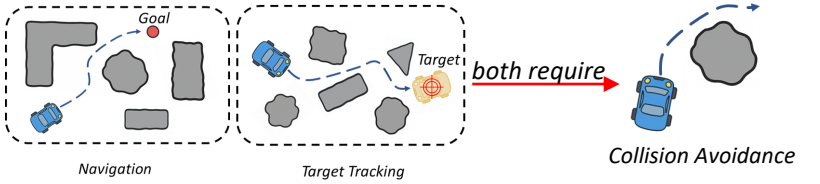
Transfer Your Safety: Learning Transferable Model-Free Safety Filters from a Single Policy to Enhance Safety Across Diverse Tasks

Junjun Xie, Siru Li*, Shuhao Zhao, Xiaochen Xie, and Liang Hu*
** Equal contribution*
Harbin Institute of Technology, Shenzhen

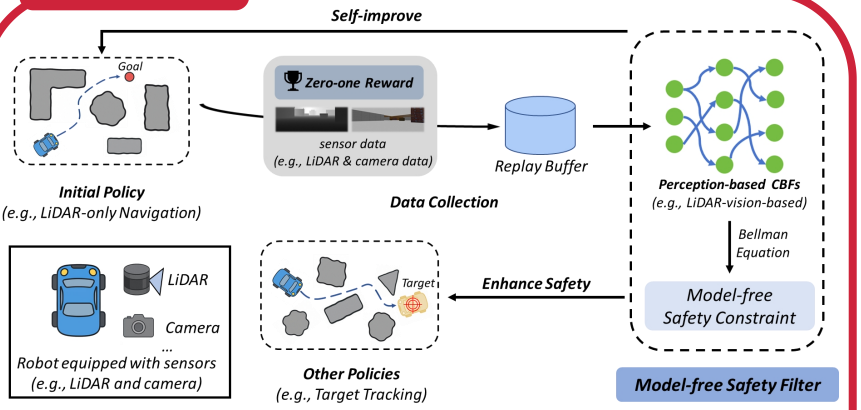
Motivation

- ***Different tasks share the same safety requirements in similar environments***



*Given a task-specific policy, can we **learn the knowledge of safety** and **transfer it to other tasks**?*

Method



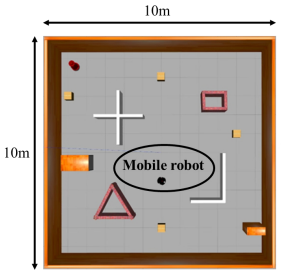
The overall pipeline is as follows:

- Collect data from an initial policy and learn a perception-based Control Barrier Function (CBF) using predefined binary rewards.
- Derive *model-free* action CBF constraints from the Bellman equation without predicting future observations.
- Use the constraints as a safety filter to correct policies for other tasks in similar environments.

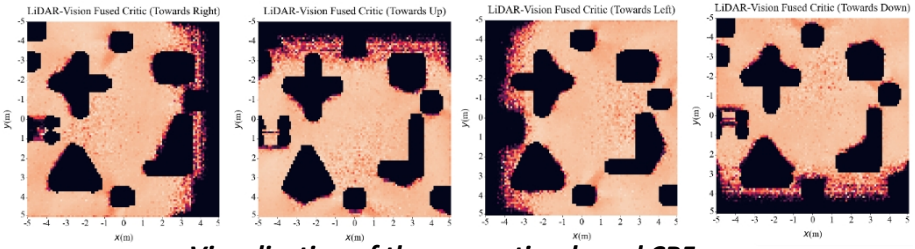
Experiments

Perception-based CBF Learning

We pretrained a **LiDAR-based navigation policy** and collected LiDAR data together with camera data unused by the original policy from randomly initialized environments, then trained a **CBF based on both LiDAR and camera** observations.



Test Scenario



Visualization of the perception-based CBF
(robot facing right, up, left, and down, respectively).

Cross-task Safety Enhancement & Self-improvement

We derive a model-free safety filter based on the LiDAR-camera-based CBF and apply it to **two obstacle-unaware task policies (cruising and tracking),** as well as the **pretrained navigation policy** used for data collection. Results show significant **safety improvements across tasks**, while also **enhancing the safety of the original navigation policy.**



Cruising



Tracking



Navigation

Cruising

Safety filters	Inputs	Min.	Max.	Average
W/o safety filters	–	35	202	87
Baseline	Single-frame	403	924	539
	Multi-frame	509	1865	987
Ours	Single-frame	462	1007	648
	Multi-frame	**524**	**2089**	**1116**

Tracking

Min. distance	Max. distance	Average	Number of collisions
0.93 m (at step 864)	9.62 m (at step 0)	2.70 m	0

Navigation

	Initial	Filtered
Success rate	13/20(65%)	17/20(85%)

ICRA IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION 2026 VIENNA
哈爾濱工業大學（深圳） HARBIN INSTITUTE OF TECHNOLOGY, SHENZHEN
CLASS-LAB Control and Learning for Autonomous SystemS Lab