

# Interaction-Aware Short-Term Marine Vessel Trajectory Prediction With Deep Generative Models

Peihua Han , *Member, IEEE*, Mingda Zhu , *Member, IEEE*,  
and Houxiang Zhang , *Senior Member, IEEE*

**Abstract**—Navigation safety is of paramount importance in areas with heavy and complex maritime traffic. Any ship navigating such a scenario should be able to foresee the future positions of other ships and adjust its path accordingly to avoid collisions. However, predicting future trajectories is a very challenging problem due to many possible future trajectories from the inherent uncertainty and the complex interaction dynamics between different ships. In this article, we propose a deep generative model based on the conditional variational autoencoder framework to learn marine vessel movement and predict future trajectories. The model is able to produce a multimodal probability distribution over future trajectories and model the complex interactions between vessels. Experiments are performed in two-vessel encounter scenarios from real-world automatic identification system data. The proposed model outperforms the baseline methods, including both kinematics-based and data-driven methods. The trajectories predicted by the proposed model are also analyzed to demonstrate the effectiveness of the model.

**Index Terms**—Automatic identification system (AIS) data, generative model, marine vessel, neural network, trajectory prediction.

## I. INTRODUCTION

OVER the past few decades, intelligent marine transportation systems have received increasing attention from the maritime industry. The development of digital twin [1], remotely operated, and autonomous ships is within such a context. They are expected to increase the efficiency of maritime transport while reducing fuel consumption and extending the operating window [2]. Understanding vessel motion is a key skill of intelligent systems, which includes predicting the future trajectories of other traffic vessels. This prediction enables a range of downstream tasks, such as predictive planning, model predictive control, and collision avoidance.

Manuscript received 30 January 2023; revised 16 June 2023; accepted 31 July 2023. Paper no. TII-23-0313. (*Corresponding author: Peihua Han.*)

The authors are with the Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, 6009 Aalesund, Norway (e-mail: peihua.han@ntnu.no; mingda.zhu@ntnu.no; hozh@ntnu.no).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TII.2023.3302304>.

Digital Object Identifier 10.1109/TII.2023.3302304

The challenge of the accurate trajectory prediction for marine vessels arises from the complexity of human behavior and its diversity of internal and external stimuli [3]. The future behaviors of marine vessels may be driven by their intent, the interaction with surrounding vessels, the environment, and traffic rules. Most factors cannot be directly observed and need to be inferred from noisy perceptual cues. Traditionally, the constant velocity model is used for vessel trajectory prediction, and the future position is simply extrapolated by its velocity and course. More advanced model-based approaches involve a Bayesian filter to estimate the acceleration or turning rate [4] and then assume these parameters remain constant. These methods apparently have difficulty in modeling the intent of the vessels as well as other stimuli and, thus, often lead to high prediction errors in the real world.

Pattern-based methods, especially machine-learning methods, are able to address the aforementioned complexities involved in trajectory prediction by learning from historical data. These methods learn motion behavior by fitting different function approximators to the data, ranging from hidden Markov models to more recently deep neural networks. Learning such models requires recording historical traffic data. For marine vessels, the automatic identification system (AIS) is used, which is an automatic tracking system that uses transceivers on the vessel and is used by vessel traffic services. Information provided by AIS includes unique identification, position, course, speed, etc. Its main purpose is to allow ships to view marine traffic in their area and to be seen by that traffic. The historical AIS data are also saved and it is usually publicly accessible from different organizations, such as coastal administration. Thus, these AIS data form a rich dataset for analyzing the behavior or traffic of marine vessels. It can also be used for learning trajectory prediction models, and different machine-learning models have been applied [5], [6] by using AIS data.

However, unlike human motion prediction [3], these models for marine vessels focus on long-term predictions (up to several hours) and, thus, do not take into account the inherent uncertainties of the predictions and interactions between ships. This may be due to the fact that long-term trajectory prediction mainly depends on the ship's destination and route. It is pretty deterministic when the destination and the route are known. However, this is not the case for the short-term prediction (up

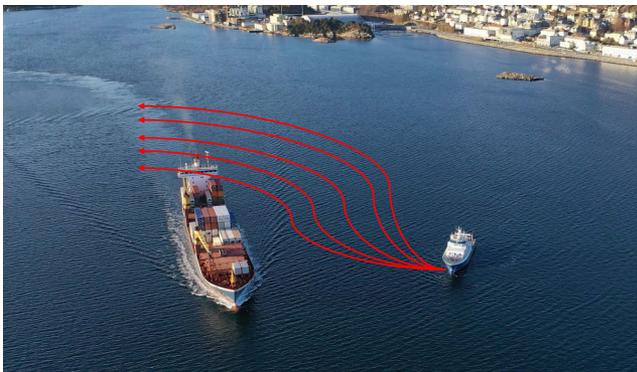


Fig. 1. Illustration of the possible trajectories in an encounter scenario. The red lines denote the possible trajectories, which depend heavily on the captain and the interactive ship.

to several minutes) as it is heavily influenced by these two factors. For example, in the case of an encountering ship, as shown in Fig. 1, whether the ship gives way or passes directly depends on the behavior of the encountering ship. Also, an aggressive captain might only alternate the course slightly in a give-way situation, while a conservative captain would alternate the course in a much larger manner, which results in the inherent uncertainties (aleatoric uncertainty) of the predictions. Therefore, it is important to consider ship interactions and prediction uncertainty for short-term ship trajectory prediction. Note that the uncertainty can be divided into aleatoric and epistemic uncertainty. Uncertainty in the predicted trajectory comes from unknown stimuli, such as the captain's intentions, and cannot be removed by collecting more data. Therefore, it is regarded as aleatoric uncertainty. To model this uncertainty, predicted trajectories are treated as distributions and modeled by a variational autoencoder.

In this article, we propose a novel model based on deep neural networks for short-term trajectory prediction of marine vessels. In particular, we approximate the predicted aleatoric uncertainty with a deep generative model: Conditional variational autoencoder (CVAE). It includes a latent random vector to represent the uncertainty. Multiple future trajectories can be generated by sampling from this latent vector. The model is implemented in a sequence-to-sequence (seq2seq) manner with the use of the recurrent neural network (RNN) to better handle the sequential data. The interaction between the vessels is encoded as context for the CVAE model. The performance of the model is demonstrated with two-vessel encounter scenarios from AIS data. Although CVAE has been applied to the trajectory prediction of people and vehicles, it is the first time that it is applied to marine vessels. The main contributions of this article can be highlighted as follows.

- 1) A novel model is developed for short-term trajectory prediction of marine vessels, which includes the prediction uncertainty and the interaction between vessels.
- 2) Extensive experiments are performed to validate the model and the detailed analyses of the future trajectory patterns generated by the model are conducted.

CVAE is a deep generative model based on autoencoders. Other generative models, such as regression generative

networks [7], lead to min-max optimization problems that are known to be unstable to train. The main focus of vanilla autoencoders is usually to compress data for downstream tasks, as shown in [8]. And for CVAE, although it uses an autoencoder architecture, its focus is on model distribution. The rest of this article is organized as follows. Section II presents the literature reviews of trajectory prediction. The illustration of the proposed prediction model is given in Section III. In Section IV, experiments are conducted with AIS data to validate the model, and the results are shown and discussed. Finally, Section V concludes this article.

## II. RELATED WORKS

### A. Human and Vehicle Trajectory Prediction

Trajectory prediction for humans and vehicles has been extensively studied in recent years, especially in the application domains of autonomous vehicles and service robots. Learning-based methods are one of the modeling approaches that have made promising progress recently. In particular, RNNs for sequence learning have become a widely popular modeling approach in such a context [9], [10], [11]. Altche and de la Fortelle [10] use a long short-term memory (LSTM) network for highway trajectory prediction. Similarly, Park et al. [11] use an LSTM network as well but use an encoder-decoder structure to generate the future trajectory sequence in highway traffic scenarios. This kind of model is also called the seq2seq model. However, these methods only produce a single deterministic trajectory output, thus neglecting to capture the uncertainty inherent in the prediction process. Predicting multiple future trajectories or the distribution of possible future outcomes is critical for safety-critical systems, as it requires reasoning over many possible future outcomes to guard against worst-case scenarios. In [9], the future position is assumed to follow a Gaussian distribution to represent the distribution of possible future outcomes. This method is simple but not able to account for the multimodal futures' distributions. In such a context, a popular approach is to use deep generative models that model the future trajectories implicitly as latent variables. Gupta et al. [12] leverage generative adversarial networks to capture future distribution. The model consists of a generator and a discriminator network, and the generator outputs trajectory samples, which are then evaluated by the discriminator. Rhinehart et al. [13] use a flow-based generative model, while Ivanovic et al. [14] use CVAE framework. These generative models show promising results in generating multimodal distributions of future trajectories. The interaction between different agents also has a nonnegligible effect on future trajectories. Many approaches attempt to aggregate information from neighboring agents. Alahi et al. [9] model the interaction of pedestrians by sharing the hidden state of each individual RNN using a pooling mechanism. Salzmann et al. [15] represent the scene as a graph and aggregate the information from neighboring agents by an elementwise sum. In addition to the above two factors, more stimuli for the trajectory prediction, such as the target destination [16], can also be included.

## B. Trajectory Prediction for Marine Vessels

In the maritime domain, the term trajectory prediction is used not only for the traffic vessel but also for the controlled vessel as in [17]. This is due to the inaccurate dynamic model and uncertain environmental effects [18] on the controlled vessel. The major difference is whether the control command or future plan is available. In this article, we focus on the traffic vessel since neither the control commands nor the future plan is accessible. Unlike human and vehicle trajectory prediction, the prediction horizon of marine vessels is usually much longer (on the order of several minutes). Also, lanes are not designated as vehicles. The learning-based approaches have received increasing attention in recent years. Gao et al. [19] apply a similarity-based method to determine the destination point of the vessel from historical data and use an LSTM network to generate multiple support points to the destination point. The predicted trajectory is the cubic interpolation of the support and destination points. Capobianco et al. [6] developed a seq2seq model, where the encoder is a bidirectional LSTM network and the decoder is a unidirectional LSTM network. The attention mechanism between the encoder and decoder is utilized. Murray and Perera [5] developed a two-step approach. The historical AIS trajectories are first clustered using a clustering method and a local prediction model is built for each cluster. The local prediction model is similar to [6], which is a seq2seq LSTM network with an attention mechanism. Nguyen and Fablet [20] learn a prediction model based on the transformer architecture. But instead of learning a regression model, they discretize position into bins and learn a classification model. Multistep prediction is made by applying this model recursively. The majority of the research for ship trajectory prediction focuses on finding a suitable network architecture, either for long-term or short-term prediction. Liu et al. [21] apply a graph convolutional neural network to aggregate the information from surrounding vessels. The future position of the vessel is assumed to follow a Gaussian distribution. As a result, all of them except Liu et al. [21] use a deterministic approach that does not consider the inherent uncertainty of the prediction. However, the multimodal distributions of future trajectories are not considered in [21]. Besides, most of them do not try to capture the interaction between different vessels, which could be nonnegligible in short-term prediction. In this article, we propose a model that includes the prediction uncertainty as well as vessel interaction and emphasize their importance.

## III. GENERATIVE MODEL FOR VESSEL TRAJECTORY PREDICTION

In this section, a general CVAE model and the gated recurrent unit (GRU) are described and we apply them in the context of vessel trajectory prediction. Then, the core characteristics of the proposed CVAE trajectory prediction model are illustrated.

### A. Conditional Variational Autoencoder

Given a dataset  $D = \{(x_i, y_i)\}_{i=1}^N$ , the conditional generative modeling tries to fit a model of the conditional probability distribution  $p(y|x)$ . Once fit, the model can be used to generate

samples  $y$  given  $x$ , which can be used to represent uncertainty for trajectory prediction. In other words, the aleatoric uncertainty is modeled as a conditional probability distribution  $p(y|x)$ . In this article, we consider  $p(y|x)$  to be defined by a fixed set of parameters, which we fit into the dataset with the objective of maximizing the likelihood of the observed data. In such a context, neural networks are often used due to their strong expressivity. The commonly used model includes CVAE [22] and conditional generative adversarial network (CGAN) [23]. We choose to use CVAE because CGAN is harder to train and may suffer from mode collapse.

A CVAE is a latent conditional generative model. The CVAE consists of an encoder  $q_\phi(z|y, x)$  parameterized by  $\phi$  and a decoder  $p_\theta(y|z, x)$  parameterized by  $\theta$ . The encoder takes the inputs  $y$  and  $x$  and produces a distribution  $p(z|x)$ , where  $z$  is the latent vector. The decoder uses  $x$  and samples from  $p_\theta(z|x)$  to produce  $y$ . The model can be described by

$$p(y|x) = \int p_\theta(y|z, x)p(z|x)dz. \quad (1)$$

To efficiently perform the marginalization in (1), a proposal distribution  $q_\phi(z|y, x)$  is used. The marginal likelihood in (1) becomes

$$\begin{aligned} p(y|x) &= \int \frac{p_\theta(y|z, x)p(z|x)}{q_\phi(z|y, x)}q_\phi(z|y, x)dz \\ &= \mathbb{E}_{q_\phi(z|y, x)} \left[ \frac{p_\theta(y|z, x)p(z|x)}{q_\phi(z|y, x)} \right]. \end{aligned} \quad (2)$$

By taking the log of both sides in (2) and using Jensen's inequality, the evidence lower bound (ELBO) is derived as follows:

$$\begin{aligned} \log p(y|x) &\geq \mathbb{E}_{q_\phi(z|y, x)} [p_\theta(y|z, x)] \\ &\quad - D_{\text{KL}} [q_\phi(z|y, x)||p(z|x)] \end{aligned} \quad (3)$$

where  $D_{\text{KL}}$  is the Kullback–Leibler (KL) divergence. Therefore, instead of maximizing the log-likelihood directly, the ELBO is maximized. By using the reparameterization trick [24], the ELBO is tractable and can be optimized via stochastic gradient descent. The negative ELBO is, therefore, minimized and the loss for a single training example  $(x, y)$  is

$$\begin{aligned} \mathcal{L}(x, y) &= -\mathbb{E}_{q_\phi(z|y, x)} [p_\theta(y|z, x)] \\ &\quad + D_{\text{KL}} [q_\phi(z|y, x)||p(z|x)]. \end{aligned} \quad (4)$$

During training, the negative log likelihood (first term) is modeled as the mean square error. The distribution  $p(z|x)$  is modeled as a standard multivariate Gaussian distribution  $p(z|x) \sim \mathcal{N}(0, I)$ . The loss is minimized to find the neural network parameters  $\phi$  and  $\theta$ .

### B. Gated Recurrent Unit

The GRU [25] is an RNN with a gating mechanism to avoid gradient vanishing problems in RNN. It is similar to LSTM but has fewer parameters. Given sequential data  $x_{1:T}$ , the GRU processes the sequence by repeating the following function:

$$r_t = \sigma(W_{ir}x_t + W_{hr}h_{t-1})$$

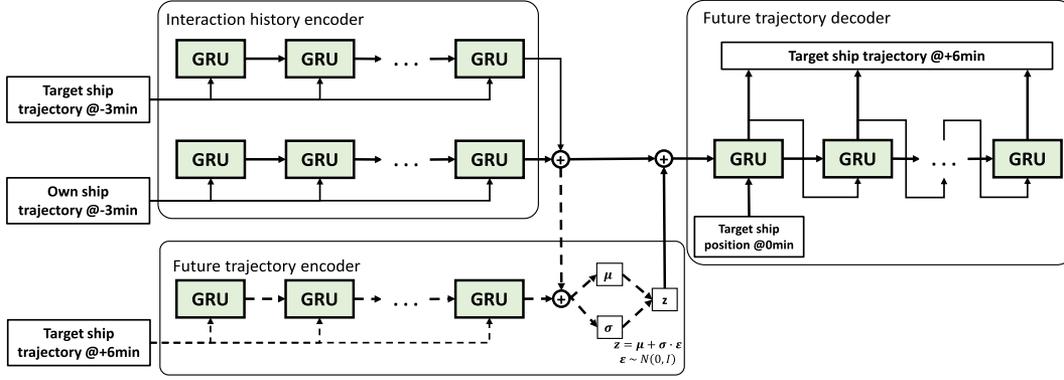


Fig. 2. Schematic illustration of the neural network architecture of a CVAE for vessel trajectory prediction. The solid lines denote all the processes for the inference, while the dashed lines represent the processes only used in training.

$$\begin{aligned}
 z_t &= \sigma(W_{iz}x_t + W_{hz}h_{t-1}) \\
 n_t &= \tanh(W_{in}x_t + r_i * (W_{hn}h_{t-z})) \\
 h_t &= (1 - z_t) * n_t + z_t * h_{t-1}
 \end{aligned} \quad (5)$$

where  $x_t$  and  $h_t$  are the input and hidden state at time  $t$ , respectively.  $r_t$ ,  $z_t$ , and  $n_t$  are the reset, update, and new gates, respectively.  $W$  is the weight of the network. Note that LSTMs can be used instead of GRUs in this article. GRU was chosen because it can consider long-term dependencies while having fewer parameters than LSTM.

### C. Interaction-Aware Trajectory Prediction

In order to model the complex trajectory prediction process of marine vessels, we are interested in learning a model for future trajectory prediction that satisfies the following desires.

- 1) The model is history dependent that the intent or future trajectory can be predicted from the past trajectory.
- 2) The interaction between the two encountering vessels is captured.
- 3) The model is able to generate multiple future trajectories to account for uncertainty in the forecasting process.

The proposed seq2seq CVAE is illustrated in Fig. 2. Here, we denote the ship that we want to predict its trajectory as the target ship, while the other one is called the own ship. Three modules are included in the model: An interaction history encoder, a future trajectory encoder, and a future trajectory decoder. All these three modules are parameterized with GRU.

1) *Interaction History Encoder*: The interaction history encoder is designed to encode the past information from the target ship and own ship into vectors. In particular, the trajectories of the target ship and own ship over the past 3 min are encoded using different GRUs. The interaction history encoder can be represented in the following function:

$$\begin{aligned}
 h_{os} &= \text{GRU}(p_{os}^{t-\delta t_1:t}) \\
 h_{ts} &= \text{GRU}(p_{ts}^{t-\delta t_1:t})
 \end{aligned} \quad (6)$$

where  $p_{os}^{t-\delta t_1:t}$  and  $p_{ts}^{t-\delta t_1:t}$  are the positions of the own ship and target ship from time  $t - \delta t_1$  to  $t$ , respectively. The  $\text{GRU}()$  is the GRU network by applying (5) recursively and, thus,  $h_{os}$  and  $h_{ts}$  are the final hidden states.

2) *Future Trajectory Encoder*: The future trajectory encoder outputs the mean and variance of the latent vector  $z$  by encoding the future trajectory conditioning on the outputs of the interaction history encoder, which can be represented as follows:

$$\begin{aligned}
 h_{ts}^f &= \text{GRU}(p_{ts}^{t:t+\delta t_2}) \\
 \mu, \sigma &= \text{Linear}(h_{ts}^f, h_{ts}, h_{os})
 \end{aligned} \quad (7)$$

where  $\text{Linear}()$  is a linear mapping.

3) *Future Trajectory Decoder*: The future trajectory decoder recursively generates the future position by taking the hidden state and the predicted position from the last time step. The initial hidden state and position are obtained as follows:

$$\begin{aligned}
 h_0 &= \text{Linear}(z, h_{ts}, h_{os}) \\
 \hat{p}_0 &= p_{ts}^t
 \end{aligned} \quad (8)$$

The future trajectory, therefore, can be generated by recursively applying the following function:

$$\begin{aligned}
 h_{t+1} &= \text{gru}(\hat{p}_t, h_t) \\
 \hat{p}_{t+1} &= \text{Linear}(h_{t+1})
 \end{aligned} \quad (9)$$

where  $\text{gru}()$  is the same function as (5).

In summary, the interaction history encoder is designed to address 1) and 2). That is, the future trajectory depends on the past trajectories of both the own ship and the target ship. To address 3), we are interested to learn a conditional probabilistic distribution  $p(p_{ts}^{t:t+\delta t_2} | p_{ts}^{t-\delta t_1:t}, p_{os}^{t-\delta t_1:t})$  based on the CVAE framework. In particular, the future trajectory encoder is only used in the training phase and not in the inference phase. In the inference phase, we instead randomly sample the latent vector  $z$  from the multivariate standard Gaussian distribution  $z \sim \mathcal{N}(0, I)$ . Taking the random vector  $z$  and the conditional vectors from the target ship and the own ship, the future trajectory decoder is able to generate the future trajectory.

**Algorithm 1: Training.**


---

**Input:** Dataset  $\left\{ \left( p_{os}^{t-\delta t_1:t}, p_{ts}^{t-\delta t_1:t}, p_{ts}^{t:t+\delta t_2} \right) \right\}_{i=1}^N$ , interaction history encoder  $\phi_1$ , future trajectory encoder  $\phi_2$ , future trajectory decoder  $\theta$ .

- 1: **while** not converge **do**
- 2: Sample  $\left( p_{os}^{t-\delta t_1:t}, p_{ts}^{t-\delta t_1:t}, p_{ts}^{t:t+\delta t_2} \right)$  from the dataset.
- 3: Calculate predicted trajectory  $\hat{p}_{ts}^{t:t+\delta t_2}$ , mean  $\mu$  and variance  $\sigma$  of the latent vector by applying (6)–(9).
- 4: Calculate the loss function as in (4).
- 5: Update  $\phi_1$  and  $\phi_2$ ,  $\theta$  by gradient descent.
- 6: **end while**
- 7: **return**  $\phi_1, \theta$

---

**Algorithm 2: Trajectory Generation.**


---

**Input:** Past trajectory  $\left( p_{os}^{t-\delta t_1:t}, p_{ts}^{t-\delta t_1:t} \right)$ , interaction history encoder  $\phi_1$ , future trajectory decoder  $\theta$ .

- 1: Sample  $z$  from  $z \sim \mathcal{N}(0, I)$ .
- 2: Calculate  $h_0$  and  $\hat{p}_0$  from (6) and (8).
- 3: Generate trajectory  $\hat{p}_{ts}^{t:t+\delta t_2}$  by recursively applying (9).
- 4: **return**  $\hat{p}_{ts}^{t:t+\delta t_2}$

---

Therefore, by randomly generating the latent vector  $z$ , we are able to generate multiple future trajectories that enable a future trajectory distribution. Algorithm 1 and Algorithm 2 present the pseudocode for training the model and generating trajectory from the model, respectively.

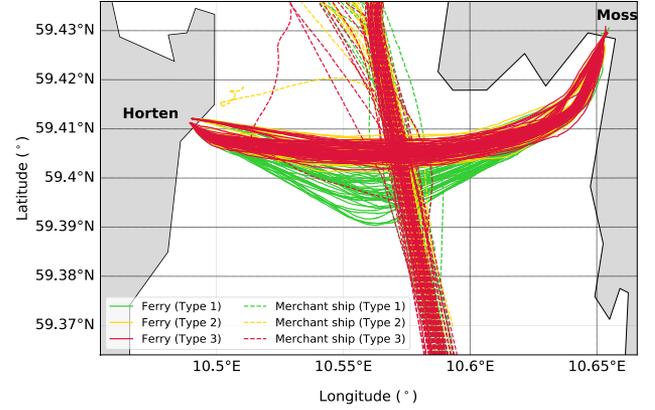
## IV. EXPERIMENTS

In this section, we present experiments on AIS data in two-vessel encounter scenarios to show the effectiveness of the proposed model.

### A. Dataset

The vessel trajectory data are collected from AIS data. The raw AIS data are retrieved from the database of the Norwegian Coastal Administration (Kystverket). The raw data contain essential information on longitude and latitude coordinates, speed over ground, course over ground, true heading, and static information, such as maritime mobile service identity (MMSI) and ship length. Since the raw AIS data are broadcast at a different frequency and contain anomaly data and stationary ship data, the anomaly and stationary data are filtered out and the resampling and downsampling are performed to bring the AIS data to 0.1 Hz.

In this study, we focus on the encounters between ferries crossing from Horten to Moss and merchant vessels navigating from the North Sea toward Oslo or Svelvik. The trajectories of the ferries and the encountered ships are extracted. The data for the whole month of January 2019 are extracted as the training set, while the first ten days of February 2019 are used as the test set. This results in 173 encounter cases in the training set



**Fig. 3.** Encountered trajectory data. The ferry travels from Horten to Moss, while the merchant ship travels from south to north. In such scenarios, the ferry is responsible to give way.

**TABLE I**  
DESCRIPTION OF THE ENCOUNTER TYPES IN THE DATASET

Encounter type	Description
Type 1	Obvious course alternation, give way to the merchant ship.
Type 2	No obvious course alternation, passing behind the merchant ship.
Type 3	No obvious course alternation, passing ahead the merchant ship.

and 63 encounter cases in the test set. Each encounter case is approximately 35 min. The trajectories of the ferry and the merchant ship are shown in Fig. 3. In this article, only longitude and latitude coordinates are used. The goal is to make future trajectory predictions for ferries only since the dataset only contains ferries from Horten to Moss and merchant ships from bottom to top (see Fig. 3). In this case, the ferry is responsible for deciding whether to give way or go straight through according to the Convention on the International Regulations for Preventing Collisions at Sea.

Although more ships can be involved in an encounter and relatively large datasets can be easily collected since the AIS data are publicly available, we only focus on this small dataset with two-ship interactions because it is labeled and reviewed by experts. It is more convenient to analyze whether the model can predict the trajectory under different encounter situations. Future work will include more ships and larger datasets. Note that there are three types of encounters as defined in [26] in the dataset. Table I lists a detailed description of these three encounter types. While Type 2 is similar to Type 3 in trajectory, they differ significantly in how ships pass.

### B. Implementation Details

For the interaction history encoder, future trajectory encoder, and future trajectory decoder, a two-layer GRU unit with a hidden size of 256 is used. The position of the vessel is normalized with  $z$ -score normalization. The Adam with decoupled weight decay regularization is used as the optimizer. A cosine annealing schedule with an initial learning rate of  $1 \times 10^{-3}$  is used. The model is trained for 1000 epochs.

TABLE II  
QUANTITATIVE RESULTS OF ALL THE METHODS ON THE DATASET (UNIT: M)

Model		KM	seq2seq	S-LSTM	proposed
ADE	Type 1	230.48	131.01	100.22	<b>93.02</b>
	Type 2	214.74	132.67	115.46	<b>106.25</b>
	Type 3	224.35	114.00	99.11	<b>87.53</b>
	Total	223.37	122.69	101.69	<b>94.79</b>
FDE	Type 1	608.05	176.46	159.83	<b>149.67</b>
	Type 2	567.20	194.55	190.45	<b>182.51</b>
	Type 3	568.62	158.41	160.44	<b>142.51</b>
	Total	587.77	172.70	163.30	<b>157.82</b>
BoN-ADE	Type 1	-	169.10	59.41	<b>50.96</b>
	Type 2	-	169.28	82.72	<b>69.81</b>
	Type 3	-	153.51	63.64	<b>50.35</b>
	Total	-	160.97	67.86	<b>56.92</b>
BoN-FDE	Type 1	-	116.66	71.98	72.53
	Type 2	-	136.49	117.37	<b>105.38</b>
	Type 3	-	108.72	85.42	<b>66.98</b>
	Total	-	118.50	91.43	<b>80.46</b>

The bold values indicates the best performance within the comparison.

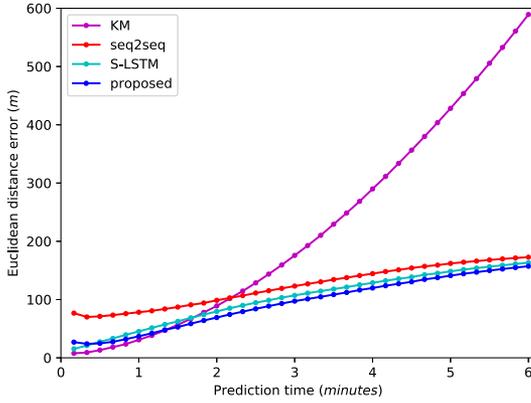


Fig. 4. Prediction performance versus prediction time horizon. Since our model produced samples on the predicted trajectories, we computed the average trajectory to calculate the Euclidean distance error.

### C. Quantitative Performance

In this part, the performance of the proposed model is evaluated quantitatively. The longitude and latitude coordinates in the dataset are actually converted to meters for training and testing the proposed model.

1) *Baselines*: The performance of our model is compared with several baselines.

- 1) *Kinematic model (KM)*: The model simply extrapolates trajectories with the assumption of constant speed and course direction.
- 2) *seq2seq model*: The seq2seq model follows an encoder–decoder structures. The encoder is used to encode the past trajectory, while the decoder is used to generate the future trajectory. The encoder and decoder are parameterized by RNNs. Note that even though this model does not consider the interaction between agents, it has been used for trajectory prediction on vehicles [11] and vessels [27]. We parameterize the model using a GRU with the same hidden size and layers as our proposed model. In addition, Monte-Carlo dropout [28] is used to generate multiple future trajectories from the model.

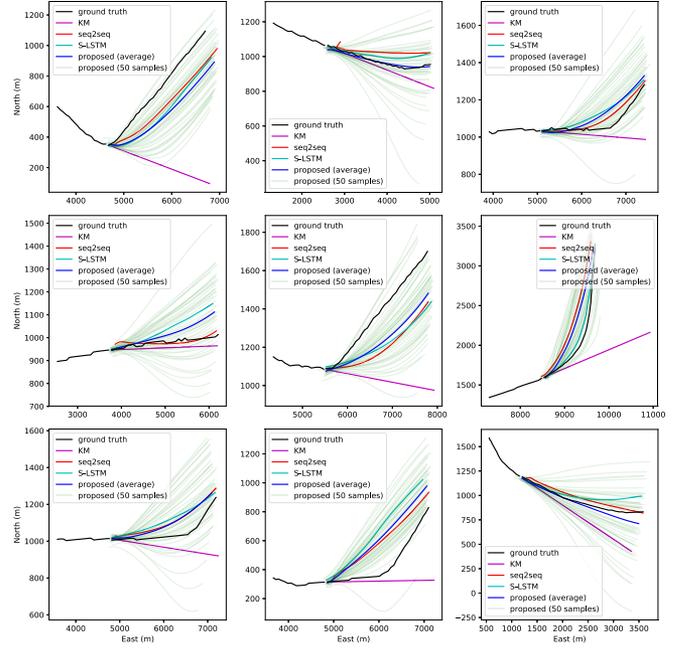


Fig. 5. Illustration of the predicting trajectories based on different methods. We randomly draw nine samples in the test dataset. The black line indicates the ground truth trajectory of the vessel.

- 3) *Social LSTM (S-LSTM)*: S-LSTM [9] models that each agent by an individual LSTM with the hidden states at each time step is shared. Since only two vessels are considered, we do not use the pooling mechanism but directly share the hidden states. Besides, the LSTM is changed to GRU. This model is trained and inferred in an autoregressive manner. The future trajectories at each time step are assumed to follow a Gaussian distribution and the model is trained with negative log likelihood.

2) *Evaluation Metrics*: Two different metrics that are widely used for trajectory prediction are considered:

- 1) *Average distance error (ADE)*: The average Euclidean distance of all estimated and true points of the trajectory.
- 2) *Final distance error (FDE)*: The Euclidean distance between the predicted final destination and the true final destination at the end of the forecast period.

Since the proposed model can predict the future trajectory distribution, we sample 50 trajectories from our model and compute the average trajectory to evaluate ADE and FDE. Furthermore, we evaluate the predictive distributions using the Best-of-N (BoN) ADE and FDE, which we denote as BoN-ADE and BoN-FDE, respectively. BoN-ADE and BoN-FDE were proposed in [12]. This is a way to evaluate whether multiple predicted trajectories cover the true one. Here, we sample 50 future trajectories and compute their errors from the true trajectories to obtain the best five trajectories. The errors for these five trajectories indicate how well our multiple predicted trajectories match the true trajectories.

3) *Results*: Fig. 4 shows the Euclidean distance error at different prediction time horizons. It is shown that the error of the KM increases dramatically with prediction time horizons.

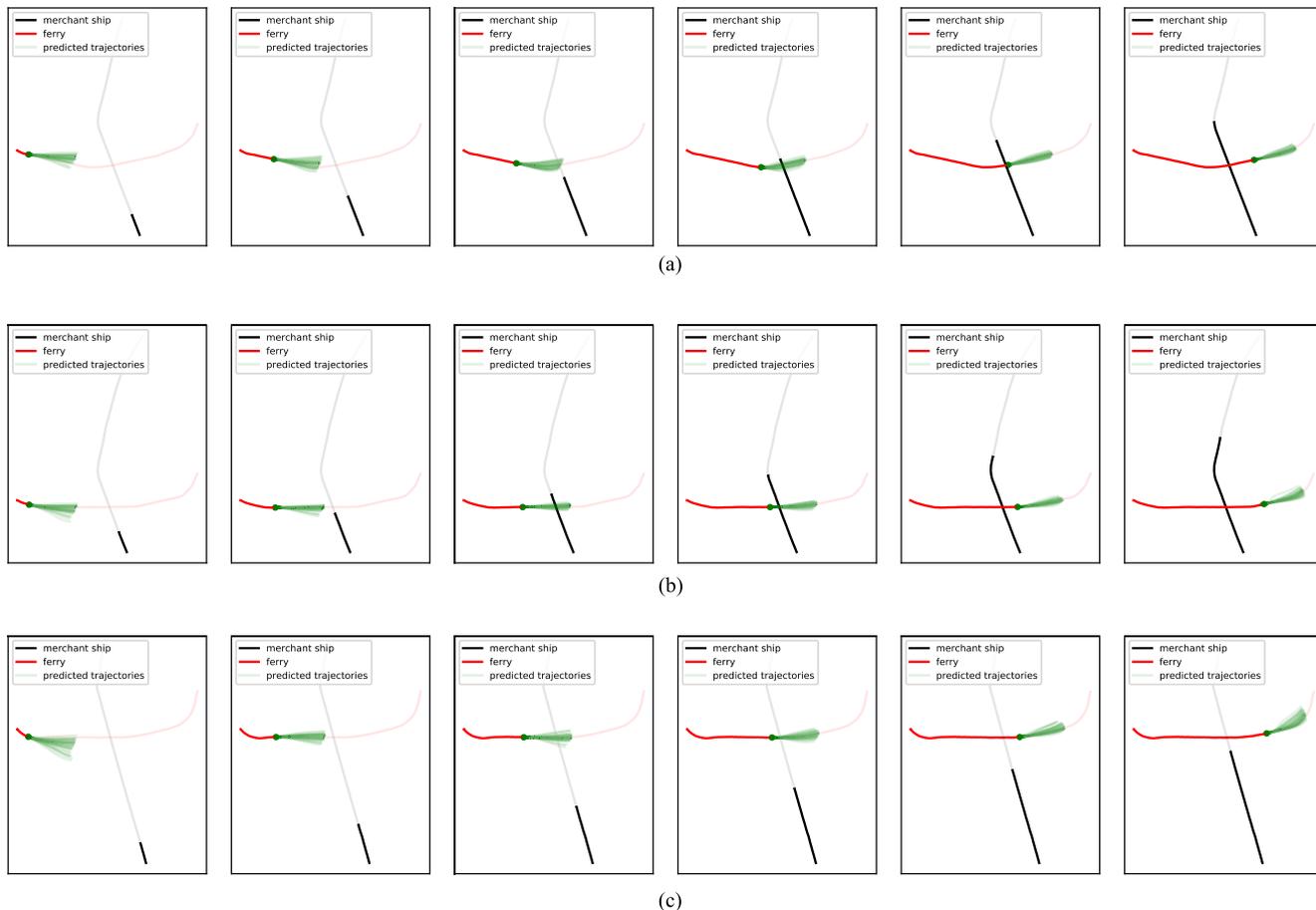


Fig. 6. Ferry trajectory predictions at different time steps for different encounter types. (a) Type 1 encounter. (b) Type 2 encounter. (c) Type 3 encounter.

When the prediction time exceeds 1 min, the proposed model provides the smallest error among all the baseline models.

In Table II, we compare the performance of our model with the baseline methods. The naive KM produces high prediction errors. The pattern-based methods clearly outperform the KM. While seq2seq has similar errors on ADE and FDE, it does not perform well on the BoN metric, suggesting that Monte-Carlo dropout has difficulty modeling predictive distributions. This may be due to the fact that the method is often used to approximate epistemic rather than aleatoric uncertainty. The S-LSTM and our proposed model outperform seq2seq models, demonstrating the importance of involving vessel interaction. The proposed model provides the smallest error among all metrics, which shows the superiority of the proposed model in modeling the behavior of marine vessels.

#### D. Trajectory Prediction Analysis

The quantitative evaluation shows that the proposed model outperforms other baseline methods. In this part, the actual behavior of the proposed model in different settings is analyzed.

1) *Trajectory Prediction With Different Baselines:* In Fig. 5, the prediction results of the proposed model and the baseline methods are shown in several random samples from the test set.

It is shown that the KM model produces high prediction errors, especially around nonlinear regions. The other learning-based methods are able to predict nonlinear behaviors. The sampled trajectories from the proposed model are able to match the real trajectory and the average of these sampled trajectories provides the least error.

2) *Trajectory Prediction on Different Time Steps:* Fig. 6 presents the predictions on random scenarios at different time steps from type 1, type 2, and type 3 encounters. The results show that across all encounter types, the predicted uncertainty is lower after the ferry passes the merchant ship, which fits with our intuition that the ferry only needs to focus on its destinations after passing. In addition, the predictions show that Type 2 and Type 3 have less uncertainty before passing than Type 1, possibly due to the ferry's ability to quickly recognize that it can safely pass the merchant ship without changing course.

3) *Trajectory Prediction With the Change of Encounter Type:* To analyze the change of encounter, we sample 3 encounter scenarios from type 1, type 2, and type 3, respectively. The type 3 scenario is linearly interpolated to the type 1 scenario, as shown in Fig. 7(a). The type 1 scenario is linearly interpolated to the type 2 scenario, as shown in Fig. 7(b). Since the ferry has almost the same trajectory for Type 2 and Type 3 encounters,

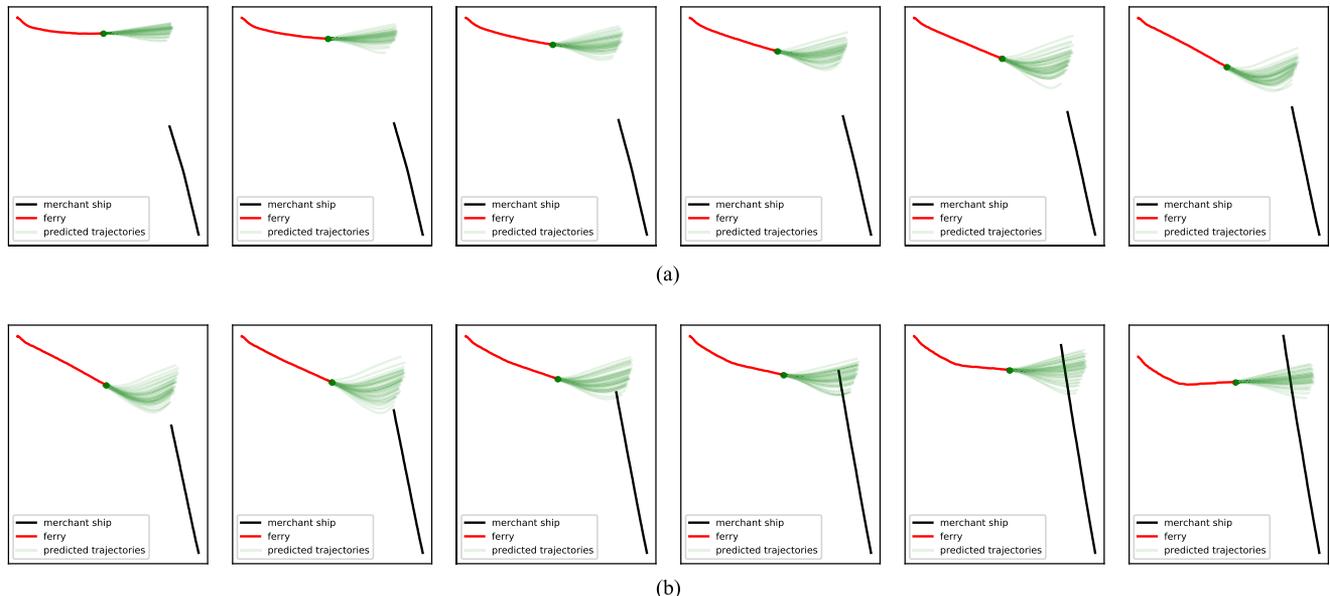


Fig. 7. Ferry trajectory predictions when the encounter type changes. (a) From type 3 to type 1 encounter. (b) From type 1 to type 2 encounter.

TABLE III  
EFFECT OF DIFFERENT AGGREGATION METHODS (UNITS: M)

Aggregation	No interaction	Sum	Max pool	Proposed
ADE	109.11	98.28	97.65	<b>94.79</b>
FDE	167.93	159.10	157.93	<b>157.82</b>
BoN-ADE	70.72	65.83	65.24	<b>56.92</b>
BoN-FDE	89.02	85.74	86.93	<b>80.46</b>

The bold values indicates the best performance within the comparison.

TABLE IV  
EFFECT OF DIFFERENT LOOKBACK WINDOWS (UNITS: M)

Lookback	@-9min	@-6min	@-3min	@-1min	@-30s
ADE	<b>92.09</b>	92.93	94.79	99.48	108.76
FDE	<b>150.36</b>	154.73	157.82	163.18	173.35
BoN-ADE	57.61	61.87	<b>56.92</b>	59.86	62.04
BoN-FDE	<b>80.22</b>	87.96	80.46	84.40	79.90

The bold values indicates the best performance within the comparison.

the transition between these two encounters is not considered. In Fig. 7(a), it can be found that the merchant ship barely changes, and when the ferry approaches the merchant ship, the predicted trajectory changes from passing directly to giving way. Similarly, in Fig. 7(b), the ferry has little changes, and when the merchant ship moves further, the predicted trajectory changes from giving way to passing. These qualitative results demonstrate that our model is able to capture the interaction between two vessels.

### E. Ablation Study

To validate the design choice of how we aggregate the interaction information, an ablation study is performed. In particular, we performed experiments on the following.

- 1) No interaction is considered.
- 2) The vectors from two vessels are summed.
- 3) The vectors from two vessels are max pooled.

Table III presents that including the interaction improves the prediction performance. In addition, simply concatenating the vectors from two vessels provide the best performance in the proposed model compared with other aggregation methods.

To evaluate the lookback windows in Fig. 2, four different lookback windows were evaluated in addition to 3 min. As shown in Table IV, as the lookback window increases, the prediction error tends to decrease. It shows that the model successfully

extracts temporal information rather than fitting temporal noise. However, the improvement in forecasts is not significant when the lookback window is longer than 3 min.

## V. CONCLUSION

In this article, an interactive-aware short-term trajectory prediction model for marine vessels was proposed. The model followed the CVAE framework and, thus, was able to model the inherent uncertainty of the forecasting process. By sampling from the latent space, it can quickly generate multiple future trajectories. The interaction was encoded into a context vector for the model. The model was implemented in a seq2seq manner to account for time-series data. Experiments were performed on real-world two-ship encounter AIS data. The proposed method outperformed the baseline methods. In addition, we qualitatively showed that our model successfully models the uncertainty as well as the interactions.

Future work will extend our model to multivessel encounter scenarios over larger areas. More stimuli can be considered, such as ship types, semantic maps, and weather conditions. More broadly, there are also architectural considerations when large datasets are involved and integration with downstream planning tasks when predictive models are available.

## REFERENCES

- [1] H. Zhang et al., "A digital twin of the research vessel Gunnerus for lifecycle services: Outlining key technologies," *IEEE Robot. Autom. Mag.*, early access, Nov 14, 2022, doi: 10.1109/MRA.2022.3217745.
- [2] E. Jokioinen et al., "Remote and autonomous ships—the next steps," Rolls-Royce plc, 2016. [Online]. Available: <http://www.rolls-royce.com/~media/Files/R/Rolls-Royce/documents/customers/marine/ship-intel/aawa-whitepaper-210616.pdf>
- [3] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrilu, and K. O. Arras, "Human motion trajectory prediction: A survey," *Int. J. Robot. Res.*, vol. 39, no. 8, pp. 895–935, 2020.
- [4] S. Fossen and T. I. Fossen, "Extended Kalman filter design and motion prediction of ships using live automatic identification system (AIS) data," in *Proc. IEEE 2nd Eur. Conf. Elect. Eng. Comput. Sci.*, 2018, pp. 464–470.
- [5] B. Murray and L. P. Perera, "A dual linear autoencoder approach for vessel trajectory prediction using historical AIS data," *Ocean Eng.*, vol. 209, 2020, Art. no. 107478.
- [6] S. Capobianco, L. M. Millefiori, N. Forti, P. Braca, and P. Willett, "Deep learning methods for vessel trajectory prediction based on recurrent neural networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 6, pp. 4329–4346, Dec. 2021.
- [7] S. Gao, Y. Dai, Y. Li, Y. Jiang, and Y. Liu, "Augmented flame image soft sensor for combustion oxygen content prediction," *Meas. Sci. Technol.*, vol. 34, no. 1, 2022, Art. no. 015401.
- [8] K. Liu, M. Zheng, Y. Liu, J. Yang, and Y. Yao, "Deep autoencoder thermography for defect detection of carbon fiber composites," *IEEE Trans. Ind. Inform.*, vol. 19, no. 5, pp. 6429–6438, May 2023.
- [9] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 961–971.
- [10] F. Alché and A. de la Fortelle, "An LSTM network for highway trajectory prediction," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst.*, 2017, pp. 353–359.
- [11] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder-decoder architecture," in *Proc. IEEE Intell. Veh. Symp.*, 2018, pp. 1672–1678.
- [12] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2255–2264.
- [13] N. Rhinehart, R. McAllister, K. Kitani, and S. Levine, "PRECOG: Prediction conditioned on goals in visual multi-agent settings," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 2821–2830.
- [14] B. Ivanovic, K. Leung, E. Schmerling, and M. Pavone, "Multimodal deep generative models for trajectory prediction: A conditional variational autoencoder approach," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 295–302, Apr. 2021.
- [15] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Eur. Conf. Comput. Vis.*, 2020, pp. 683–700.
- [16] H. Zhao et al., "TNT: Target-driven trajectory prediction," in *Proc. Conf. Robot Learn.*, 2021, pp. 895–904.
- [17] T. Wang, G. Li, L. I. Hatledal, R. Skulstad, V. Æsøy, and H. Zhang, "Incorporating approximate dynamics into data-driven calibrator: A representative model for ship maneuvering prediction," *IEEE Trans. Ind. Inform.*, vol. 18, no. 3, pp. 1781–1789, Mar. 2022.
- [18] P. Han, G. Li, X. Cheng, S. Skjong, and H. Zhang, "An uncertainty-aware hybrid approach for sea state estimation using ship motion responses," *IEEE Trans. Ind. Inform.*, vol. 18, no. 2, pp. 891–900, Feb. 2022.
- [19] D.-W. Gao, Y.S. Zhu, J.-F. Zhang, Y.-K. He, K. Yan, and B.-R. Yan, "A novel MP-LSTM method for ship trajectory prediction based on AIS data," *Ocean Eng.*, vol. 228, 2021, Art. no. 108956.
- [20] D. Nguyen and R. Fablet, "TraISformer—A generative transformer for AIS trajectory prediction," 2021, *arXiv:2109.03958*.
- [21] R. W. Liu et al., "STMGCN: Mobile edge computing-empowered vessel trajectory prediction using spatio-temporal multigraph convolutional network," *IEEE Trans. Ind. Inform.*, vol. 18, no. 11, pp. 7977–7987, Nov. 2022.
- [22] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Proc. 28th Neural Inf. Process. Syst.*, 2015, pp. 3483–3491.
- [23] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014. [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [24] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. 2nd Int. Conf. Learn. Representations*, 2014. [Online]. Available: <http://arxiv.org/abs/1312.6114>.
- [25] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," in *Proc. {SSST}-8, 8th Workshop Syntax, Semantics Struct. Stat. Transl.*, Association for Computational Linguistics, 2014, pp. 103–111. [Online]. Available: <https://aclanthology.org/W14-4012>
- [26] M. Zhu, C. Wang, P. Han, R. Skulstad, H. Zhang, and G. Li, "An AIS data-driven based hybrid approach to ship trajectory prediction—A real case in Oslofjord," submitted to *Reliability and System Engineering*.
- [27] N. Forti, L. M. Millefiori, P. Braca, and P. Willett, "Prediction of vessel trajectories from AIS data via sequence-to-sequence recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 8936–8940.
- [28] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.



**Peihua Han** (Member, IEEE) received the bachelor's in civil engineering and master's degree in structural engineering from the Department of Architecture and Civil Engineering, Zhejiang University, Hangzhou, China, in 2016 and 2019, respectively, and the Ph.D. degree in engineering from the Norwegian University of Science and Technology (NTNU), Aalesund, Norway, in 2022.

He is currently a Postdoctor with Intelligent Systems Laboratory, NTNU. His current research interests include data mining, machine learning, time-series modeling, and uncertainty qualification.



**Mingda Zhu** (Member, IEEE) received the M.Sc. degree in ship design in 2021 from the Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology (NTNU), Aalesund, Norway, where he is currently working toward the Ph.D. degree in engineering with Intelligent Systems Lab, Department of Ocean Operations and Civil Engineering.

His research interests include ship modeling and simulation, onboard support system, and machine learning.



**Houxiang Zhang** (Senior Member, IEEE) received the Ph.D. degree in mechanical and electronic engineering and the Habilitation degree in informatics from the University of Hamburg, Hamburg, Germany, in 2003 and 2011, respectively.

Since 2004, he has been a Postdoctoral Fellow and a Senior Researcher with the Institute of Technical Aspects of Multimodal Systems, Department of Informatics, Faculty of Mathematics, Informatics and Natural Sciences, University of Hamburg, Germany. In April 2011, he joined the Norwegian University of Science and Technology (NTNU), Aalesund, Norway, where he is currently a Professor of mechatronics. He is currently a Full Professor with the Department of Ocean Operations and Civil Engineering, Faculty of Engineering, NTNU. From 2011 to 2016, he also held a Norwegian National GIFT Professorship on product and system design funded by the Norwegian Maritime Centre of Expertise. He has applied for and coordinated more than 30 projects supported by the Norwegian Research Council (NFR), German Research Council (DFG), EU, and industry. In these areas, he has authored or coauthored more than 250 journals and conference papers as the author or coauthor. He has engaged in two main research areas, including control, optimization, and AI application, especially on autonomous vehicles, marine automation, digitalization, and ship intelligence. In 2019, he was elected as a member of the Norwegian Academy of Technological Sciences.

Dr. Zhang was a recipient of four Best Paper Awards and five Finalist Awards for Best Conference Paper at International Conference on Robotics and Automation.