上海科技大学
ShanghaiTech University

Towards Safer Navigation: Reward Shaping with Prior Topographic Knowledge in Point Goal Navigation

Jiajie Zhang Linkai Zu
Jintian Hu Tianyang Zhao

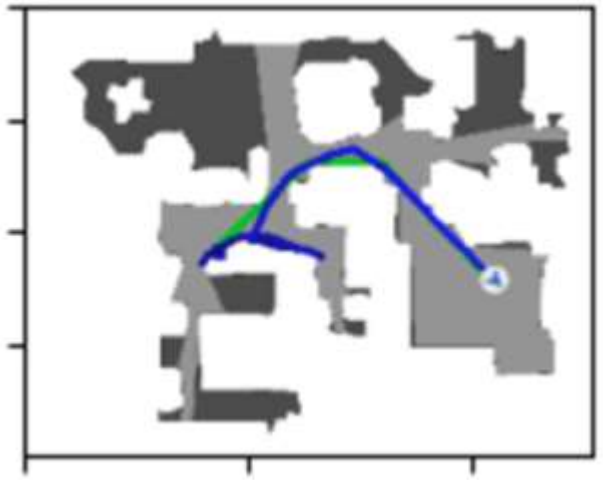
ShanghaiTech University

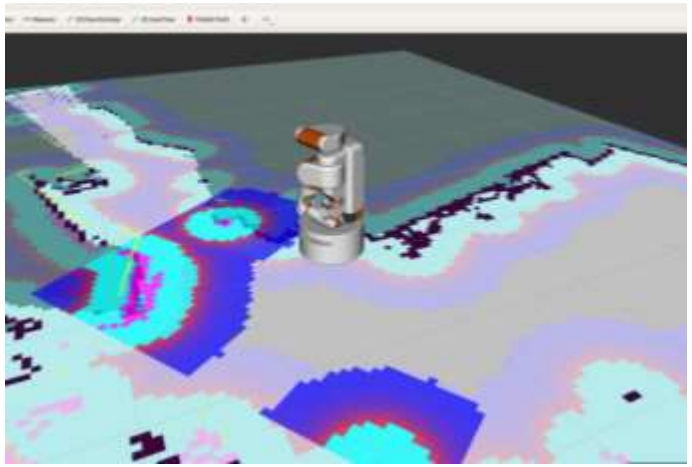
2024.1.06

Problem Statement & Motivation

- **Limitations of Traditional Navigation Methods:**
  - Traditional robot navigation systems rely on modular pipelines, including mapping, localization, path planning, and control.
  - Often cannot adapt to different complex real-world environments, unless tuning dozens of parameters by hand.
- **Shortcomings of Reinforcement Learning (RL) Agents:**
  - Typically do not use explicit map information, relying on representations learned from sensor data, which can lead to insufficient obstacle avoidance and poor path planning.
  - RL agents trained in virtual environments often encounter collision issues when deployed in the real world.

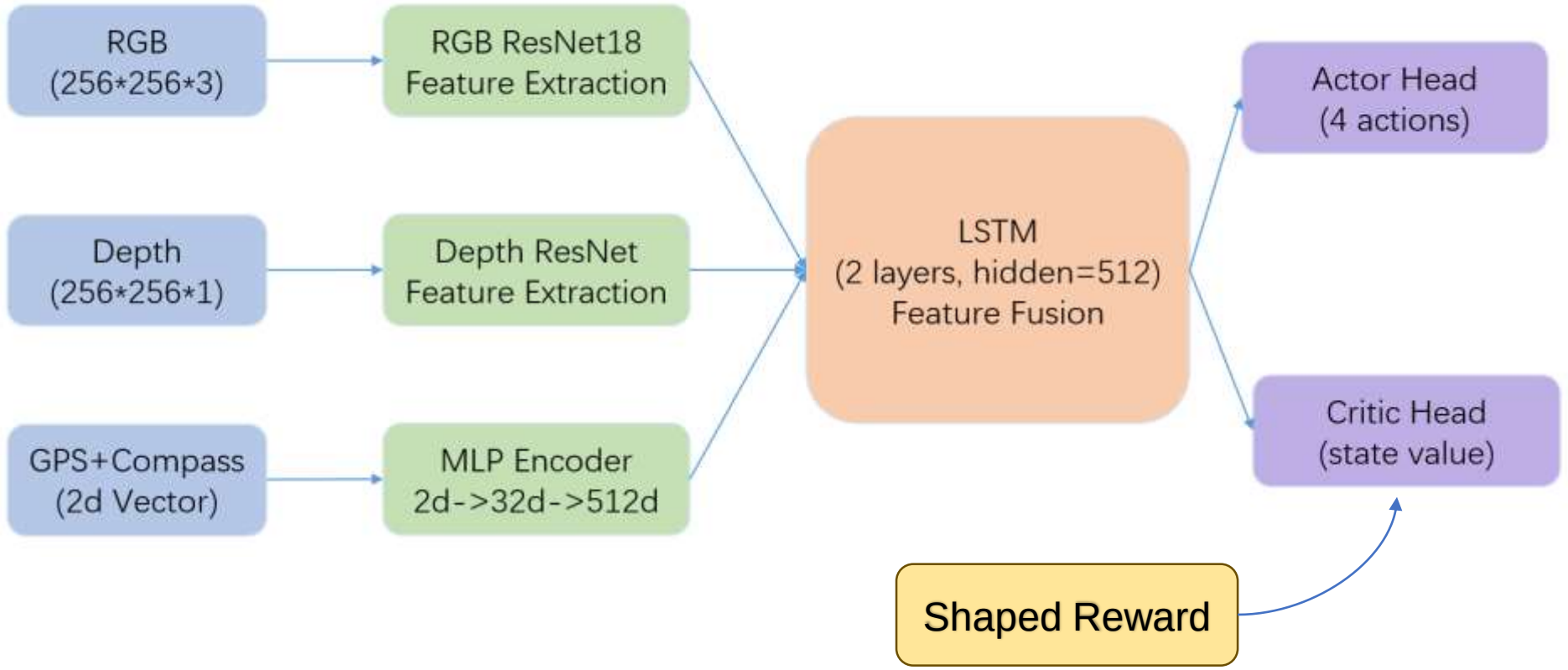
- **Core Problem:** How to improve the safety of navigation agents within the reinforcement learning framework, enabling them to operate reliably in real-world environments and avoid collisions?





Method

- **Integration with Habitat PPO Framework:**
  - Combines shaping rewards with baseline rewards for policy and value function updates using the PPO algorithm, ensuring the stability and optimization efficiency of the algorithm.



Method

- **Reward Shaping Technique:**
  - **Core Idea:** Use topographic information to guide the agent to learn safe navigation strategies by designing reward functions, rather than directly inputting map information into the agent.
- **Dual Reward Mechanism:**
  - **Exploration Reward (R_exploration):**
    - Encourages the agent to explore unknown areas and avoid stagnation
  - **Safety Reward (R_safety):**
    - Provides tiered rewards based on the agent's distance from obstacles, encouraging a safe distance and penalizing actions t

$$R_{\text{total}}(s_t) = r_t + \alpha R_{\text{exploration}}(s_t) + \beta R_{\text{safety}}(s_t)$$

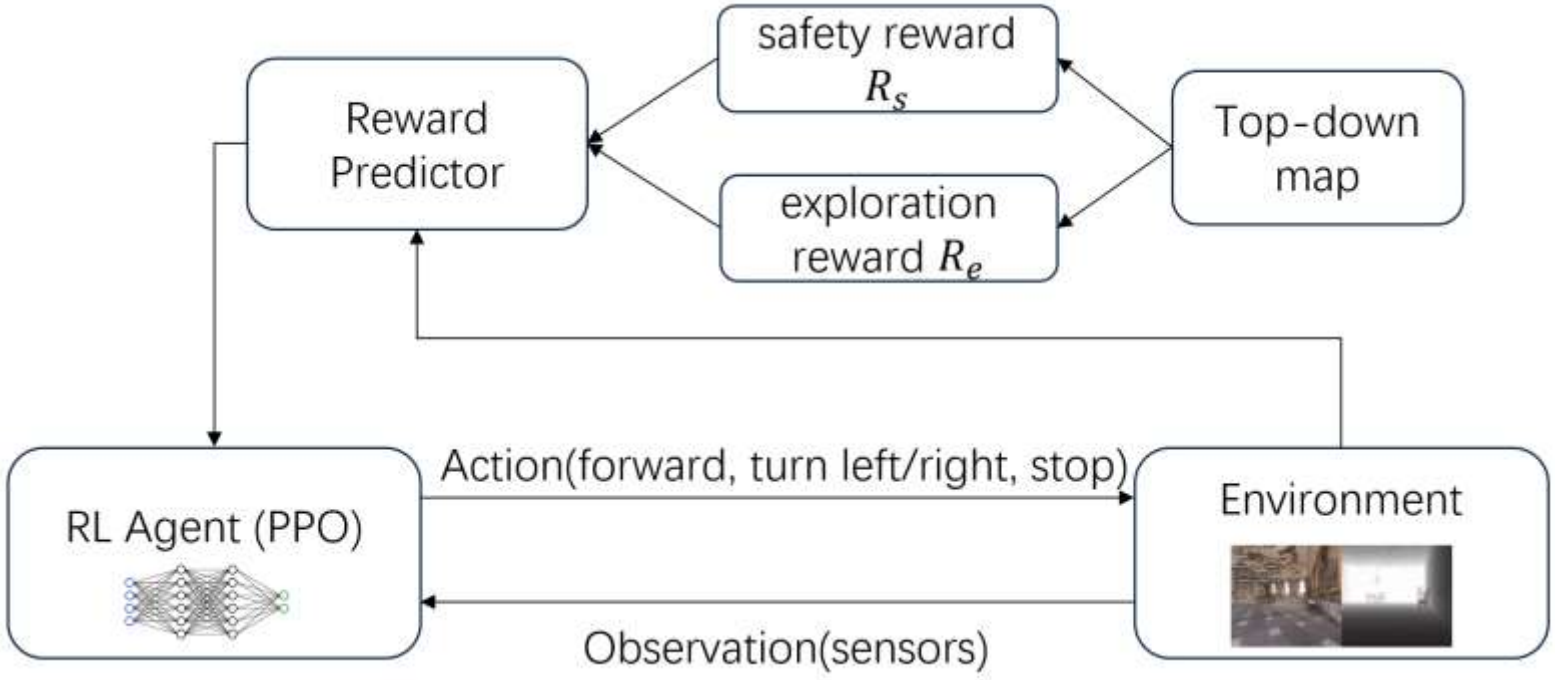
$$r_t = \begin{cases} s + d_{t-1} - d_t + \lambda & \text{if goal is reached} \\ d_{t-1} - d_t + \lambda & \text{otherwise} \end{cases}$$

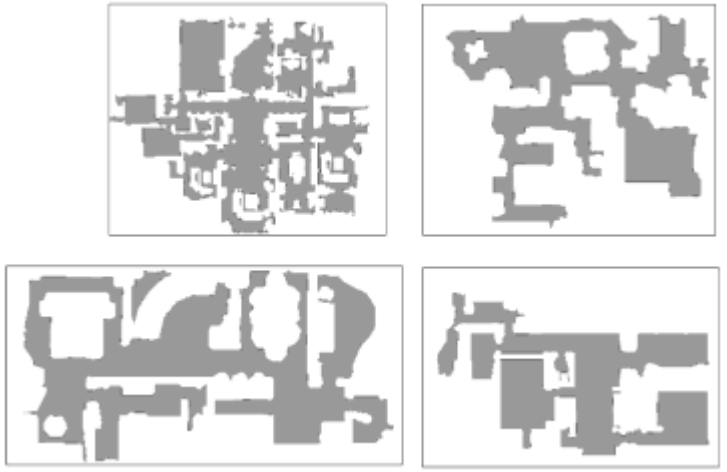
$$R_{\text{exploration}} = \lambda_1 \Delta d_{\text{goal}} + \lambda_2 \mathbb{I}(s_t \in \mathcal{V}) - \lambda_3 \mathbb{I}(s_t \in \mathcal{H})$$

$$R_{\text{safety}}(d) = \begin{cases} -0.2 & \text{if } d < 0.5 \text{ m} \\ 0.1 & \text{if } 0.5 \text{ m} \leq d < 2.0 \text{ m} \\ 0.0 & \text{if } d \geq 2.0 \text{ m} \end{cases}$$

Method

- **Core Method:** The reward shaping technique transforms map information into reward signals, guiding reinforcement learning agents to learn safe navigation strategies, while also using a new path safety metric for evaluation.





Experiment & Result

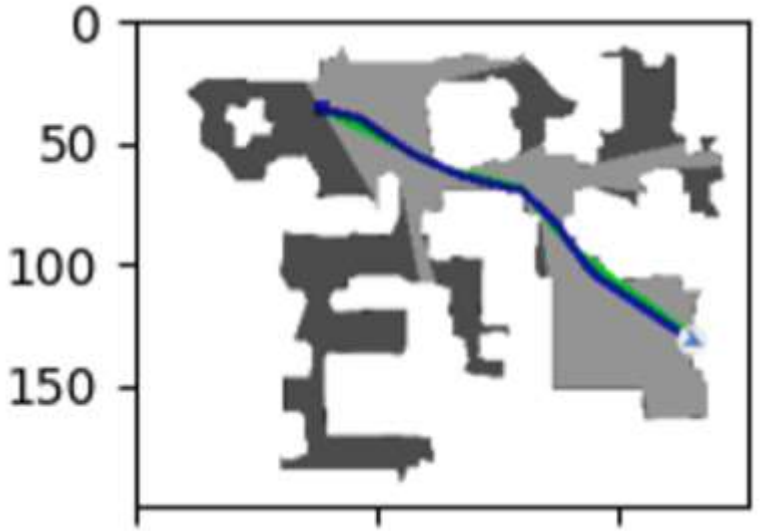
- **Task Definition:**
  - **PointGoal Navigation Task:** The agent needs to navigate to a specified target location without a map input.
  - **No Map Input:** Map information is only used for reward shaping.
- **Agent Configuration:**
  - **Sensors:** RGB and depth sensors (RGBD agents) or only RGB sensors (RGB agents).
  - **Action Space:** Includes four actions: turn left, turn right, move forward, and stop.
- **Experimental Setup:**
  - **Simulation Environment:** Uses the AI Habitat Simulator and Matterport3D dataset.
  - **Training Process:** Training is done in parallel using 4 multiple threads, for a total of **5 million** steps. Training time is approximately **10** GPU hours for RGBD agents and **9.5** GPU hours for RGB agents.

Experiment & Result

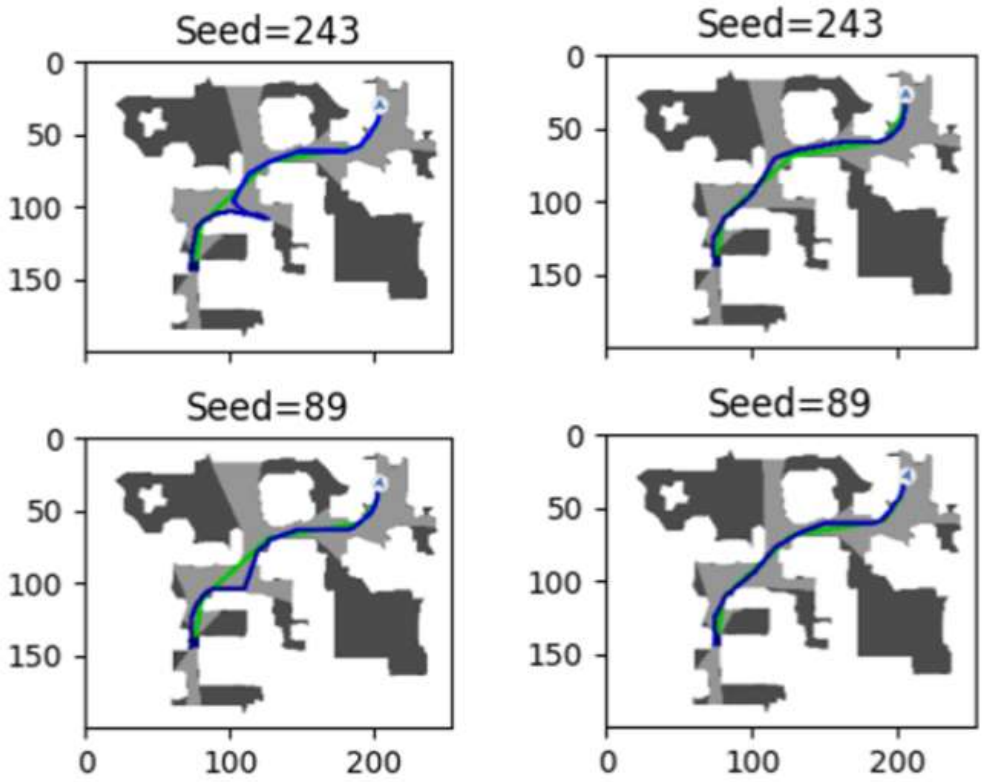
- **Evaluation Metrics:**
  - **Success Rate:** Whether the agent reaches the target within a specified distance.
  - **SPL (Success weighted by Path Length):** Considers both success rate and path efficiency.
  - **Path Safety:** Average distance between the agent's trajectory and obstacles.

$$\text{SPL} = S \cdot \frac{l}{\max(p, l)}$$

$$\texttt{path\_safety}(\tau) = \frac{1}{T+1} \sum_{t=0}^{T} D(u_t, v_t).$$



Experiment & Result



Agent	Success Rate	SPL
gibson_rgbd	0.230	0.262
gibson_rgbd_shaped	0.236	0.279
mp3d_rgbd	0.343	0.341
mp3d_rgbd_shaped	0.353	0.339

Table 1. Average Success Rate and SPL for different Agents (averaged over multiple seeds).

Experiment & Result

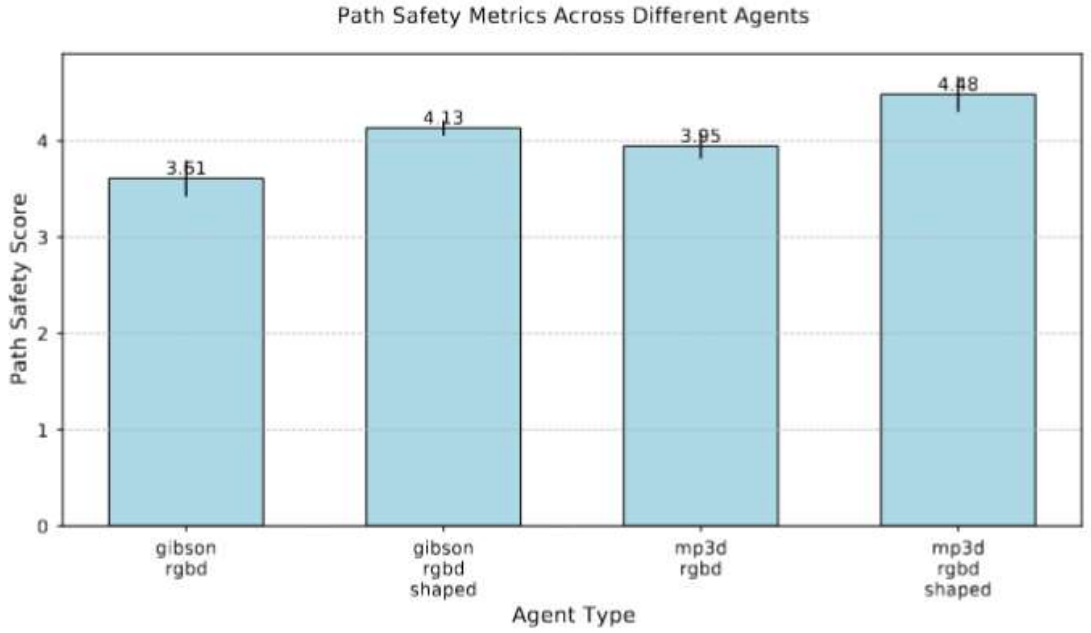


Figure 4. Comparison of path safety scores between RGBD agents with and without reward shaping. The figure shows that, with our reward shaping, the paths taken by the RGBD agents maintain a larger margin from obstacles. We trained the Gibson agent and MP3D agent, then tested them on the MP3D dataset. The results demonstrate that both RGBD agents with reward shaping perform better in an unseen environment

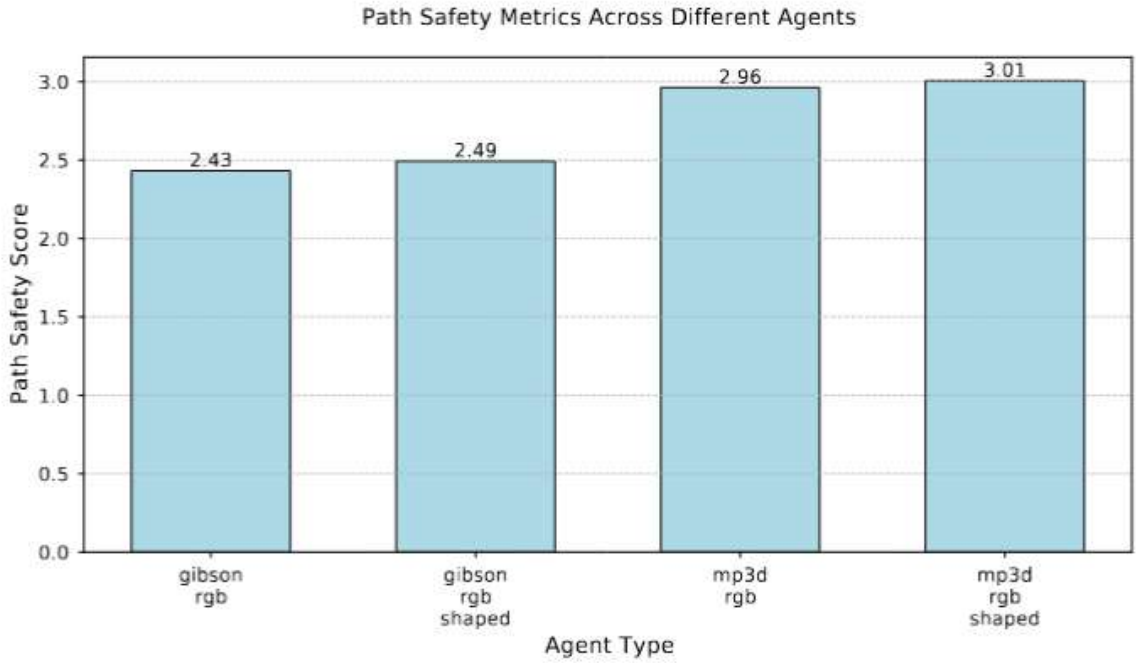


Figure 5. Comparison of path safety scores between RGB agents with and without reward shaping. A limitation we discovered is that reward shaping did not significantly impact the path safety performance of RGB-based agents during testing, particularly when the agents encountered unseen environments.

Discussion

- **Improvements in RGBD Agents:**
  - **Learned Safe Strategies:** Even during tests without active reward shaping, RGBD agents maintained high path safety scores, indicating they had learned safe navigation strategies.
- **Limitations of RGB Agents:**
  - **Insignificant Safety Improvement:** The improvement in path safety for RGB agents using reward shaping was not significant, especially in unseen environments.
  - **Importance of Depth Information:** This shows that depth information is crucial for effective reward shaping, and RGB agents may require other mechanisms to improve safety.

- **Core Findings:**
  - Reward shaping can effectively improve the safety of RL navigation agents, especially when depth information is used.
  - Depth information is crucial for safe navigation; agents that only use RGB information have limitations in safe navigation.