SSY226: Design Project in Systems, Control and Mechatronics

Real-Time Obstacle Avoidance for Mobile Robots using DRL and MPC

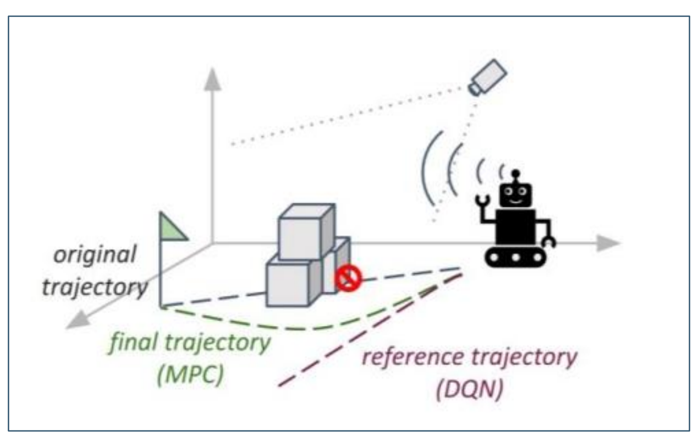
Shao-Hsuan Huang, Valdemar Samuelsson, Mathanesh Vellingiri Ramasamy, Rishikesh Vishnu Sivakumar , Yu Kang

CHALMERS UNIVERSITY OF TECHNOLOGY

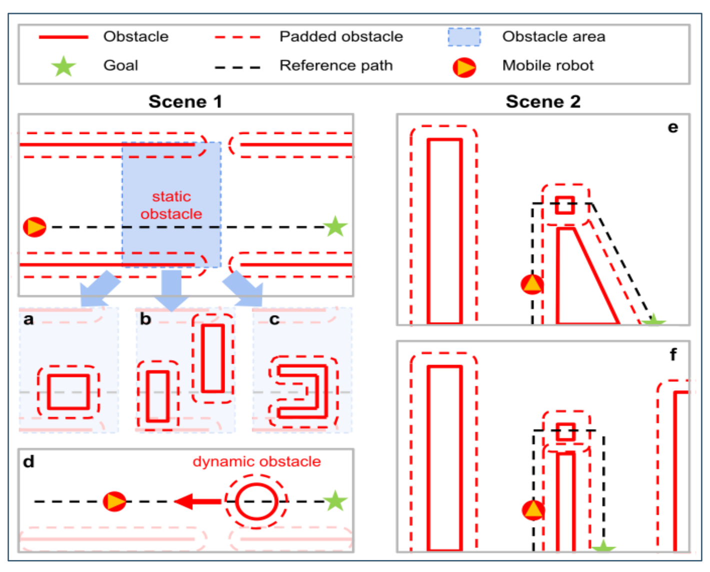
Background

AMR may encounter unforeseeable situations during operation. In this complex scenario, we have integrated multiple frameworks to achieve a better balance between security, efficiency, and accuracy. Some ideas and evaluation criteria considered are as follows:

- Q-function integration: incorporating the Q-function of DRL into the optimization process of MPC, replacing the switching mechanism.
- Action space optimization: Optimize the action space of DRL to align with the control objectives of MPC, reducing computational overhead.
- Stability and feasibility check: Ensure that the control measures generated by DRL comply with MPC's stability and feasibility standards.



Robot and its Environment



Scenes for single-agent evaluation

Methodologies

Q-function as final cost for Model Predictive Control:

Discarding the existing final cost of MPC and replacing it with the learned Q-function:

$$J_{terminal} = q_w \cdot \left(Q_{max} - Q_{ip}(s_H, a_H, t) \right)^2$$

- Generating a look up table to create approximations of the Q-function.
- Building a point grid that spans the operational area to discretize the state action space.
- Calculating the Q-value of each state action combinations using the DRL model.

Including MPC within the learning process of DRL:

- Calculating the original action prediction of the DRL model as a reference path and sending it as the input to the MPC model.
- The above results will be used to assign rewards for the DRL training process (solver time, MPC model cost, and model penalty).
- However, the training process becomes highly complex.

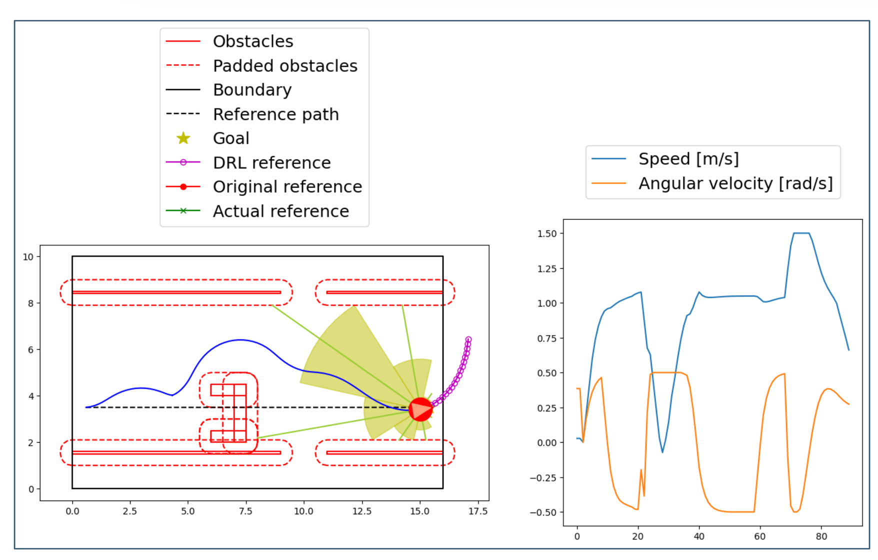
Generating reference path using DRL:

- A new hybrid model DRL reference path is generated by multiple DRL actions, with the same horizon length as that of the MPC.
- This eliminates the need to switch between two different references.
- Thus, the DRL agents take steps in simulated environments without moving obstacles.

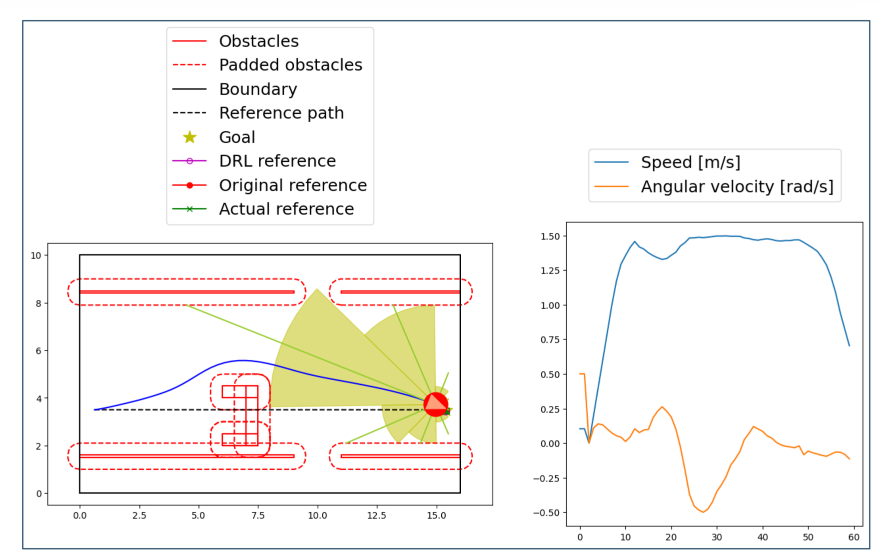
Evaluation and Results

Scene	Obstacle Type (Scene ID)	Method	Computation Time (ms/step)			Deviation (m)		Action smoothness		Other	
			mean	max	std	mean	max	speed	angular speed	finish time step	success rate (%)
Scene 1	(a) Rectangular obstacle (1-1-2)	MPC	245.85	1253.47	83054.14	0.39	1.49	0.04	0.06	69	94
		DDPG-L	0.33	0.72	0.01	0.84	1.75	0.09	0.2	58	100
		REF-DDPG-L	65.35	112.24	237.11	0.81	1.65	0.02	0.04	56	100
		HYB-DDPG (OG)	29.97	224.2	1527.18	0.84	2.29	0.02	0.06	80	100
		HYB-DDPG (M1)	1611.77	5081.67	1281916.35	0.87	1.7	0.07	0.18	113	94
		HYB-DDPG (M2)	112.95	897.04	26705.77	1	2.16	0.01	0.05	81	92
	(b) Two obstacles (1-2-1)	MPC	356.59	1259.07	83640.48	0.43	1.27	0.05	0.04	72	100
		DDPG-L	0.3	0.61	0.01	0.92	1.84	0.11	0.12	82	4
		REF-DDPG-L	81.76	181.8	606.51	0.71	1.64	0.04	0.04	80	100
		HYB-DDPG (OG)	38.78	150.82	866.08	0.82	2.14	0.04	0.05	98	100
		HYB-DDPG (M1)	1936.81	5026.06	1633470.2	0.74	1.6	0.12	0.27	165	2
		HYB-DDPG (M2)	174.27	633.61	15187.8	0.83	2.14	0.03	0.05	97	100
	(c) U-shape obstacle (1-3-2)	MPC	583.39	3989.42	1113825.28	0.14	0.3	0.15	0.2	-1	0
		DDPG-L	0.31	0.6	0.01	1.08	2.38	0.11	0.18	61	98
		REF-DDPG-L	72.52	109.06	234.7	0.95	1.99	0.02	0.04	58	100
		HYB-DDPG (OG)	73.47	842.56	25109.03	0.28	0.82	0.03	0.04	-1	0
		HYB-DDPG (M1)	2085.14	5031.06	1293683.37	0.77	1.71	0.15	0.4	98	64
		HYB-DDPG (M2)	86.9	1152.91	41671.47	0.3	0.85	0.03	0.04	87	2
Scene 2	(e) Sharp turn with an obstacle (2-1-2)	MPC	50.95	340.89	4043.25	0.12	0.56	0.04	0.07	147	8
		DDPG-L	0.31	0.76	0.01	0.92	2.03	0.09	0.18	120	100
		REF-DDPG-L	63.59	130.12	181.09	0.9	1.99	0.02	0.03	121	100
		HYB-DDPG (OG)	20.29	131.52	483.99	0.38	1.56	0.01	0.02	146	100
		HYB-DDPG (M1)	1207.57	5007.26	967383.62	0.55	3.75	0.01	0.01	-1	0
		HYB-DDPG (M2)	47.42	309.39	2000.97	1.01	2.54	0.01	0.05	186	66
	(f) U-turn with an obstacle (2-1-3)	MPC	135.55	295.76	4863.15	0.08	0.27	0.02	0.02	-1	0
		DDPG-L	0.32	0.83	0.01	0.73	1.48	0.12	0.18	122	92
		REF-DDPG-L	64.11	130.65	183.18	0.75	1.51	0.01	0.02	121	100
		HYB-DDPG (OG)	28.92	373.5	2030.02	0.41	1.45	0.02	0.02	145	100
		HYB-DDPG (M1)	1293.66	5043.25	1716473.7	0.48	3.31	0.04	0.03	-1	0
		HYB-DDPG (M2)	65.51	1189.7	11791.2	0.75	1.51	0.02	0.05	164	42

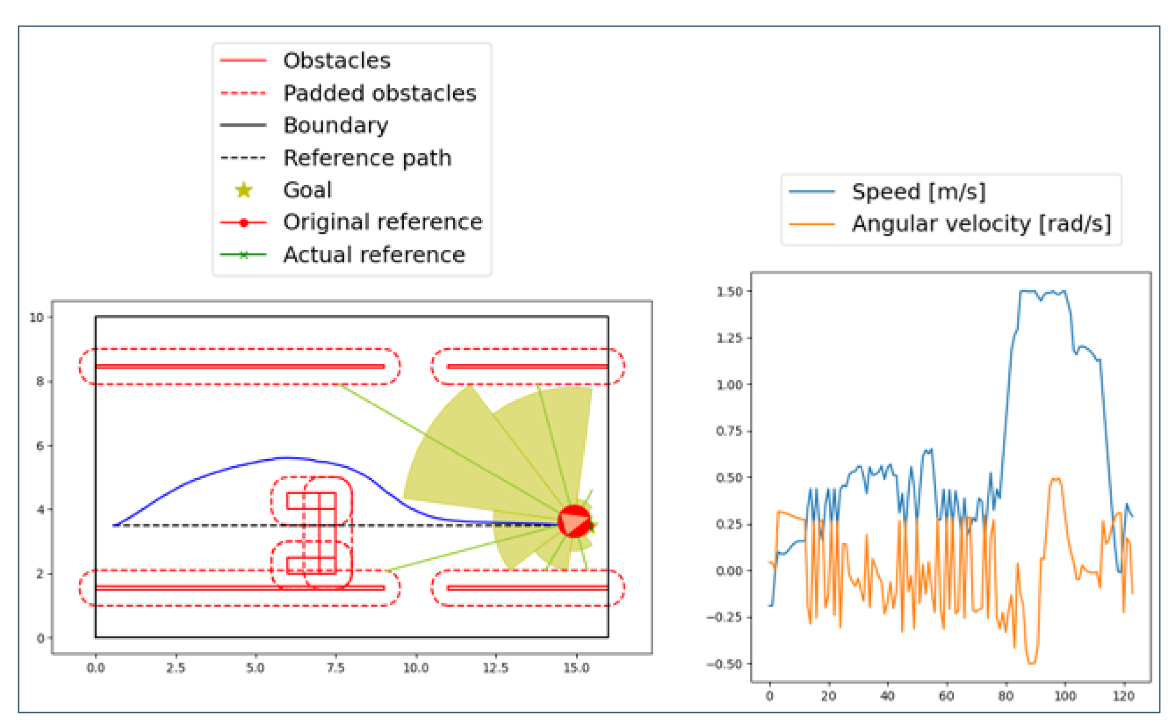
Evaluation results for different scenarios and methods



HYB-DDPG(OG)



REF-DDPG-L



HYB-DDPG-L(M1)

Conclusion and Future Work

In this project, different methods were attempted to improve the existing model.

- The use of Q-learning as the final cost of MPC has shown promising results in some test cases, but there are issues with long computation time and some instability.
- Adding MPC to the training loop did not improve the hybrid model. This may be because MPC adds another layer of separation between DRL agent decisions and rewards.
- Using multiple DRL actions as reference paths for MPC shows promise as it improves completion time, success rate, and maximum computation time. One drawback is that it has a longer average value, which may not be possible in real-world applications.

References

➢ Collision-Free Trajectory Planning of Mobile Robots by Integrating Deep Reinforcement Learning and Model Predictive Control (2023) (CASE). (Authors: Ze Zhang, Yao Cai, Kristian Ceder, Arvid Enliden, Ossian Eriksson, Soleil Kylander, Rajath Sridhara and Knut Åkesson).

➢ Bird's-Eye-View Trajectory Planning of Multiple Robots usingContinuous Deep Reinforcement Learning and Model Predictive Control (2024) (IROS). (Authors: Kristian Ceder, Ze Zhang, Adam Burman, Ilya Kuangaliyev, Krister Mattsson, Gabriel Nyman, Arvid Petersén, Lukas Wisell and Knut Åkesson).

➢ Temporal Difference Learning for Model Predictive Control. (Authors: Nicklas Hansen, Xiaolong Wang, Hao Su).

MPSYS

SYSTEMS, CONTROL AND MECHATRONICS, MSC PROGR