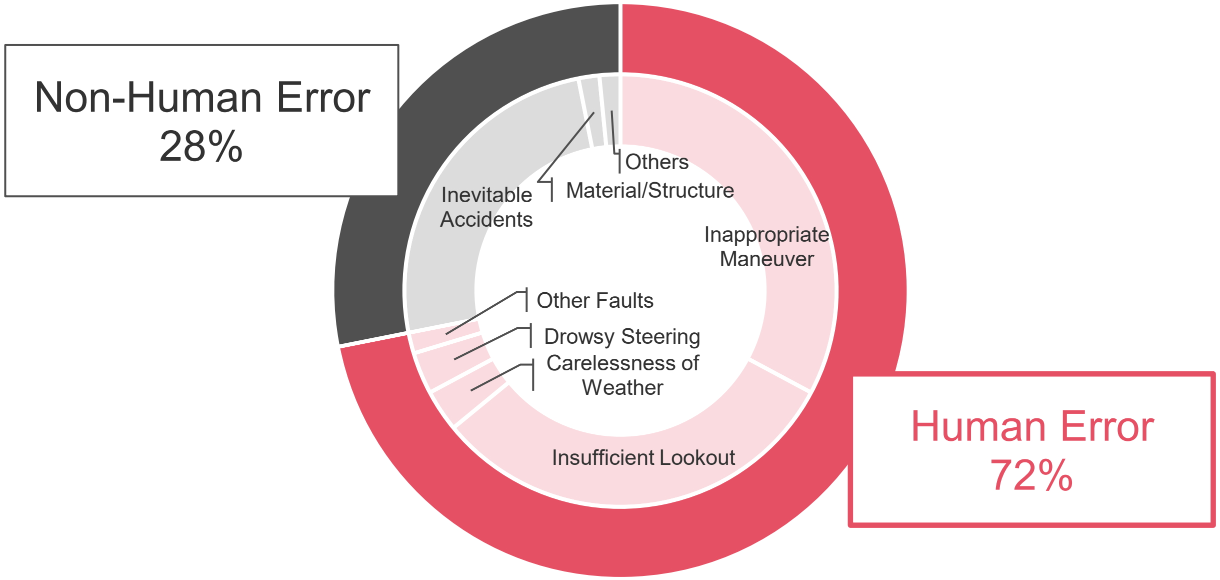
Data-Driven Approaches for Autonomous Shipping

Takefumi Higaki, Ph.D.
Assistant Processer, Osaka Metropolitan University

2026/03/03 JST ASPIRE Kick-off MTG

Importance of Automatic Collision Avoidance

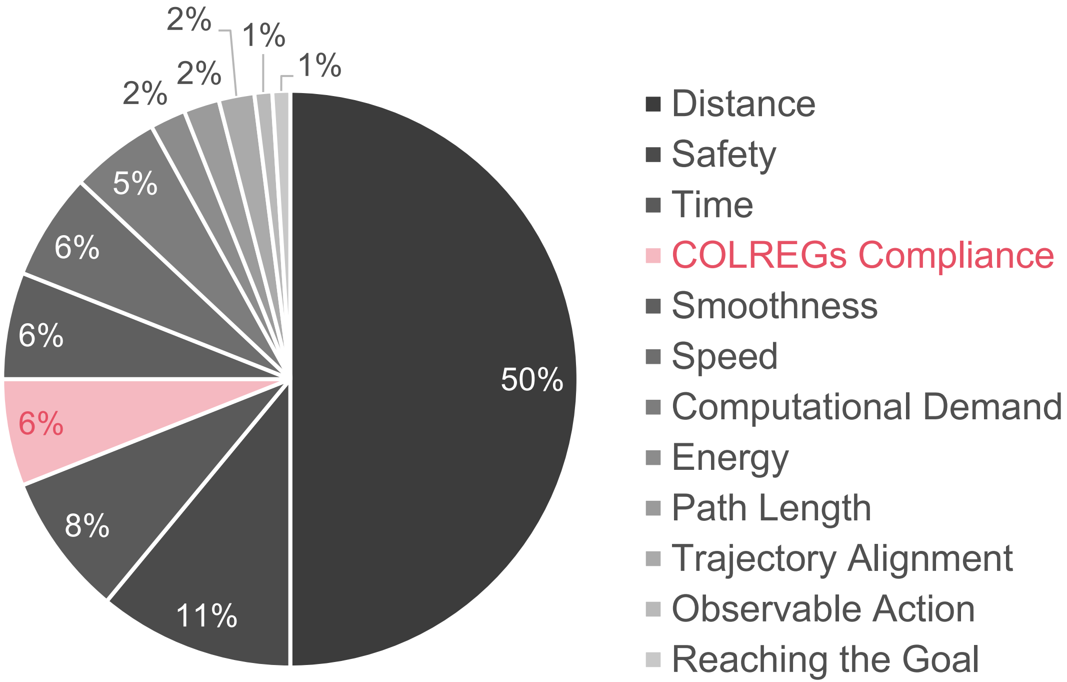
Most maritime collisions are triggered by human error.
→ Automatic collision avoidance is expected to eradicate maritime accidents.



Rate of collision accidents of cargo ship in Japan (2021)[1]

[1] Japan Cost Guard (2021). Current situation and measure of maritime accidents in 2021.

Difficulty of Optimization in Collision Avoidance



Types of objective functions defined in previous studies[2]

Collision avoidance involves various objectives, and therefore route optimization is challenging.

Considering the navigation rules (COLREGs) is especially difficult as they have inherent ambiguity.

Previous studies[3-5] have tried to formulate COLREGs, but only a part of rules have been considered.

[2] Öztürk et al. (2022). *Ocean Eng. 251*, 111010.
[3] Du et al., (2020). *Ocean Eng. 218*, 107866.
[4] Bakdi & Vanem (2022). *IEEE Trans. Intell. Transp. Syst. 23*, 18433–18445.
[5] Gleeson et al. (2024). *Ocean Eng. 313*, 119552.

Motivation & Approaches

Objective

We aim to address the difficulty in formulating objective functions for ship collision avoidance.

Approaches

To the end, we have presented three data-driven approaches that learn directly from expert data without explicit objectives.

- ✓ Inverse reinforcement learning (IRL): we inferred COLREGs-compliant rewards from skilled captains' maneuvering data[6].
- ✓ Imitation learning (IL): we achieved human-like collision avoidance without defining objective functions[7,8].
- ✓ Diffusion policy (DP): we developed a route planning method for realistic situations with multiple ships and geometric constraints[9].

[6] Higaki et al. (2022), *J. Japan Soc. Nav. Archit. Ocean Eng., 36*, 137–148.
[7] Higaki & Hashimoto (2023), *Appl. Ocean Res., 138*, 103620.
[8] Higaki et al. (2023), *Proceedings of SMATECH 2023*, 104–109.
[9] Higaki et al. (2025), *Proc. Annu. Spring Meet. JASNAOE 2025, 40*, 13–15.

Approach 1: Inverse Reinforcement Learning (IRL)

Reinforcement Learning (RL)

Aims to obtain an optimal policy π^* that maximizes the summation of rewards r.

$$\pi^* = \underset{\pi}{\operatorname{argmax}}\, E\left[\sum_{t'=t}^{\infty} \gamma^{t'-t} r_{t'+1}\right]$$

What kind of reward function should we define?
→ Setting appropriate rewards is difficult.

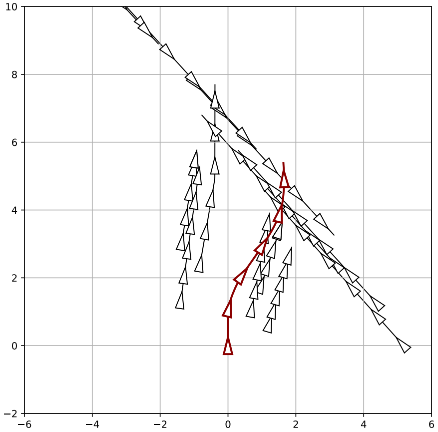
Inverse Reinforcement Learning (IRL)

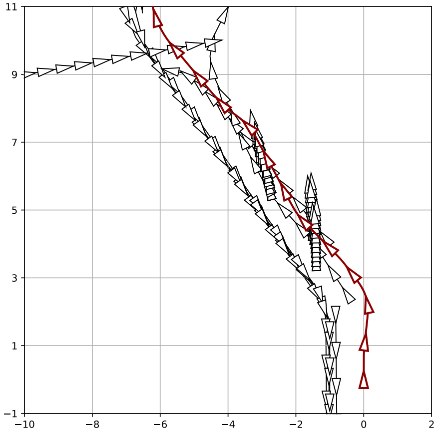
IRL estimates an optimal reward r^* from given expert data.

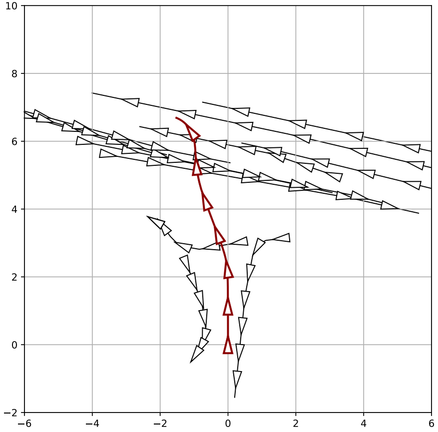
$$r^* = \underset{r}{\operatorname{argmax}}\Big(\underbrace{E_{\pi^E}[R(s)]}_{\substack{\text{Reward summation} \\ \text{under an expert's policy } \pi_E}} - \max \underbrace{E_{\pi}[R(s)]}_{\substack{\text{Reward summation} \\ \text{under an agent's policy } \pi}}\Big)$$

Reward represents how desirable the state is.
→ We can quantify what kind of states are desirable for human captains…?

Collection of Experts’ Maneuvering Data







Types of Target Ships

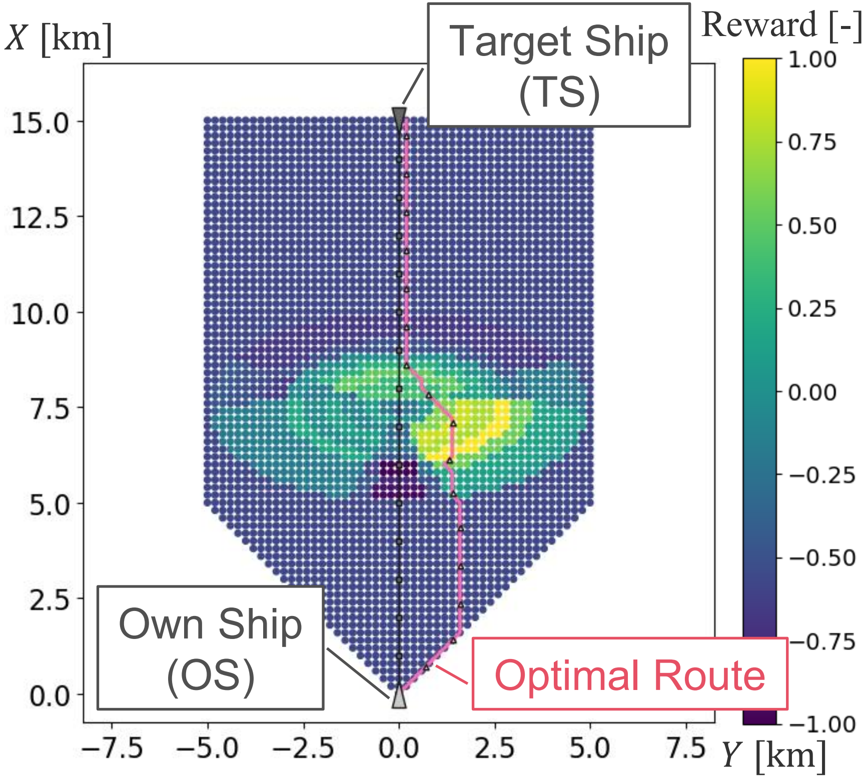
Encounter Type	N_{ship}
Over-taken	15
Head-on	9
Crossing from left	12
Crossing from right	8
Total	44

We conducted an experiment on a ship maneuvering simulator.

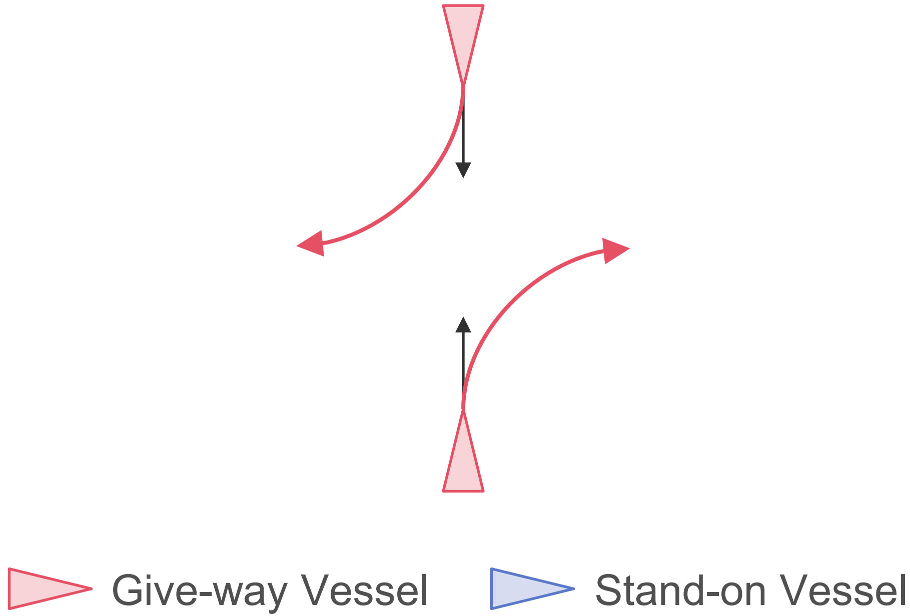
Skilled captains in a shipping company (NYK Line) steered a container ship ($L_{OA} = 350$ [m]) in a congested sea area.

[6] Higaki et al. (2022), *J. Japan Soc. Nav. Archit. Ocean Eng.*, *36*, 137–148.

Reward Distribution in a Head-on Case



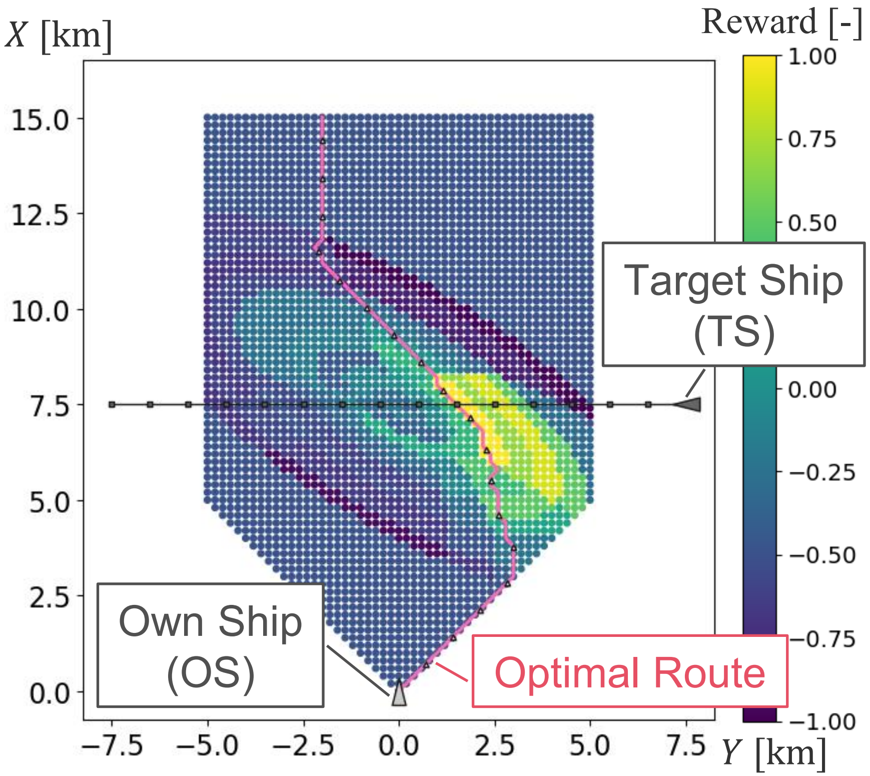
COLREGs Rule 14 (Head-on)



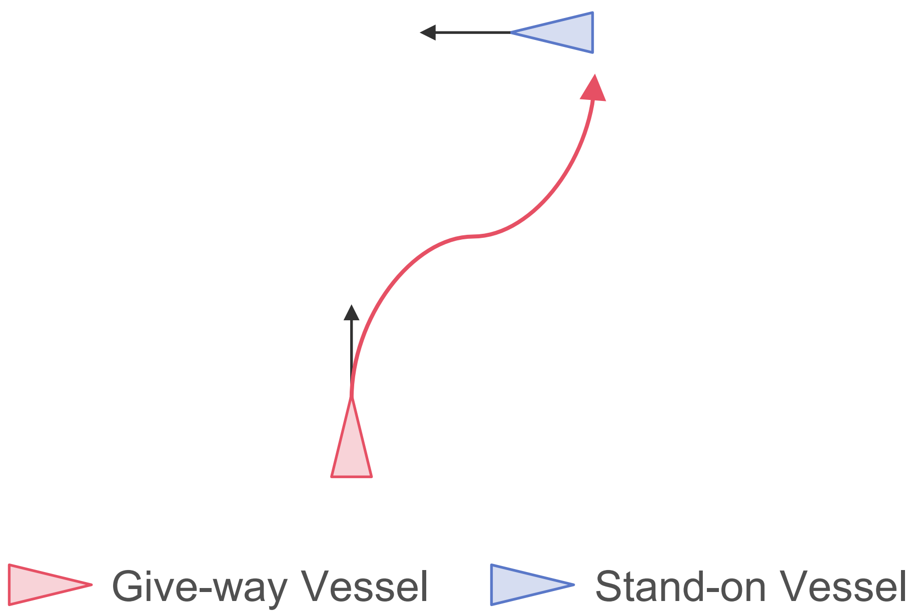
The reward distribution implies the experts likely turn right in a head-on case.
This trend is consistent with the right-hand traffic rule defined in COLREGs Rule 14.

[6] Higaki et al. (2022), *J. Japan Soc. Nav. Archit. Ocean Eng., 36*, 137–148.

Reward Distribution in a Crossing-from-Right Case



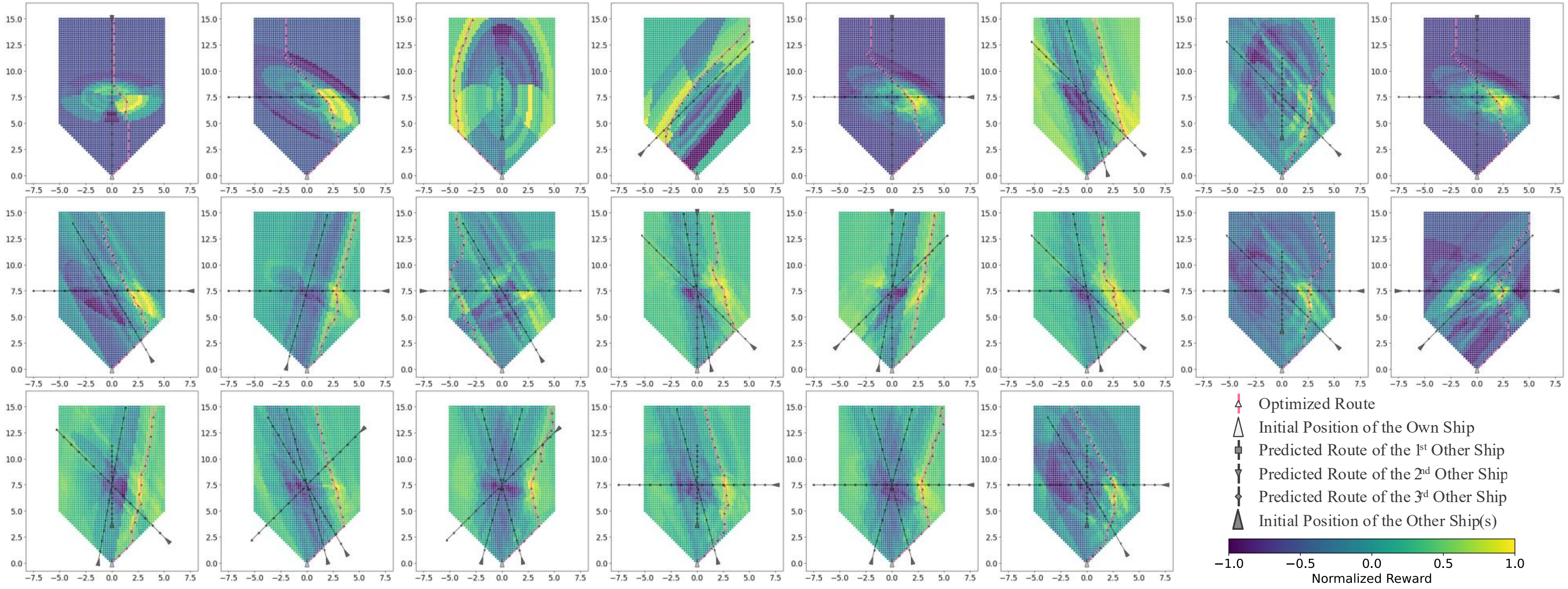
COLREGs Rule 15 (Crossing)



By passing the stern of the target ship, the own ship can get high rewards.
The rewards represent the duty of a give-way vessel in Rule 15.

[6] Higaki et al. (2022), *J. Japan Soc. Nav. Archit. Ocean Eng.*, *36*, 137–148.

Reward Distributions in Various Encounter Situations



We derived rewards that reflect human experts' preference.
No explicit violations of COLREGs Rules 13–17 were observed in the optimized routes.

[6] Higaki et al. (2022), *J. Japan Soc. Nav. Archit. Ocean Eng.*, *36*, 137–148.

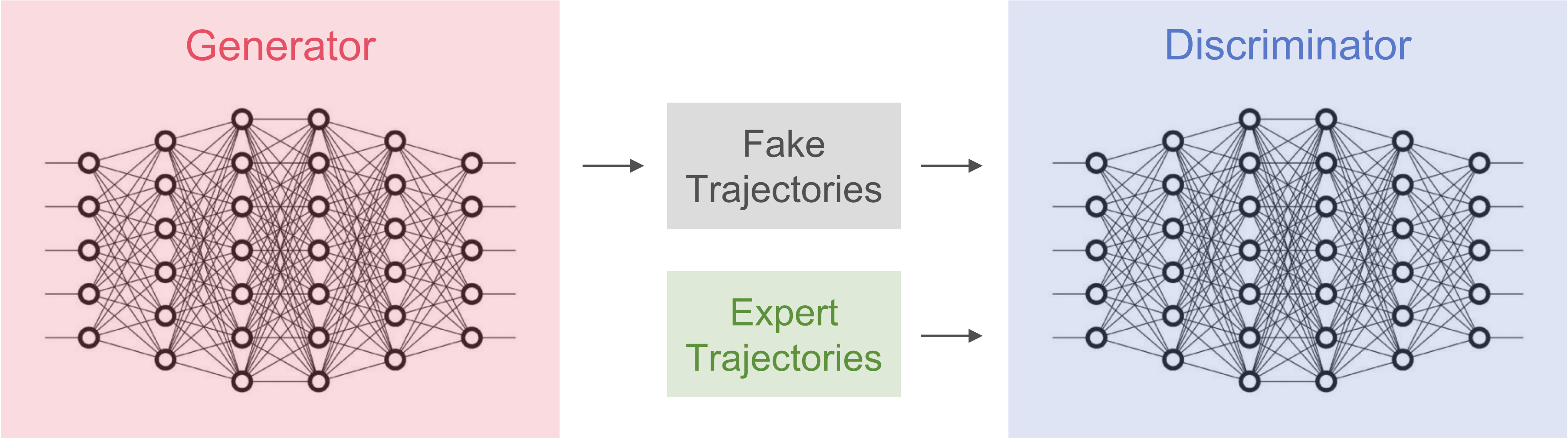
Approach 2: Imitation Learning (IL)

Inverse Reinforcement Learning (IRL)	Imitation Learning (IL)
✓ Quantifies / visualizes rewards.	✗ Cannot visualize rewards (intention).
✗ Limited to small & discrete states.	✓ Applicable to large & continuous states.

Introduction to Imitation Learning (IL)

Generative Adversarial Imitation Learning (GAIL[10])

IL method that trains two adversarial networks: generator and discriminator.

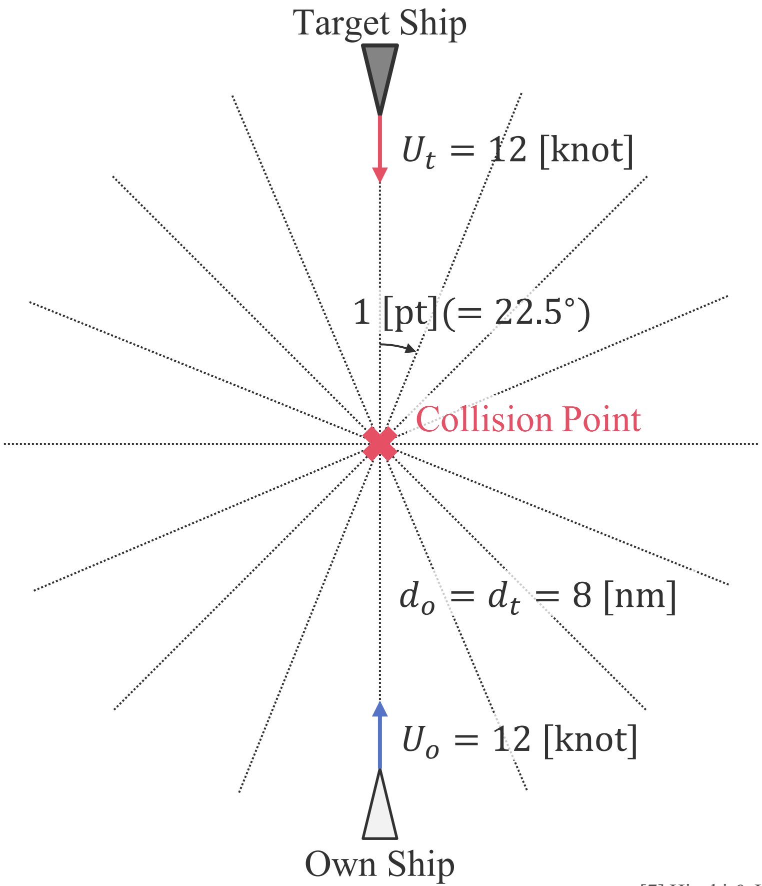


Learns fake actions so as not to be distinguished from true expert data.

Discriminates whether given data is made by true expert or generator.

[10] Ho & Ermon (2016). *Advances in Neural Information Processing Systems, 29*.

Collection of Experts' Maneuvering Data

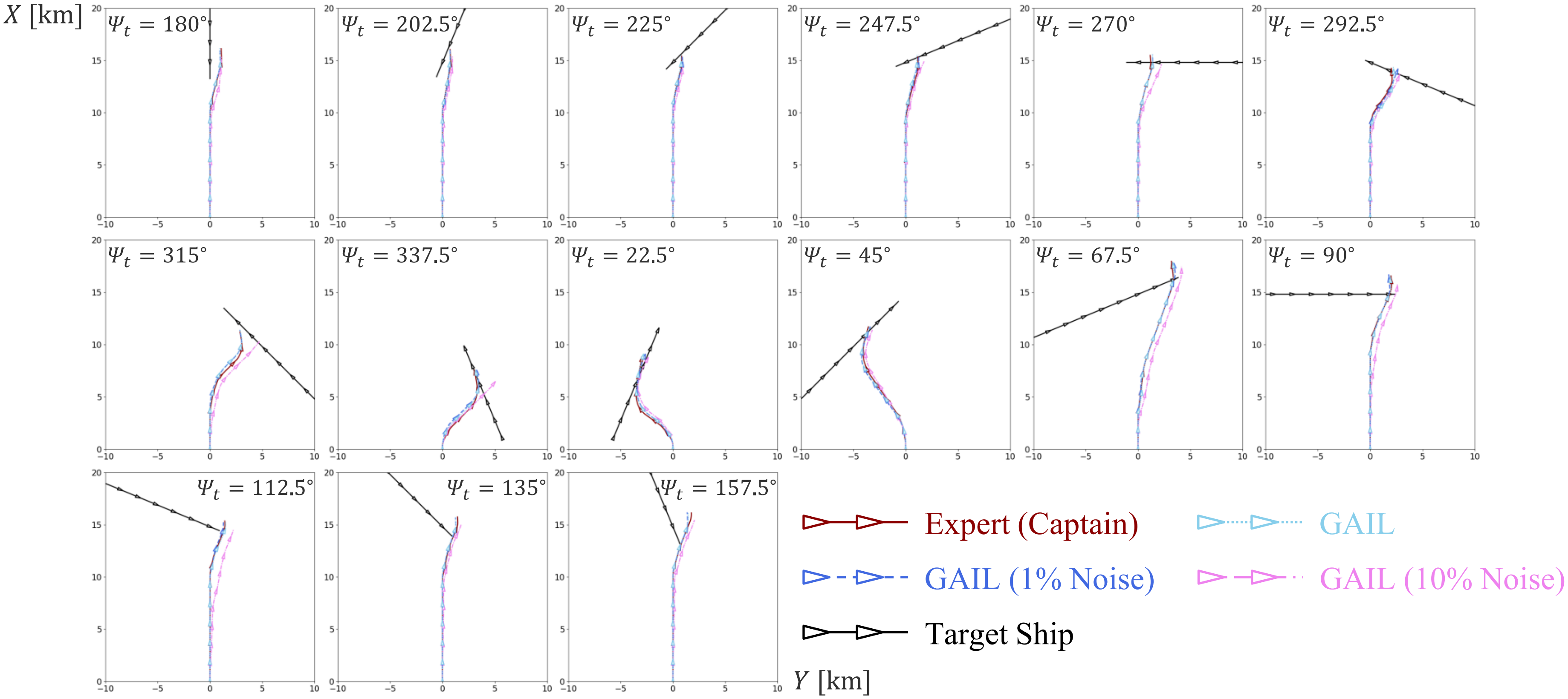


We conducted another experiment on the ship maneuvering simulator.

A skilled human captain steered the container ship ($L_{OA} = 350$ [m]) in 15 scenarios with a target ship.

[7] Higaki & Hashimoto (2023), *Appl. Ocean Res., 138*, 103620.

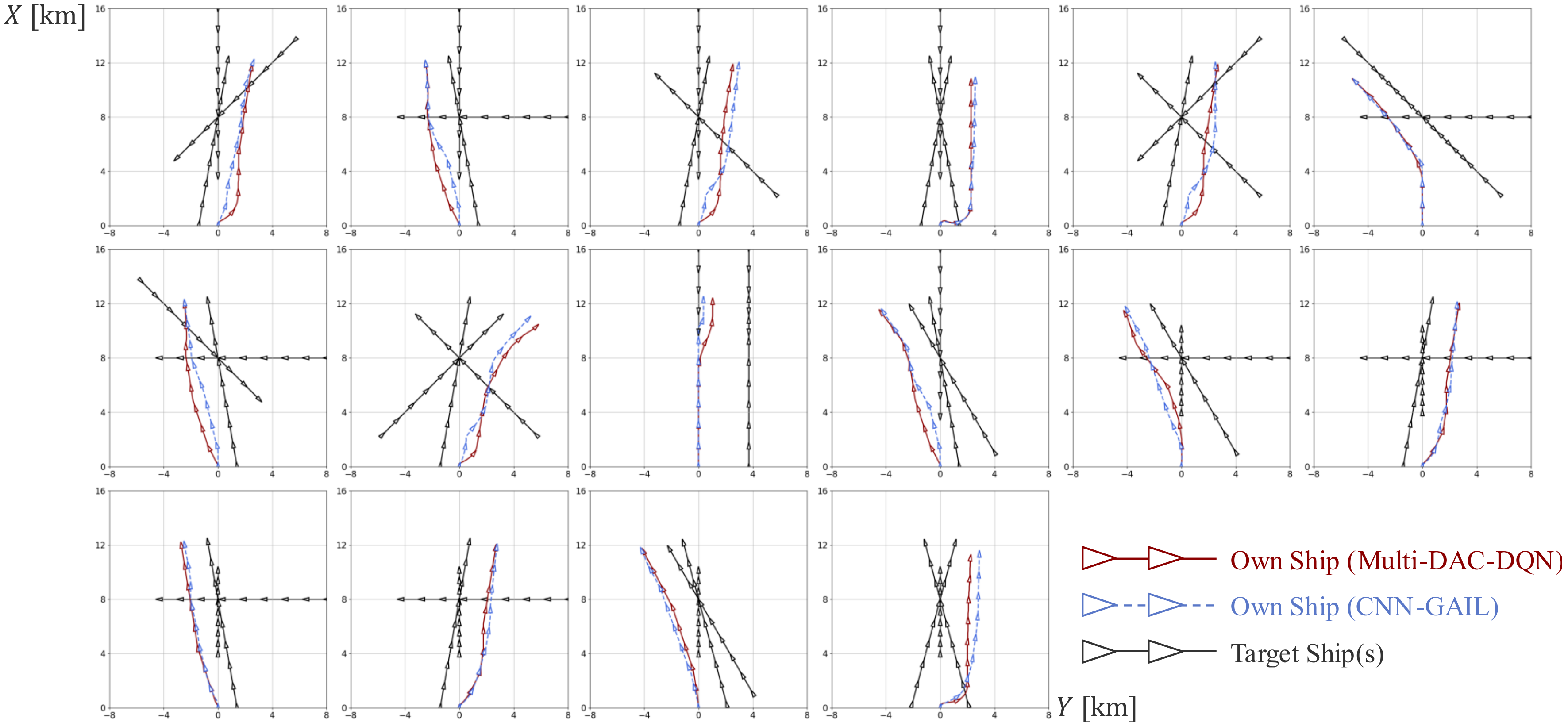
Collision Avoidance Trajectories (1 on 1 Scenarios)



The imitative route planner appropriately avoided collisions like human captain.

[7] Higaki & Hashimoto (2023), *Appl. Ocean Res.*, *138*, 103620.

Collision Avoidance Trajectories (Congested Scenarios)



The imitative route planner showed its applicability to congested situations.

[8] Higaki et al. (2023), *Proceedings of SMATECH 2023*, 104–109.

Approach 3: Diffusion Policy (DP)

IRL / IL	Diffusion Policy (DP)[12]
✓ Requires only a small amount of expert data. ✓ Prediction is fast.	✗ Requires moderate amount of data. ✗ Prediction takes relatively longer time.
✗ Discrepancy between the actual and virtual environments can degrade the learning performance. ✗ The policy's expressiveness is limited.	✓ No need for modeling the environment. ✓ Represents multi-modal policies via an energy function.

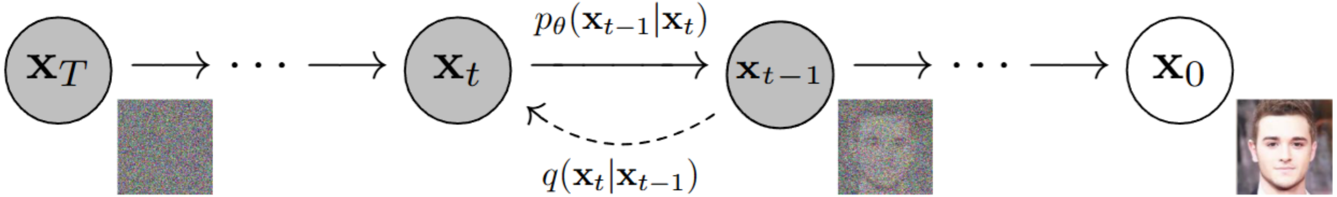
[12] Chi et al. (2024). *Int. J. Robot. Res*.

Introduction to Diffusion Policy (DP)

Diffusion Model[13]

Generative AI that creates data by learning to reverse a process of gradually adding random noises to original data.

In an image generation task, it commonly generates an image from natural language.



Diffusion Policy[12]

Learns to generate an action a_t conditioned on an observation o_t within the framework of diffusion model.

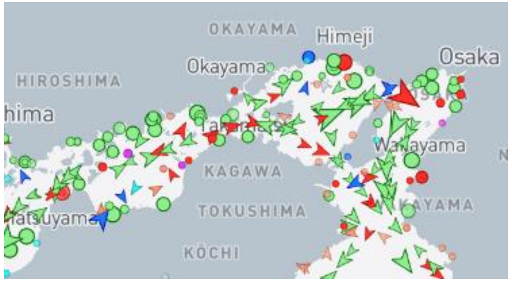
$$a_t^{(k-1)} = \frac{1}{\sqrt{1-\beta^{(k)}}}\left(a_t^{(k)} - \frac{\beta^{(k)}}{\sqrt{1-\bar{\alpha}^{(k)}}}\varepsilon_\theta\left(o_t, a_t^{(k)} k\right)\right)$$

[12] Chi et al. (2024). *Int. J. Robot. Res.*

[13] Ho et al. (2020). *Adv. Neural Inf. Process. Syst.*, *33*, 6840–6851.

Collection of Experts’ Maneuvering Data

Fukae-Maru



Seto Inland Sea

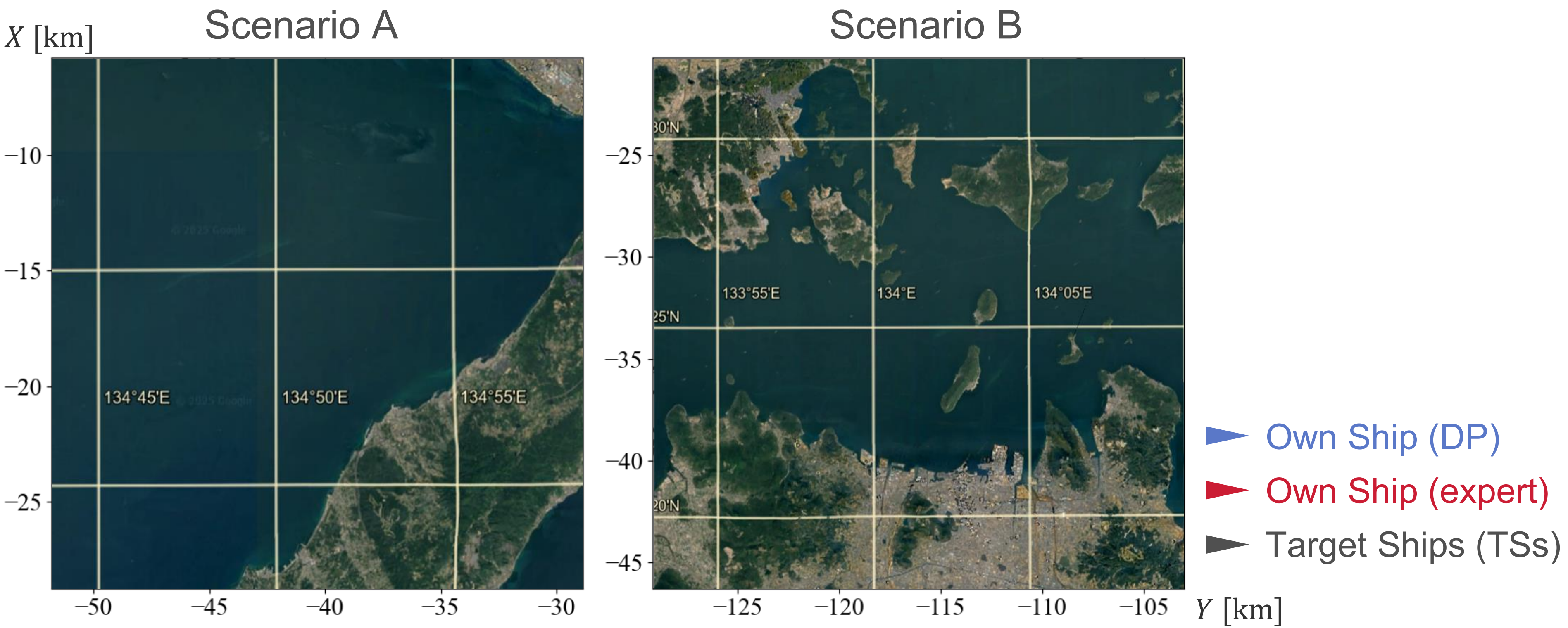
We collected 182 days of navigation data using the onboard acquisition system[14] installed on the training ship named Fukae-Maru at Kobe University.

The ship sailed in the Seto Inland Sea, one of the most topographically intricate and traffic-dense waterways in Japan.

Since DP behaves probabilistically, we tested the DP 10 times from the same initial condition.

[14] Wakabayashi et al. (2018). *Proceedings of the Twenty-eighth International Ocean and Polar Engineering Conference*, 922–928.

Collision Avoidance Trajectories (Realistic Scenarios)



The DP generally presented similar trajectories with the experts.

However, the routes had variance and some of them were dangerous.

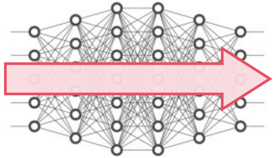
Challenge & Breakthrough of DP in Collision Avoidance

✗ In a safety-critical task like collision avoidance, the DP's probabilistic behavior, which could generate dangerous routes, is not acceptable.

✗ DP is not transparent because it does not formulate a policy.

DP

Observation $\boldsymbol{O}_t$ + Noise ε



Action a_t

← Is this action trustworthy? Rare action?

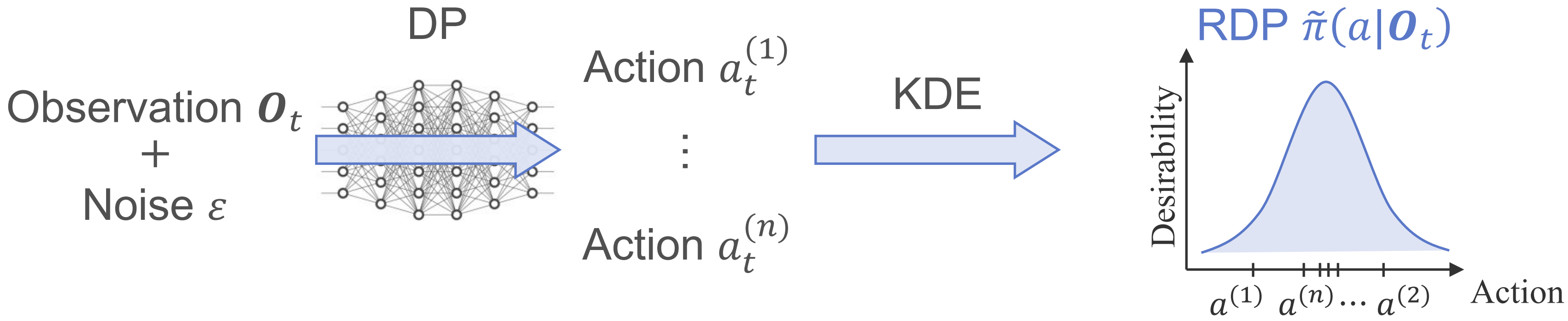
Challenge & Breakthrough of DP in Collision Avoidance

✗ In a safety-critical task like collision avoidance, the DP's probabilistic behavior, which could generate dangerous routes, is not acceptable.

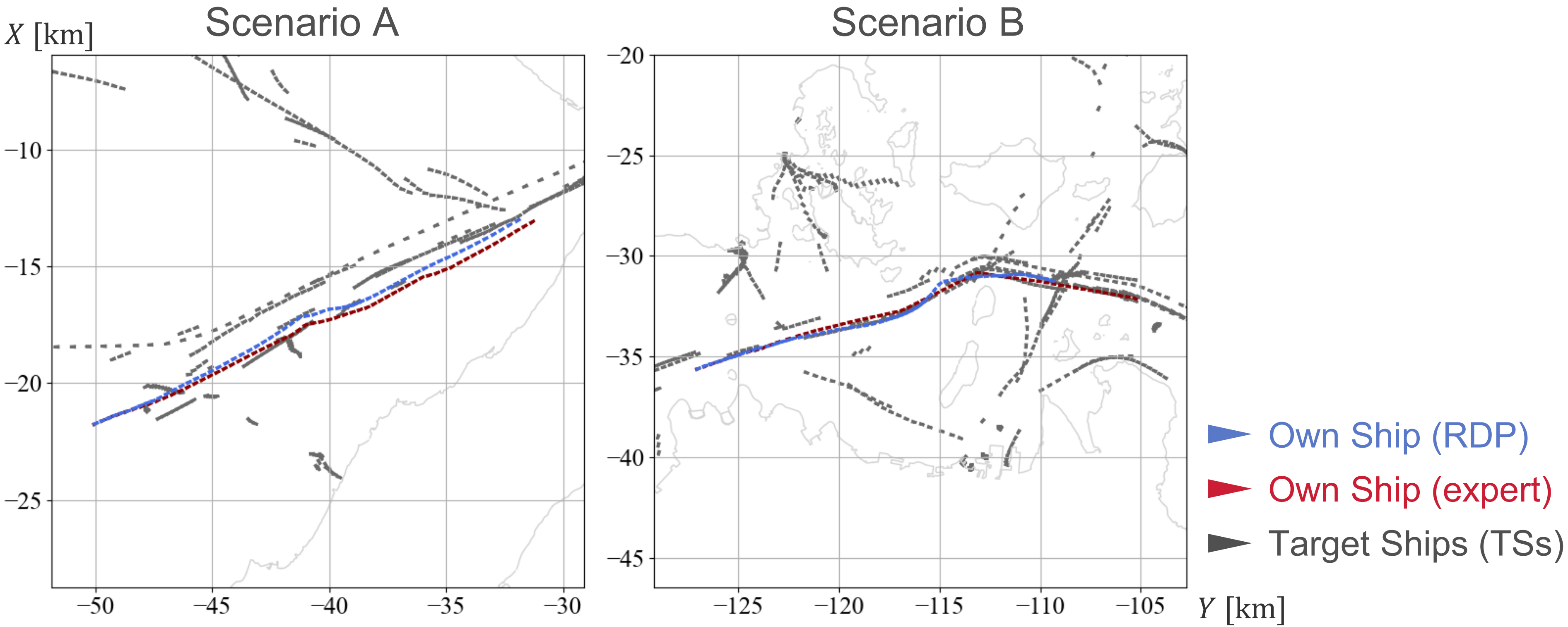
✗ DP is not transparent because it does not formulate a policy.

→ Representative Diffusion Policy (RDP): we made DP to sample many actions simultaneously $\left[a_t^{(1)}, \cdots, a_t^{(n)}\right]$, and obtained a policy-like function with kernel density estimation (KDE).

$$\tilde{\pi}(a|\boldsymbol{O}_t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{a - a_t^{(i)}}{h}\right) \quad \text{where } K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



Collision Avoidance Trajectories (Realistic Scenarios)



Even in the congested scenarios under geometric constraints, the RDP successfully generated experts-like routes.

Summary

Contributions

- ✓ To address the difficulty in formulating objective functions for ship collision avoidance, we have presented three data-driven approaches—IRL, IL, and RDP.
- ✓ The biggest advantage is that they do not require explicit specification of objective functions or constraints. This can be particularly beneficial when managing the remaining COLREGs rules and various local traffic regulations.

Future Perspectives

- ☐ Strict safety assurance: our approaches did not incorporate explicit objectives or constraints for collision risk.
- ☐ Real-world experiment: even in the realistic scenarios, TSs did not react to the OS's actions, and emergent risks remain unassessed.