arXiv:1911.03074v2 [cs.RO] 5 Mar 2020

Mapless Navigation among Dynamics with Social-safety-awareness: a reinforcement learning approach from 2D laser scans

Jun Jin*,†, Nhat M. Nguyen†, Nazmus Sakib*,†, Daniel Graves†, Hengshuai Yao† and Martin Jagersand*

***Abstract*— We consider the problem of mapless collision-avoidance navigation where humans are present using 2D laser scans. Our proposed method uses ego-safety to measure collision from the robot's perspective and social-safety to measure the impact of robot's actions on surrounding pedestrians. Specifically, the social-safety part predicts the intrusion impact of the robot's action into the interaction area with surrounding humans. We train the policy using reinforcement learning on a simple simulator and directly evaluate the learned policy in Gazebo and real robot tests. Experiments show the learned policy smoothly transferred to different scenarios without any fine tuning. We observe that our method demonstrates time-efficient path planning behavior with high success rate in the mapless navigation task. Furthermore, we test our method in a navigation task among dynamic crowds, considering both low and high volume traffic. Our learned policy demonstrates cooperative behavior that actively drives our robot into traffic flows while showing respect to nearby pedestrians. Evaluation videos are at https://sites.google.com/view/ssw-batman**

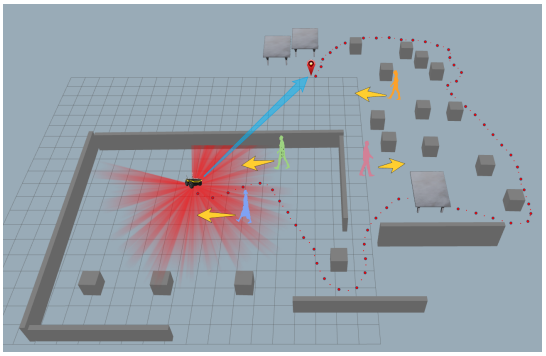


Fig. 1: Given a target location, humans can navigate locally without knowing a detailed map of surroundings. We typically make decisions from sensory inputs to avoid obstacles and walk-in pedestrians while showing respect to other people. Our decisions come from both our egocentric and allocentric spatial cognition. Can we train a robot with a similar behavior?

I. Introduction

The problem of autonomously navigating a robot in a map-unknown (mapless) environment while avoiding both static and dynamic obstacles, is important but challenging in applications like delivery robots, indoor service robots, etc. The *path planning and static obstacle avoidance* parts in this problem are often formatted as *mapless navigation* [1, 2] where a robot is driven by observed sensory data from the unknown environment, assuming the relative pose from robot to target is given by a third party localization module (e.g., GPS for outdoor cases and UWB / Wifi / Zigbee for indoors.). The *dynamic obstacle avoidance* part in this problem is more complicated since it requires future prediction on unknown surrounding dynamics like moving pedestrians, vehicles or other robots [3, 4]. As complexity of surrounding dynamics increases, the prediction may result in occupying a large portion of free space, successively causing no solution in path planning, namely the freezing robot problem [5]. Moreover when the moving obstacles are human, not only collision avoidance, but also *human-awareness* [6, 7] should be considered. While recent approaches using multi-modal sensing [8] or high-end object/pedestrian detection pipelines [4, 9, 10] have been proposed to tackle part of the problem, we are still curious to ask: is it possible to solve for all parts purely based on 2D laser scans?

Recent works on deep reinforcement learning have proven the capability of using 2D laser scans in mapless navigation[2, 11] and multi-agent / dense crowd[12, 13] collision avoidance tasks. Such methods share a similar reward structure by adding a reaching target reward and a collision penalty, but differ in the training process. For example, in a mapless navigation task, policy training is done by using a fixed number of floor plans[2, 11]; In multi-agent / dense crowd[12, 13] collision avoidance task, the training involves interactions with randomized autonomous agents. It is observed that the collision penalty combined with sufficient exploration in training, makes the learned policy show cooperative behaviors that solve the freezing robot problem. However, this collision penalty only measures near impact of the robot from surrounding agents and obstacles. When humans are present nearby, the inverse directional impact on human safety should also be considered: namely human-awareness [6, 7].

Inspired by the concept of egocentric and allocentric spatial cognition [14], we propose a framework using ego-safety to measure collision from the robot's perspective and social-safety to measure the impact of our robot's actions on surrounding pedestrians. Specifically, social-safety is defined as the intrusion into surrounding human's interaction area considering a look-ahead distance (fig. 2A). We train a policy in our specially designed simulator with both randomized maps, obstacles, and moving agents of randomized numbers, sizes, shapes, and locations. Our learned policy is then directly evaluated in Gazebo and real robot test. Results

†Noah's Ark Lab, Huawei Technologies Canada, Ltd., Edmonton AB., Canada, {jun.jin1, minh.nhat.nguyen, daniel.graves, hengshuai.yao}@huawei.com

*Department of Computing Science, University of Alberta, Edmonton AB., Canada {jjin5, nazmus, mj7}@ualberta.ca

show the learned policy demonstrates both time-efficient path planning behavior in a mapless navigation task and cooperative behavior which actively drives the robot into the crowd flows while showing respect to nearby pedestrians.

II. Related Works

This work is inspired by researches in the following topics.

Mapless Navigation: The problem of mapless navigation [2, 11] studies how to autonomously navigate a robot in unknown environments given a predefined target. The assumption of accessibility to a 3rd party localization module is commonly applied, however limits its practical applications. For example, in outdoor scenarios, there are situations when GNSS signals are weak or denied by surroundings; in indoor scenarios, deploying a localization solution is tedious. Visual navigation [15–18] approaches remove the assumption, however, is challenging to learn dynamic collision avoidance behavior simply from visual inputs. One potential solution is a *hierarchical approach* [17] which has both a high-level global path planner and low-level motion controller. The high-level global path planner, is obtained from either a visual navigation module or simply GPS waypoints, that enables long-term navigation. The low-level motion controller learns both static and dynamic collision avoidance behavior to reach a local target defined by the global path planner. In this paper, we assume such high-level path planner for a long-term navigation is given, the focus is to learn a local navigation controller.

Social-aware/ Human-aware robot navigation: Following the same approach as CADRL, *social-aware* navigation problem is further considered which views more human-robot interactions in the navigation behavior. Chen et al. [9] propose SACADRL which considers human-like social norm behavior by adding a complex social norm reward. Similarly, Tai et al. [19] train a social-compliant policy from RGB-D raw data inputs. These methods consider *human-aware* robot navigation with social rules (norm) and human comfort. Apart from these complicated considerations, a simplified version is practical by only considering learning a cooperative ability to solve the freezing robot problem [5]. Pfeiffer et al. [10] propose a framework to learn cooperative motion planning behavior by modeling human-robot interaction using maximum entropy methods. Chen et al. [4] jointly model human-robot and human-human interaction followed by a similar value function estimation in CADRL [3]. Above methods depend on explicit pedestrian detection.

Navigation learning from observations: Recently, learning the navigation strategy end-to-end directly from 2D laser scan readings proves possible. Long et al. [12] propose a framework to train a multi-agent collision avoidance policy using PPO [20] which maps 3 frames of historical laser scans to continuous robot actions. They show the learned model can be further directly applied in dense crowd navigation tasks [13]. Apart from using 2D laser scans, navigation learning from visual inputs, a.k.a. visual navigation problem [16, 17] has also been intensively studied.

III. Approach

A. Problem Formation

We formulate the problem as a Partially-Observable Markov Decision Process (POMDP) defined by a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \Omega, \mathcal{O})$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the state transition model, $\mathcal{R}$ is the reward function, Ω is the observation space, $\mathcal{O}$ is the probability function defining how observations are obtained from the underlying environment state ($\mathbf{o}^t \sim \mathcal{O}(\mathbf{s}^t)$). The robot can only observe laser scan $\mathbf{o}_z^t$ from surroundings and relative location $\mathbf{o}_g^t$ to the target.

We define $\mathbf{o}^t = [\mathbf{o}_m^t, \mathbf{o}_g^t]$. $\mathbf{o}_m^t$ is a motion feature which encodes the laser scan response changes along time axis due to surrounding motions. Robot action $\mathbf{a}^t$ is sampled from a stochastic policy π given observation $\mathbf{o}^t$: $\mathbf{a}^t \sim \pi_\theta(\mathbf{a}^t|\mathbf{o}^t)$, where $\mathbf{a}^t = [\mathbf{a}_x^t, \mathbf{a}_y^t]$ is within range [-1.5,1.5]. Considering a nonholonomic kinematics model, $\mathbf{a}^t$ is further converted to linear and angular velocity $[\mathbf{v}_l^t, \mathbf{v}_w^t]$ with $\mathbf{v}_l^t = sqrt(\mathbf{a}_x^t, \mathbf{a}_y^t)$ and $\mathbf{v}_w^t = atan(\mathbf{a}_y^t/\mathbf{a}_x^t) \in [-\pi, \pi]$.

1) Motion features: $\mathbf{o}_m^t$ plays as a prediction feature to learn a cooperative path planning behavior among dynamics. We consider disentangling robot ego motion out while constructing $\mathbf{o}_m^t$ from historical laser scans $\{\mathbf{o}_z^{t-k}, ..., \mathbf{o}_z^t\}$. We observe laser scan response can be changed by either robot or surrounding motions. By disentangling, motion feature $\mathbf{o}_m^t$ will only encode surrounding motions, which helps our policy training. While there are advanced methods to detect dynamic objects from laser scans [21, 22], we observe that during a small step duration, laser reading is affected mostly by rotation rather than translations. For the efficiency of simulator training, we simply calibrate previous laser scan $\mathbf{o}_z^{t-i}$ based on the difference between robot heading angles at time t-i and t by a shift operation in the scan array. Fig. 3 shows the comparison between calibrated and uncalibrated results.

2) Reward function: The reward at each step is obtained from system state $\mathbf{S}_t$ which is fully accessible in simulator training. Our reward function is a sum of ego-safety R_e, social-safety R_s and reaching target R_g rewards:

$$\mathcal{R}(\mathbf{s}_t) = R_e(\mathbf{s}_t) + R_s(\mathbf{s}_t) + R_g(\mathbf{s}_t) \tag{1}$$

We define the ego-safety zone (fig. 2B) of the robot as a circle around with radius $(r_i + 0.4)$, where r_i is the physical dimension of the robot. Given a set of nearby pedestrians $\{p_j^t\}$ and obstacles $\{B_k\}$, their closest distance to robot is d^t. $R_e(\mathbf{s}_t)$ is defined as:

$$R_e(\mathbf{s}_t) = \begin{cases} -10 & \text{if collision happens} \\ -0.25 * (1 - \frac{d^t}{r_i + 0.4}) & \text{else if } d^t < r_i + 0.4 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

We define the social-safety zone (fig. 2C) of each pedestrian as their interaction region stretching along current moving direction with a minimum safe headway distance in Δt. For computational efficiency, only pedestrians within 5m are considered. This zone is represented as a rectangle defined

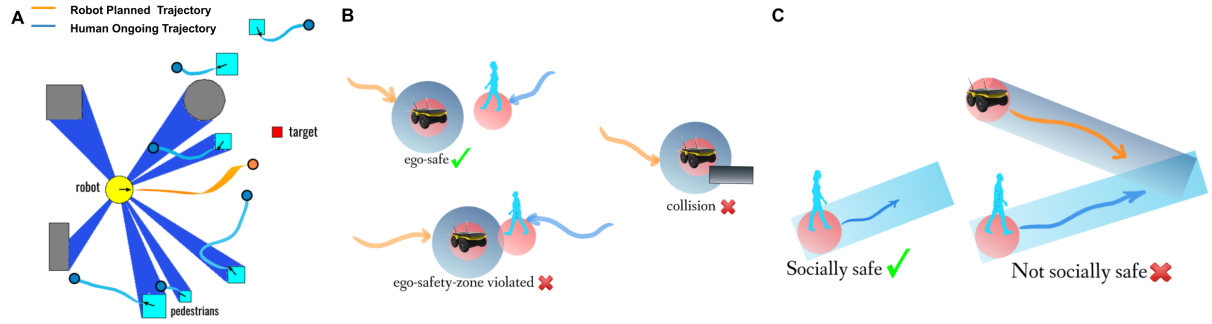


Fig. 2: **A**: An example of robot and surrounding pedestrians' movements during a time window of 40 steps. **B**: Illustration of robot ego-safety zone. **C**: Illustration of pedestrian's social-safety zone.

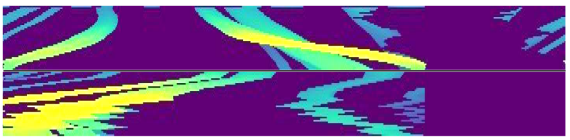


Fig. 3: Visualization of $\mathbf{o}_m^t$ with calibration (top) and without (bottom). When either the robot or a nearby pedestrian moves, the laser scan response will change, so will $\mathbf{o}_m^t$. We disentangle robot ego motion out to help the policy training.

in pedestrian's reference frame with $width = -r_{object}$ and $height = r_{object}/2 + d_{min} + \Delta t.\mathbf{v}_{object}$, where r_{object} is the radius of pedestrian bounding circle, $d_{min} = 0.5m$ is the minimum safe distance, $\mathbf{v}_{object}$ is the current speed. We set $\Delta t = 0.77s$ in training. The safety zones are then projected back into the world frame. Then we also construct the ego robot's social-safety zone likewise, and check if it intersects with that of other moving pedestrians. If intersects, we count it as one violation. Now we define $R_s(\mathbf{s}_t)$ as:

$$R_s(\mathbf{s}_t) = -0.1 * \frac{\text{number of violations}}{\text{total number of pedestrians}} \tag{3}$$

Let p^t be position of robot and p^* be the target. The last term $R_g(\mathbf{s}_t)$ is defined as:

$$R_g(\mathbf{s}_t) = \begin{cases} +10 & \text{if reach the target} \\ -0.01 * \frac{\|p^t - p^*\|}{\|p^0 - p^*\|} & \text{otherwise} \end{cases} \tag{4}$$

B. Parameterization

Our special consideration in parameterization is to make our network less sensitive to object shapes, which are mostly encoded in data along column axis in $\mathbf{o}_m^t$. So we design a large kernel size on the column direction followed by a max-pooling layer. We use DDPG [23] for training. Our actor and critic network structure are illustrated in fig. 4.

C. Training the Policy

1) Simulator: The policy is trained on our hand designed simulator (fig. 5) that runs a laser scanner at 40Hz and an inner differential drive controller at 20Hz. We model the laser scan simulation according to Hokuyo UTM-30LX with a 270° scanning range and 0.1 meter and 10 meters scanning distance. Historical 40 frames of laser scan are used to construct $\mathbf{o}_m^t$. The number, size and shape of static obstacles are randomized in simulation. Pedestrian behavior is also randomized with the following considerations: number of pedestrians in the scene, each pedestrian's current pose / velocity, each pedestrian's desired target and direction, each pedestrian's geometric shape and size, the behavior of stop-and-go, new walk-in pedestrians from random directions. Each pedestrian's behavior is controlled using ORCA [24] but ignores the robot since otherwise pedestrians will always avoid the robot.

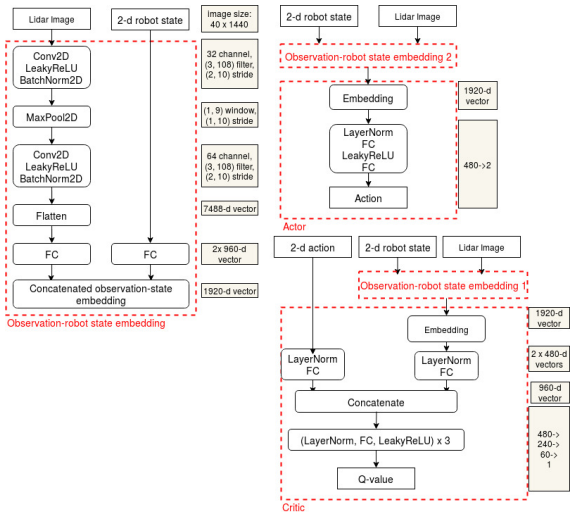


Fig. 4: Structure of our actor and critic network.

2) Training strategy: We train an *ego policy* by only including the ego-safety reward and a *social policy* by adding both ego-safety and social-safety rewards for comparison. The policy is trained using DDPG [23]. To facilitate training, the social policy is trained based on learned ego policy parameters.

IV. Evaluation

We evaluate the learned policy on *three tasks* that all require the robot reaching a target: (1) Mapless path planning: The robot is required to find a path while avoiding obstacles in an unknown environment without pedestrians. (2) Navigation among dynamic crowds: Given a dynamic

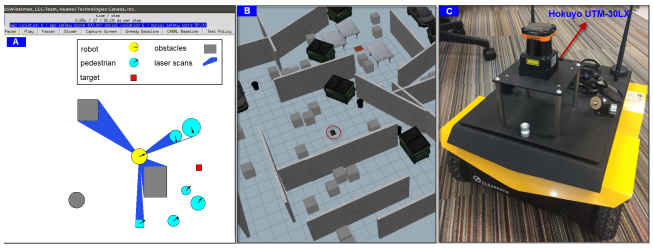


Fig. 5: The policy is trained in our specially designed simulator (**A**) and tested under various settings on our simulator, Gazebo(**B**) and a Jackal robot(**C**) in real world environment respectively.

TABLE I: Results of mapless path planning test in simulator. For each test, the robot needs to reach a fixed target in 10 randomly generated maps.

	Success Rate	Arriving Time	Ego Score
Greedy Baseline	80 %	7.0 ± 1.6 s	100
(Ours) Ego Policy	100 %	3.1 ± 0.3 s	100
(Ours) Social Policy	100 %	3.3 ± 1.3 s	100

crowd with random behaviors, the robot needs to cross the crowd while showing awareness of social-safety without the 'freezing robot' behavior. (3) Mapless path planning + navigation among dynamic crowds: A combination of the above two.

Baselines: To the authors' best knowledge, there are no current methods that tackle both the mapless navigation and collision avoidance among dynamic crowds using 2D laser scans. However we still hand designed two baselines for comparison. (1) The greedy baseline takes input of laser scan and mimics human intuition on reaching the target while avoiding collisions locally. In the planner we take robot's direction and the difference of angle between the robot and the destination. We apply 1D convolution window with laser scan and find the index having the higher convoluted value with a p controller to turn the robot to the direction. (2) The CADRL [3] baseline takes the full state of surrounding pedestrians as input. It's not designed for avoiding static obstacles in path planning, so it's only used in navigation among the dynamics crowds. Though it can't handle laser scan observations, we still feed in full states of all humans obtained from simulator, while testing our methods using only laser scans.

Metrics: We designed four metrics for quantitative evaluation: (1) Success Rate (%) after 10 runs in random settings. (2) Arriving Time (s). (3) Ego Score (0-100) to measure ego-safety awareness of robot. Let k be the number of ego-safety violation steps, and N be total steps to reach the target, $Ego_Score = (1 - k/N) * 100$. (4) Social Score (0-100) to evaluate social-safety awareness of nearby human. Let m be the number of social-safety violation steps, $Social_Score = (1 - m/N) * 100$.

We compare both our ego policy and social policy to the above two baselines. Experiments are conducted in both our simulator and Gazebo for quantitative evaluation, while in real world tests for qualitative evaluation (fig. 3).

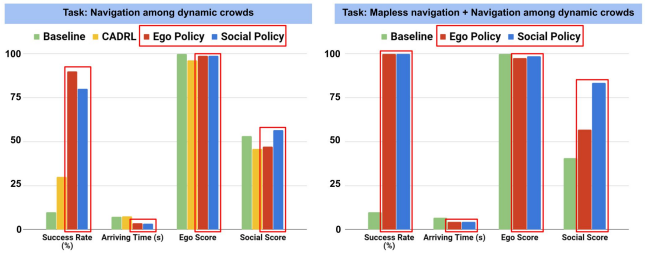


Fig. 6: Comparison between our method (ego policy, social policy) and the baseline in simulator. **Left**: results on the navigation among dynamic crowds task. **right**: results of the combined task.

A. *In the simulator*

1) Mapless path planning task: In this task, we fix the target while generating 10 random maps that only have static obstacles. Since CADRL can't handle this task, we only compare with greedy baseline (Table 1). Results show both our ego and social policy reach the target more efficiently with a higher success rate.

2) Navigation among dynamics task: We evaluate using 10 random crowds for each method. Each crowd has at least 8 pedestrians in a 5m×5m area with a chance of walk-in humans from a random direction. All pedestrians have no sense of the robot so they won't avoid it. We compare our methods to both baselines (Fig. 6-left).

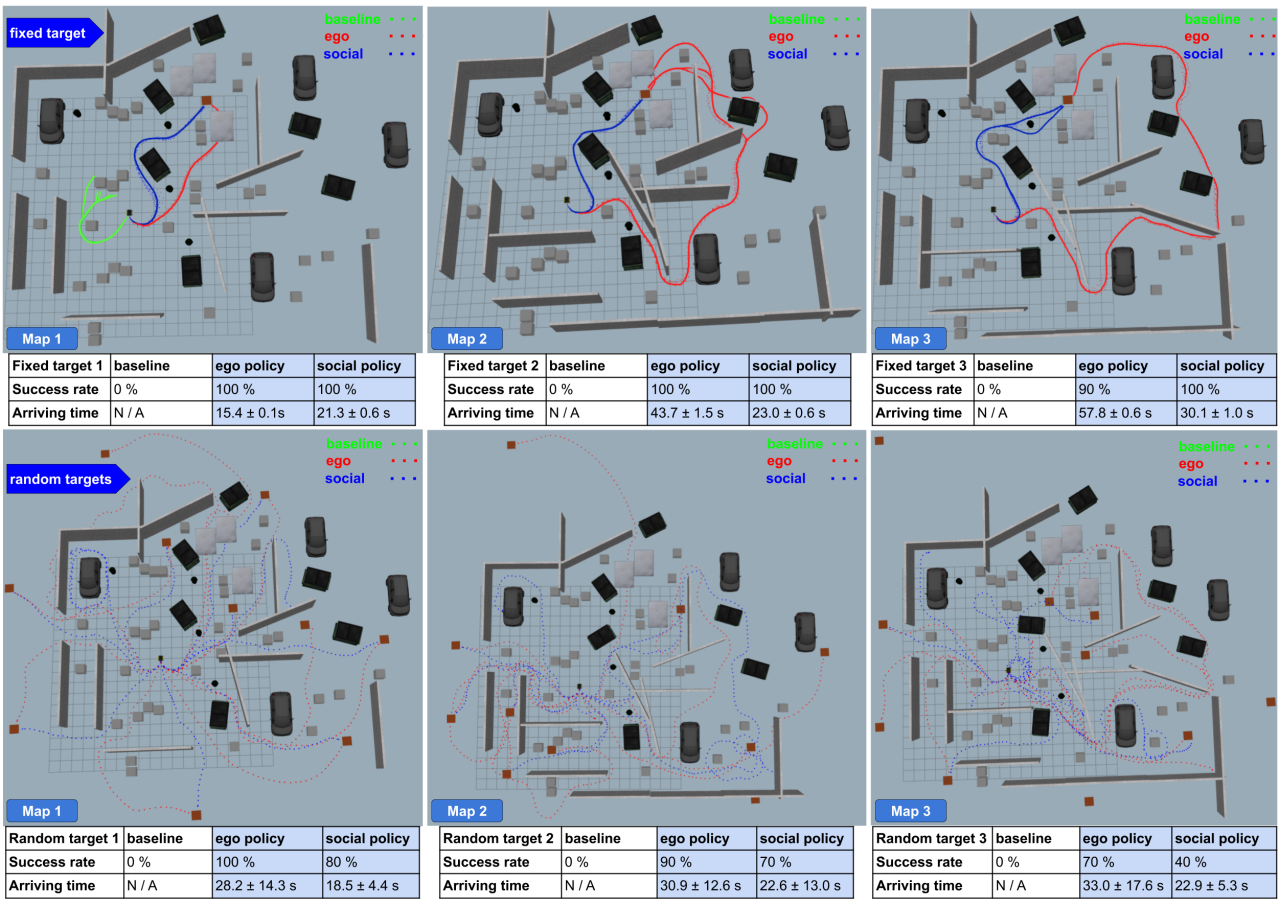
3) The combined task settings: Now we combine the above two task settings. Since CADRL can't work in mapless path planning, we only compare with the greedy baseline. Results are plotted in fig. 6-right.

Our results show the CADRL baseline typically has a passive 'wait and go' behavior which is not efficient and easily get trapped by crowds. In comparison, both our policies show more cooperative behaviors. This cooperative behavior helps the robot reach target more efficiently with higher success rate. While all methods achieve high ego score, the social policy shows better performance in social-safety awareness. By observation, the ego policy is more aggressive compared to our social policy. Detailed videos are on our website.

B. *In Gazebo environment*

We used the Gazebo maps and the ROS platform for our experiments.

1) Mapless path planning task: We evaluate the mapless navigation ability of our learned policy in three maps from simple to complex. Complexity increases from map 1 to map 3 by blocking key passages and adding more random obstacles. All tasks require the robot reaching the target (orange square) without knowing the map. We test in two scenarios: fixed target and random targets. Each scenario for each map runs 10 times on each method. Our results show the greedy baseline always fail as the map is too complex compared to simulator. Results (Fig. 7) show that all methods achieve ego score 100 in successful runs. We observe both ego and social policy show the behavior of slowing down when obstacles are highly aggregated. The



Fixed target 1	baseline	ego policy	social policy
Success rate	0 %	100 %	100 %
Arriving time	N / A	15.4 ± 0.1s	21.3 ± 0.6 s

Fixed target 2	baseline	ego policy	social policy
Success rate	0 %	100 %	100 %
Arriving time	N / A	43.7 ± 1.5 s	23.0 ± 0.6 s

Fixed target 3	baseline	ego policy	social policy
Success rate	0 %	90 %	100 %
Arriving time	N / A	57.8 ± 0.6 s	30.1 ± 1.0 s

Random target 1	baseline	ego policy	social policy
Success rate	0 %	100 %	80 %
Arriving time	N / A	28.2 ± 14.3 s	18.5 ± 4.4 s

Random target 2	baseline	ego policy	social policy
Success rate	0 %	90 %	70 %
Arriving time	N / A	30.9 ± 12.6 s	22.6 ± 13.0 s

Random target 3	baseline	ego policy	social policy
Success rate	0 %	70 %	40 %
Arriving time	N / A	33.0 ± 17.6 s	22.9 ± 5.3 s

Fig. 7: Evaluation results on the mapless path planning task in three Gazebo maps on a ROS environment. Map complexity increases from left to right. Up and down rows show result of a fixed target and a random target respectively.

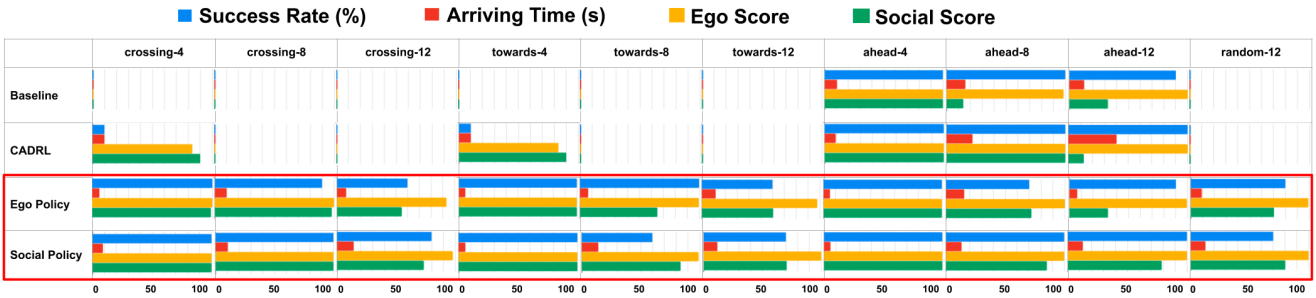


Fig. 8: Comparison between our method and two baselines on navigation among dynamic crowds tasks in Gazebo.

ego policy performs better since it positively runs via small gaps between obstacles. However, it easily gets trapped into dead loops, where it will run circle loops instead of tracing back to find a way out. As comparison, the social policy is more cautious even towards static obstacles. It always runs into wider open spaces and won't take the risk to go through small gaps, in which case, it will switch to other directions. If the gaps in all directions are small, it will run a circle loop.

2) Navigation among dynamics crowds: We designed four random behaved crowd scenarios (Fig. 9) considering crowd size 4, 8, 12 in a 5m×5m area. Each method on each scenario / crowd size combination needs to run 10 times to get evaluation results, as shown in Fig. 8. Results show our method reaches the target in a shorter time with a higher success rate. We observe the CADRL baseline prefers moving slowly to wait for crowds pass, however, may collide with human approaching from other directions. In comparison, our ego and social policies typically move into the crowd flow while avoiding pedestrians dynamically. In all tests, the social policy has the highest social score. Among the four scenarios, the 'towards' one is most difficult. Our ego policy has the best performance since it's more agile running through small gaps. The 'ahead' scenario is the easiest one. Both baselines simply move slowly behind the crowds. In contrast, our ego and social policy positively try to pass the crowd ahead. So they are significantly faster than the two baselines.

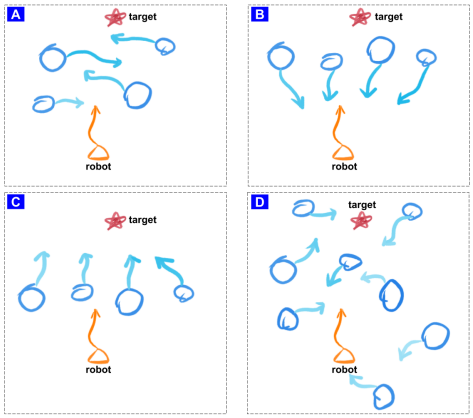


Fig. 9: The four crowd behavior scenarios. **A** crossing: The robot needs to move through the random crossing flow. **B** towards: Pedestrians are walking towards the robot. The robot needs to react fast since the relative velocity is large. **C** ahead: Pedestrians are walking ahead. The robot needs to decide how to pass. **D** random: All pedestrians' intention are randomized.

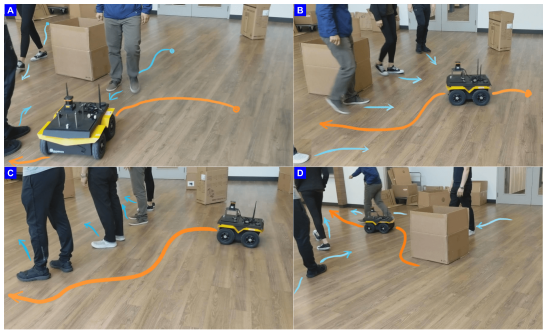


Fig. 10: Our observed robot behavior using social policy on the four crowd scenarios. **Blue** curve shows human trajectory; **Orange** curve shows robot trajectory. Our observation of robot behaviors in real world experiments aligns with our Gazebo test results. More evaluation results are included in our video.

C. In real world tests

We also conduct qualitative evaluations on a Jackal robot. It is worth noting the robot pose we used in training is accurate since it is obtained from the simulator. However, in our real world test, we don't have a precise 3rd party localization module due to resource limits. Using laser scan based SLAM localization[25] methods may help, but the localization output contains errors when running among dynamic human crowds. As a trade-off, we rely on robot's built-in odometry module which computes robot pose using wheel encoders and on-board IMU. To avoid drifting errors long time, we reset the odometry module after each run. We observe that, even with this coarse localization input, the robot performs relatively well.

We tested two tasks in an indoor room. For each task, the robot needs to navigate to the target while avoiding obstacles and humans. For each run, we assign the target coordinates in the robot's initial odometry frame when we reset it. In general, the target is around 3 to 6 meters ahead of the robot.

1) Mapless path planning task: We design two maps for this task without pedestrians. In each map, the robot needs to navigate to the target while avoiding static obstacles. We also test our method's performance while humans push obstacles into the robot's path. Evaluation results are shown in our video. For static obstacle avoidance in the two maps, both ego and social policies perform well. For pushing obstacles setting, the ego policy agilely avoids human kicking boxes while the social policy behaves more cautiously.

2) Navigation among dynamics task: We also test our social policy in the four crowd scenarios (Fig. 9). Fig. 10 shows the observed robot behavior. Our empirical results show the policy generalizes from regular geometric shape based obstacles in simulation to human legs in real world. This is achieved by specially designed convolution filter and stride size on the historical laser scan observations $\mathbf{o}_m^t$. The row axis of $\mathbf{o}_m^t$ encodes changes along the time line, while the column axis encode changes because of geometry shapes of surrounding objects. We designed large convolution filter and stride size on the column axis while small scales on the row axis.

In experiments we observed the circling behavior when the robot tries to pass an exit point out of an obstacle closure or avoid a near-close pedestrian, but restricted to its nonholonomic kinematics constraints. The policy learns to gradually change its twist velocities until precisely hit the exit point to escape the obstacle closure or avoid close pedestrians. Ideally, such behavior is not the most efficient one since the robot can simply stop, rotate and move. We leave it as a future problem to learn stop behaviors without sacrificing efficiency.

V. Conclusion

We propose a method to tackle the problem of mapless collision-avoidance navigation where humans are present using 2D laser scans by the formation of ego safety and social safety. Extensive experiments were conducted with both quantitative and qualitative evaluation. Results show our method can demonstrate cooperative path planning behavior by predicting the future motion of humans, while considering the impact of the robot actions on surrounding humans, namely the awareness of social-safety.

Though some success is achieved in this project, other practical considerations are worth pointing out: (1) To define the target, mapless navigation relies on a third party localization module. This is usually unrealistic considering it is expected to work in unknown environments. As discussed in section 2, a high level abstract path planner with vision sensor support may solve the problem. (2) We find the training process is still tedious to fine tune the parameters. Though [12, 13] propose a practical multi-stage training strategy which trains the policy from simple scenario to complex ones, hitherto, it lacks theoretical guarantees. We expect finite horizon reinforcement solutions may help since path planning problem essentially needs to consider multiple steps ahead.

The authors would like to give thanks to all our colleagues participating in the real world tests.

REFERENCES

[1] C. Giovannangeli, P. Gaussier, and G. Désilles, "Robust mapless outdoor vision-based navigation," in *2006 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2006, pp. 3293–3300.

[2] L. Tai, G. Paolo, and M. Liu, "Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 31–36.

[3] Y. F. Chen, M. Liu, M. Everett, and J. P. How, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 285–292.

[4] C. Chen, Y. Liu, S. Kreiss, and A. Alahi, "Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6015–6022.

[5] T. Fan, X. Cheng, J. Pan, P. Long, W. Liu, R. Yang, and D. Manocha, "Getting robots unfrozen and unlost in dense pedestrian crowds," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1178–1185, 2019.

[6] E. A. Sisbot, L. F. Marin-Urias, R. Alami, and T. Simeon, "A human aware mobile robot motion planner," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 874–883, 2007.

[7] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.

[8] G.-H. Liu, A. Siravuru, S. Prabhakar, M. Veloso, and G. Kantor, "Learning end-to-end multimodal sensor policies for autonomous navigation," *arXiv preprint arXiv:1705.10422*, 2017.

[9] Y. F. Chen, M. Everett, M. Liu, and J. P. How, "Socially aware motion planning with deep reinforcement learning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1343–1350.

[10] M. Pfeiffer, U. Schwesinger, H. Sommer, E. Galceran, and R. Siegwart, "Predicting actions to act predictably: Cooperative partial motion planning with maximum entropy models," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 2096–2101.

[11] O. Zhelo, J. Zhang, L. Tai, M. Liu, and W. Burgard, "Curiosity-driven exploration for mapless navigation with deep reinforcement learning," *arXiv preprint arXiv:1804.00456*, 2018.

[12] P. Long, T. Fanl, X. Liao, W. Liu, H. Zhang, and J. Pan, "Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6252–6259.

[13] T. Fan, X. Cheng, J. Pan, D. Monacha, and R. Yang, "Crowdmove: Autonomous mapless navigation in crowded scenarios," *arXiv preprint arXiv:1807.07870*, 2018.

[14] M. J. Proulx, O. S. Todorov, A. Taylor Aiken, and A. A. de Sousa, "Where am i? who am i? the relation between spatial cognition, social cognition and individual differences in the built environment," *Frontiers in psychology*, vol. 7, p. 64, 2016.

[15] J. J. Guerrero, R. Martinez-Cantin, and C. Sagüés, "Visual mapless navigation based on homographies," *Journal of Robotic Systems*, vol. 22, no. 10, pp. 569–581, 2005.

[16] Y. Zhu, R. Mottaghi, E. Kolve, J. J. Lim, A. Gupta, L. Fei-Fei, and A. Farhadi, "Target-driven visual navigation in indoor scenes using deep reinforcement learning," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 3357–3364.

[17] W. Gao, D. Hsu, W. S. Lee, S. Shen, and K. Subramanian, "Intention-net: Integrating planning and deep learning for goal-directed autonomous navigation," *arXiv preprint arXiv:1710.05627*, 2017.

[18] K. Chen, J. P. de Vicente, G. Sepulveda, F. Xia, A. Soto, M. Vázquez, and S. Savarese, "A behavioral approach to visual navigation with graph localization networks," *arXiv preprint arXiv:1903.00445*, 2019.

[19] L. Tai, J. Zhang, M. Liu, and W. Burgard, "Socially compliant navigation through raw depth inputs with generative adversarial imitation learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1111–1117.

[20] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[21] C. Mertz, L. E. Navarro-Serment, R. MacLachlan, P. Rybski, A. Steinfeld, A. Suppé, C. Urmson, N. Vandapel, M. Hebert, C. Thorpe, *et al.*, "Moving object detection with laser scanners," *Journal of Field Robotics*, vol. 30, no. 1, pp. 17–43, 2013.

[22] Z. Wang, "Laser-based detection and tracking of dynamic objects," Ph.D. dissertation, University of Oxford, 2014.

[23] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[24] J. Van Den Berg, S. J. Guy, M. Lin, and D. Manocha, "Reciprocal n-body collision avoidance," in *Robotics research*. Springer, 2011, pp. 3–19.

[25] S. Kohlbrecher, J. Meyer, O. von Stryk, and U. Klingauf, "A flexible and scalable slam system with full 3d motion estimation," in *Proc. IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*. IEEE, November 2011.