Adaptive and Explainable Deployment of Navigation Skills via Hierarchical Deep Reinforcement Learning

Kyowoon Lee*, Seongun Kim*, and Jaesik Choi

Ulsan National Institute of Science and Technology, Ulsan, Korea
Korea Advanced Institute of Science and Technology, Daejeon, Korea

Background and Motivations

Background

- Existing methods focus on learning a single navigation policy with a fixed reward function which are difficult to be deployed in a wide range of real-world scenarios.

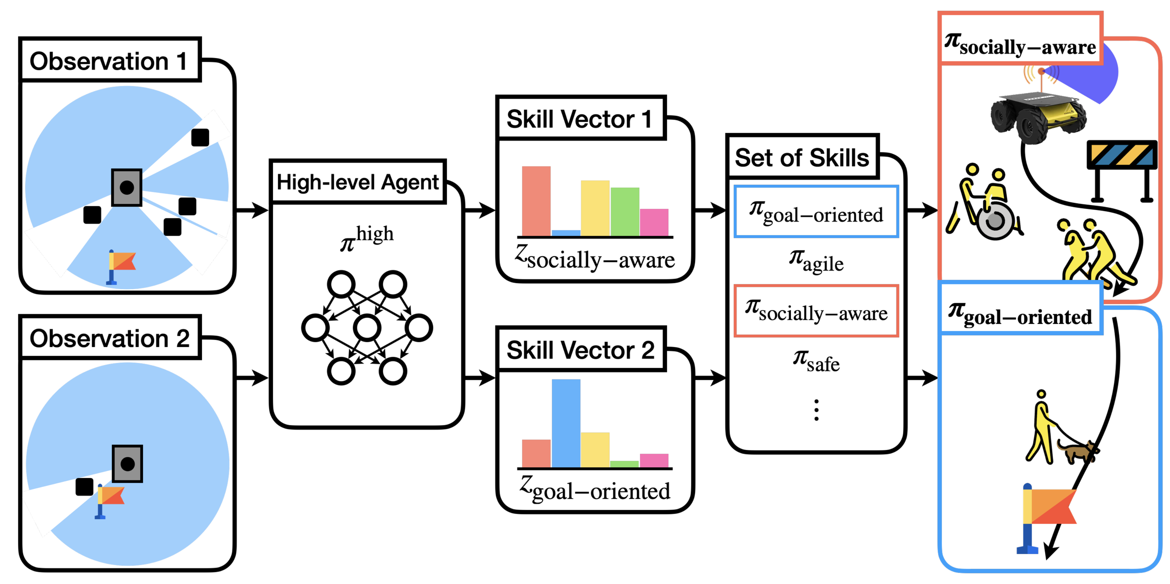
Motivation

- Fixed reward function are easy to get stuck in local optima which makes a wide range of complex real-world scenarios unsolvable.
- A navigation policy represented by a deep neural network often lacks transparency and could not provide explanations on decision making reasons.

Contributions

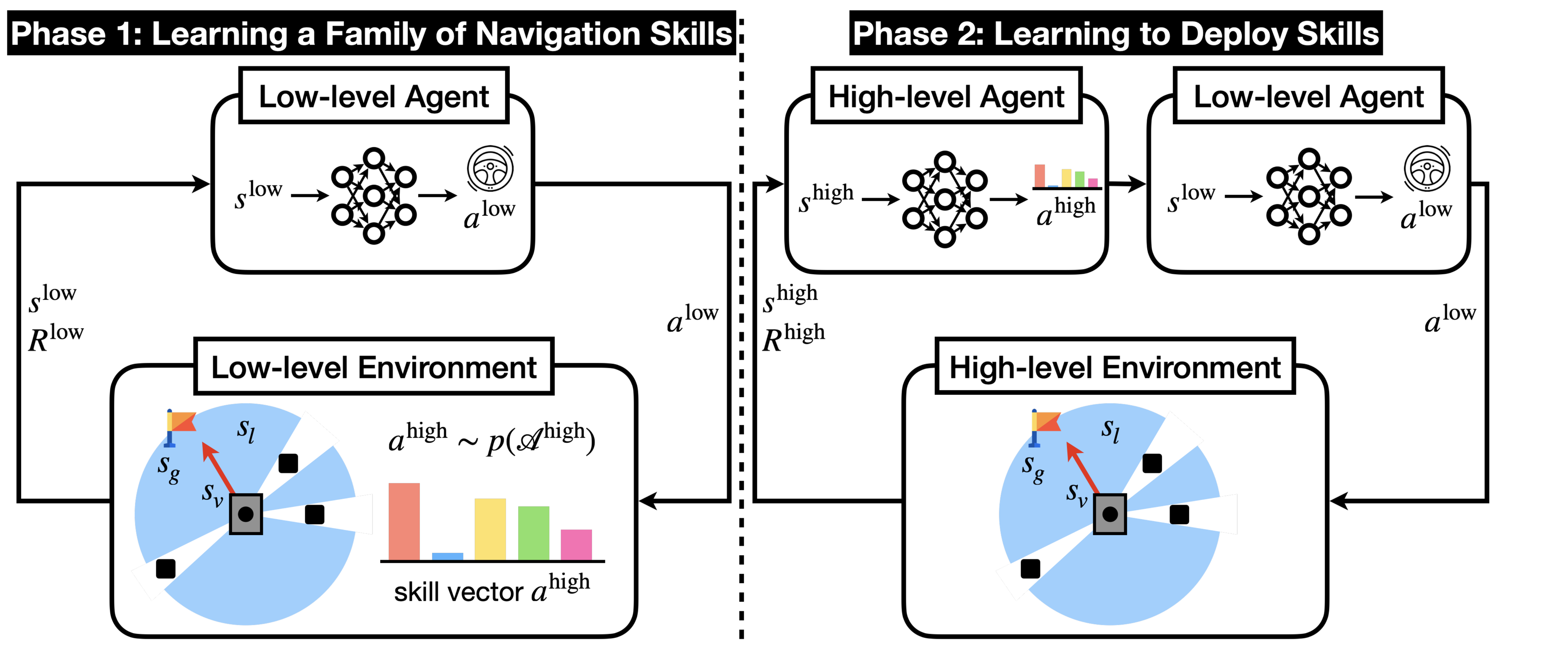
- We propose a hierarchical reinforcement learning approach which learns diverse navigation skills and deploys them.
- Experiments results show effectiveness and explainability of our approach on various scenarios.

Schematic of Hierarchical Framework of Navigation Skills



- A high-level policy invokes low-level navigation skills from a raw sensory observation.
- A low-level policy is adopted from a continuous skill vector and drives a robot.

Problem Formulation



- We decompose a problem of learning navigation skills into a hierarchy of two sub-problems as a goal-conditioned Markov Decision Process, and learn a family of low-level navigation policies and a high-level policy which adaptively deploy the learned navigation skills.

Learning a Family of Navigation Skills

We simultaneously learn a family of low-level navigation policies that exhibit different behaviors with a wide range of reward functions.

Reward Function

- The reward function is parameterized by the skill vector a^{high} which is associated with a corresponding reward function which comprises six distinct components: success, collision avoidance, progress, driving, turning, and safety.
- Skill vector a^{high} induces a specific behavior by weighting a number of reward terms.

$$R_t^{\text{low}}(s_t^{\text{low}}, a_t^{\text{low}}, a_t^{\text{high}}) = r_{\text{success}} + a^{\text{high}}[r_{\text{collision}}\ r_{\text{progress}}\ r_v\ r_w\ r_{\text{safety}}]^\top.$$

Training procedure

- At the beginning of the episode, we sample a skill vector from predefined distribution and fix it during the rollout to train policy $\pi_{a_t^{\text{high}}} : s_t^{\text{low}} \to a_t^{\text{low}}$, which we call a skill.

Learning to Deploy Skills

We learn a high-level policy to adaptively deploy the learned navigation skills.

Reward Function

- The goal is to minimize the time to reach the desired goal.
- We use the sparse reward function where the agent gets the reward of 0 when the goal is achieved, and -1 otherwise.

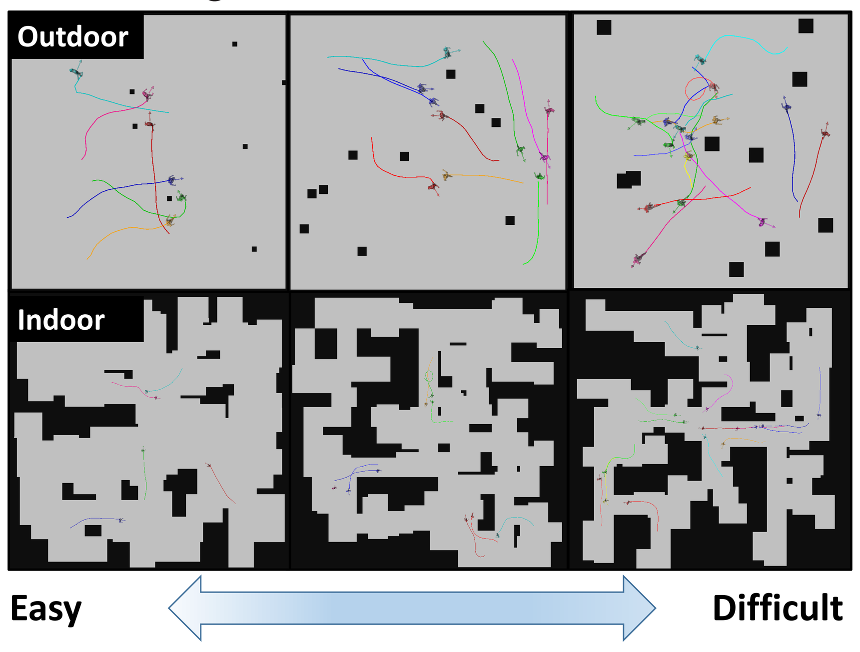
$$R_t^{\text{high}}(s_t^{\text{high}}, a_t^{\text{high}}, a_t^{\text{low}}) = \begin{cases} 0 & \text{if reach the goal} \\ -1 & \text{otherwise.} \end{cases}$$

Training procedure

- During a rollout, the high-level policy predicts the skill vector which decides the behavior characteristic.
- Skill vector is observed as an additional input to the low-level policy which outputs the command velocity of a robot.
- We use the hindsight experience replay (HER) technique to handle the sparse reward challenge.

Experimental Results

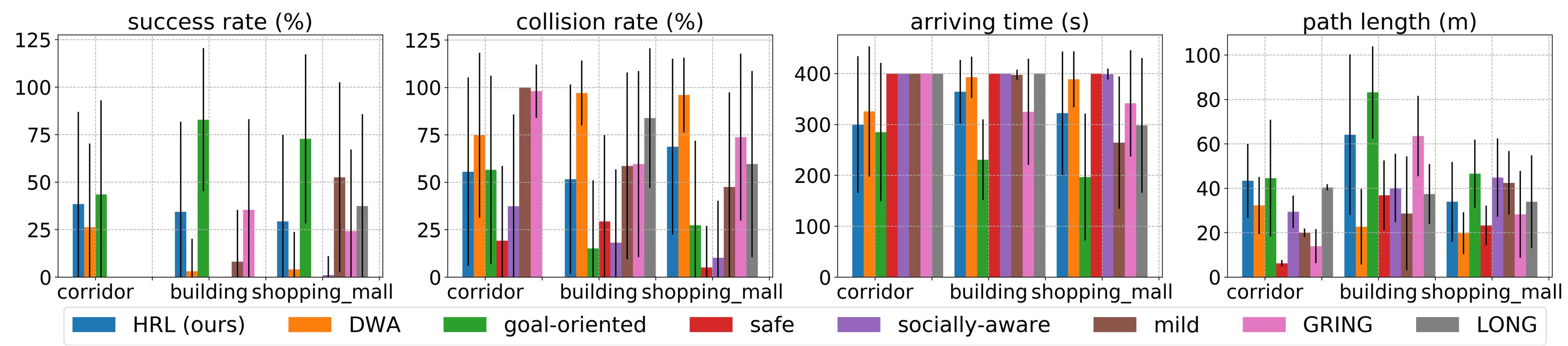
- Training environments.



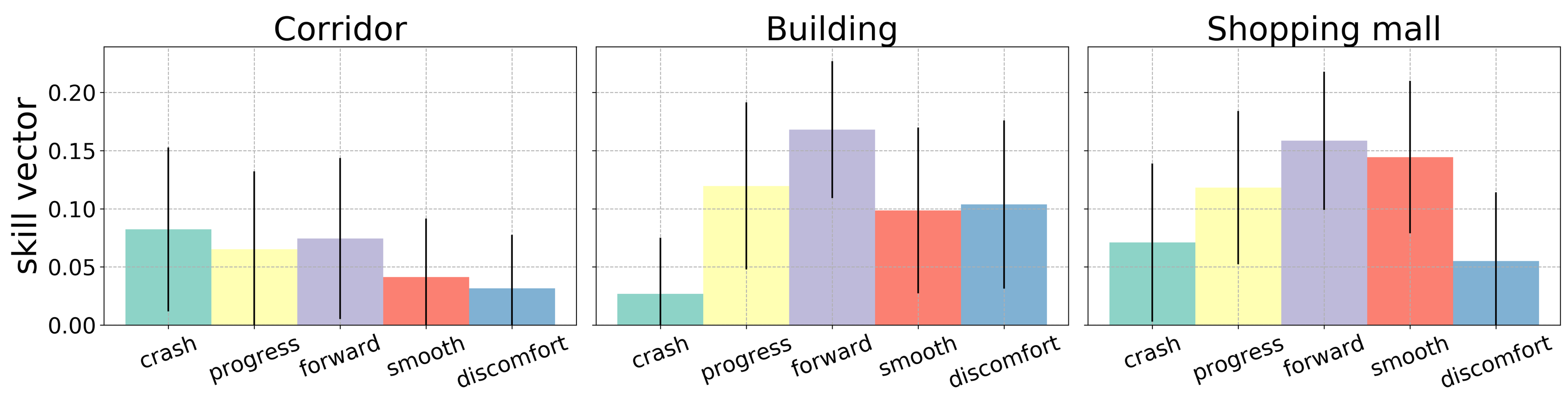
- Unseen evaluation environments.

(a) Corridor 35 m x 39 m
(b) Building 50 m x 110 m
(c) Shopping mall 50 m x 110 m

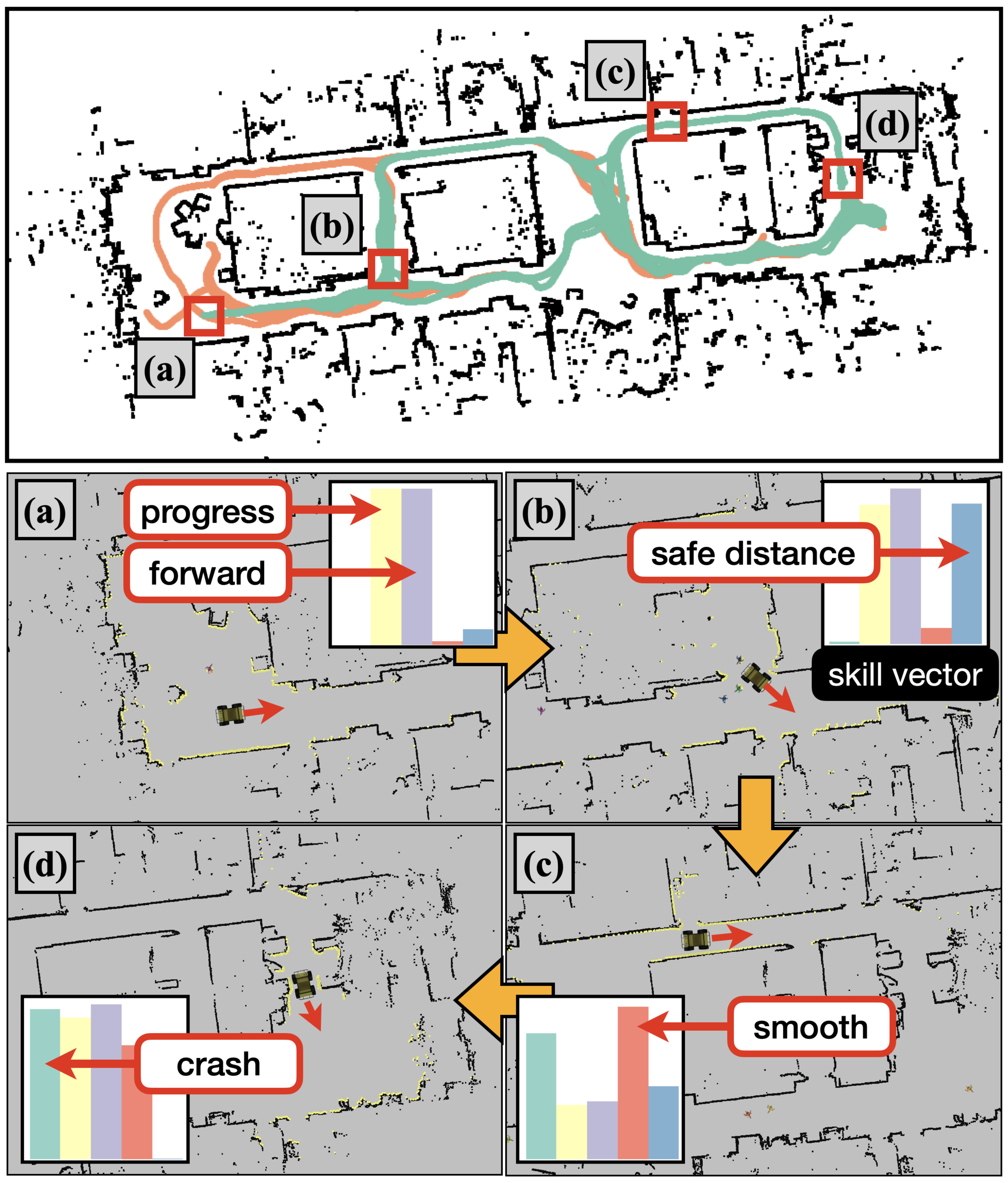
- Our approach shows comparative performance to other baseline and can reduce the effort to design hand-engineered reward functions.



- High-level policy adaptively deploys the most suitable navigation skills on unseen environments.



- Our approach presents explainability by providing semantics of a behavior of an autonomous agent.



Acknowledgements

- This work was supported by the Industry Core Technology Development Project, 20005062, Development of Artificial Intelligence Robot Autonomous Navigation Technology for Agile Movement in Crowded Space, funded by the Ministry of Trade, Industry & Energy (MOTIE, Republic of Korea) and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-00984, Development of Artificial Intelligence Technology for Personalized Plug-and-Play Explanation and Verification of Explanation, No.2019-0-00075, Artificial Intelligence Graduate School Program (KAIST)).

*Equal contribution

Email: leekwoon@unist.ac.kr, {seongun,jaesik.choi}@kaist.ac.kr