessel traffic flow prediction with a spatio-temporal multigraph convolutional network. *IEEE Transactions on Intelligent Transportation Systems* 23, 12 (2022), 23694–23707.

[22] Ryan Wen Liu, Maohan Liang, Jiangtian Nie, Wei Yang Bryan Lim, Yang Zhang, and Mohsen Guizani. 2022. Deep Learning-Powered Vessel Trajectory Prediction for Improving Smart Traffic Services in Maritime Internet of Things. *IEEE Transactions on Network Science and Engineering* 9, 5 (2022), 3080–3094. https://doi.org/10.1109/TNSE.2022.3140529

[23] Ryan Wen Liu, Maohan Liang, Jiangtian Nie, Wei Yang Bryan Lim, Yang Zhang, and Mohsen Guizani. 2022. Deep Learning-Powered Vessel Trajectory Prediction for Improving Smart Traffic Services in Maritime Internet of Things. *IEEE Transactions on Network Science and Engineering* 9, 5 (2022), 3080–3094. https://doi.org/10.1109/TNSE.2022.3140529

[24] Shaohua Liu, Haibo Liu, Huikun Bi, and Tianlu Mao. 2020. Col-gan: Plausible and collision-less trajectory prediction by attention-based gan. *IEEE Access* 8 (2020), 101662–101671.

[25] Yuejiang Liu, Qi Yan, and Alexandre Alahi. 2021. Social nce: Contrastive learning of socially-aware motion representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 15118–15129.

[26] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. 2023. DPM-Solver++: Fast Solver for Guided Sampling of Diffusion Probabilistic Models. arXiv:2211.01095 [cs.LG] https://arxiv.org/abs/2211.01095

[27] Jie Ma, Chengfeng Jia, Yaqing Shu, Kezhong Liu, Yu Zhang, and Yulong Hu. 2021. Intent prediction of vessels in intersection waterway based on learning vessel motion patterns with early observations. *Ocean Engineering* 232 (2021), 109154.

[28] Kun Ma, Qilong Han, Jingzheng Yao, and Yuntao Zhang. 2024. DM-BDD: Real-time Vessel Trajectory Prediction Based on Diffusion Probability Model Balancing Diversity and Determinacy. In *2024 International Joint Conference on Neural Networks (IJCNN)*. 1–8. https://doi.org/10.1109/IJCNN60899.2024.10649942

[29] Osama Makansi, Julius von Kügelgen, Francesco Locatello, Peter Gehler, Dominik Janzing, Thomas Brox, and Bernhard Schölkopf. 2021. You Mostly Walk Alone: Analyzing Feature Attribution in Trajectory Prediction. arXiv:2110.05304 [cs.LG] https://arxiv.org/abs/2110.05304

[30] Weibo Mao, Chenxin Xu, Qi Zhu, Siheng Chen, and Yanfeng Wang. 2023. Leapfrog Diffusion Model for Stochastic Trajectory Prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5517–5526.

[31] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. 2020. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 14424–14432.

[32] Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. 2021. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International Conference on Machine Learning*. PMLR, 8857–8868.

[33] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10684–10695.

[34] H Rong, AP Teixeira, and C Guedes Soares. 2019. Ship trajectory uncertainty prediction based on a Gaussian Process model. *Ocean Engineering* 182 (2019), 499–511.

[35] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. 2019. Sophie: An attentive gan for predicting paths compliant to social and physical constraints. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1349–1358.

[36] Tim Salimans and Jonathan Ho. 2022. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512* (2022).

[37] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. 2020. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*. Springer, 683–700.

[38] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. 2015. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. arXiv:1503.03585 [cs.LG] https://arxiv.org/abs/1503.03585

[39] Jiaming Song, Chenlin Meng, and Stefano Ermon. 2020. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502* (2020).

[40] Huang Tang, Yong Yin, and Helong Shen. 2022. A model for vessel trajectory prediction based on long short-term memory neural network. *Journal of Marine Engineering & Technology* 21, 3 (2022), 136–145.

[41] Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. 2021. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. *Advances in Neural Information Processing Systems* 34 (2021), 24804–24816.

[42] Enmei Tu, Guanghao Zhang, Lily Rachmawati, Eshan Rajabally, and Guang-Bin Huang. 2018. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology. *IEEE Transactions on Intelligent Transportation Systems* 19, 5 (2018), 1559–1582. https://doi.org/10.1109/TITS.2017.2724551

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention Is All You Need. arXiv:1706.03762 [cs.CL] https://arxiv.org/abs/1706.03762

[44] Zihao Wang, Haitong Xu, Li Xia, Zaojian Zou, and C Guedes Soares. 2020. Kernel-based support vector regression for nonparametric modeling of ship maneuvering

motion. *Ocean Engineering* 216 (2020), 107994.

[45] Chenxin Xu, Maosen Li, Zhenyang Ni, Ya Zhang, and Siheng Chen. 2022. Groupnet: Multiscale hypergraph neural networks for trajectory prediction with relational reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6498–6507.

[46] Chenxin Xu, Robby T. Tan, Yuhong Tan, Siheng Chen, Yu Guang Wang, Xinchao Wang, and Yanfeng Wang. 2023. EqMotion: Equivariant Multi-agent Motion Prediction with Invariant Interaction Reasoning. arXiv:2303.10876 [cs.CV] https://arxiv.org/abs/2303.10876

[47] Pei Xu, Jean-Bernard Hayet, and Ioannis Karamouzas. 2022. Socialvae: Human trajectory prediction using timewise latents. In *European Conference on Computer Vision*. Springer, 511–528.

[48] Dong Yang, Lingxiao Wu, and Shuaian Wang. 2021. Can we trust the AIS destination port information for bulk ships?–Implications for shipping policy and practice. *Transportation Research Part E: Logistics and Transportation Review* 149 (2021), 102308. https://doi.org/10.1016/j.tre.2021.102308

[49] Ye Yuan, Xinshuo Weng, Yanglan Ou, and Kris M. Kitani. 2021. AgentFormer: Agent-Aware Transformers for Socio-Temporal Multi-Agent Forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 9813–9823.

[50] Hai Zhou, Yaojie Chen, and Sumin Zhang. 2019. Ship trajectory prediction based on BP neural network. *Journal of Artificial Intelligence* 1, 1 (2019), 29.