Teaching Personalized Robot Navigation through Virtual Reality Demonstrations: A Learning Framework and User Study

Jorge de Heuvel[1] Nathan Corral[1] Lilli Bruckschen[2] Maren Bennewitz[1]

***Abstract*— For the most comfortable, human-aware robot navigation, subjective user preferences need to be taken into account. This paper presents a novel reinforcement learning framework to train a personalized navigation controller along with an intuitive virtual reality demonstration interface. The conducted user study provides evidence that our personalized approach significantly outperforms classical approaches with more comfortable human-robot experiences. We achieve these results using only a few demonstration trajectories from non-expert users, who predominantly appreciate the intuitive demonstration setup. As we show in the experiments, the learned controller generalizes well to states not covered in the demonstration data, while still reflecting user preferences during navigation. Finally, we transfer the navigation controller without loss in performance to a real robot.**

I. Introduction

Robot personalization to specific user-preferences will become increasingly important, as robots find their way into our everyday life. Harmonic human-robot interactions build trust and satisfaction with the user [1], whereas negative interaction experiences can quickly lead to frustration [2]. A cause for negative user experiences can be algorithms that do not reflect personal preferences.

Where mobile household robots navigate in the vicinity of a human, basic obstacle avoidance approaches fail to capture individual user preferences. While collision avoidance is undoubtedly crucial during navigation, the navigation policy should furthermore be human-aware and take into account user preferences regarding proxemics [2] and privacy, compare Fig. 1 (bottom). Subjective preferences may vary depending on the environment [3], [4] and social context, e.g., navigation preferences could reflect in the robot's approaching behavior [5], or always driving in front or behind the human. In addition following a certain speed profile and maintaining a certain distance from humans and other obstacles in the environment might play a role. The resulting navigation objective for the robot is to reach the navigation goal, not necessarily by only following the shortest path, but also by taking personal robot navigation preferences into account.

Extensive research has been done on both human-aware navigation [6] and on robot personalization [1], [7], but surprisingly, very few can be found at the intersection of both disciplines. Recent advances in learning socially-aware navigation behavior from human demonstrations have been made with inverse reinforcement learning, where the parameters of a proxemics-encoding reward function were inferred [8]. Limited by the initial shaping of the reward function [9], such approaches lack the ability for navigation style personalization beyond the scope of the reward function. Similar drawbacks hold for learning or inferring cost-maps [10], [11]. For smooth navigation, reinforcement learning (RL) based continuous control has lead to promising results on mobile robots [12], [13]. Furthermore, off-policy RL methods can be complemented with demonstration data to greatly improve learning speed on a given task, even outperforming the resourcefulness of the original demonstrations [14]. However, RL robot navigation policies learn most efficient trajectories to the goal. These trajectories do not necessarily reflect the original demonstration behavior, which contains user preferences. To more precisely imitate behavior from demonstrations, behavioral cloning (BC) can be used [15]. However, the final policy is limited by the quality and amount of demonstration data [16]. The dataset would need to cover most of the state space to generalize fluently in unseen environments. This poses a problem, as human demonstrators can only provide limited amounts of demonstration data due to their finite patience [17]. The question crystallizes, how do we efficiently record personal preferences and teach them to the robot, without being limited by the quality and quantity

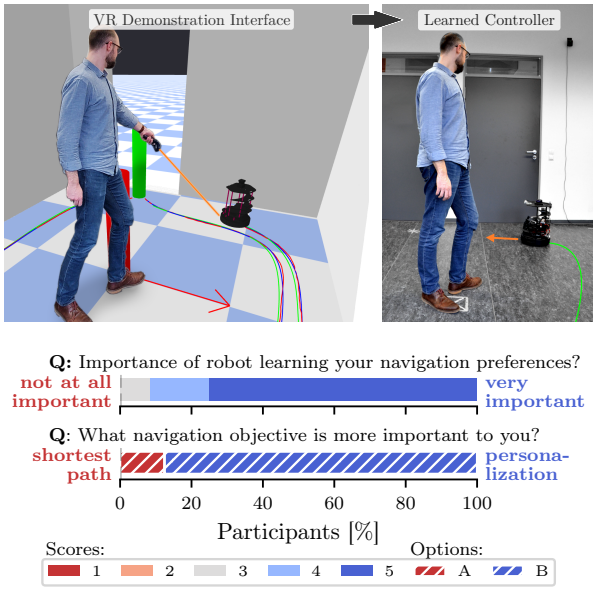

Fig. 1. **Top:** We propose a virtual reality (VR) interface to intuitively demonstrate robot navigation preferences by drawing trajectories onto the floor with a handheld controller. **Bottom:** User study survey results on the importance of personalized navigation behavior. Participants strongly expressed their preference for personalization of robot navigation behavior, even at the possible cost of longer trajectories.

[1] University of Bonn, Germany. [2] Fraunhofer FKIE, Bonn, Germany.

This work has partially been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the grant number BE 4420/2-2 (FOR 2535 Anticipating Human Behavior).

An extended version of this work is under review at an IEEE conference.

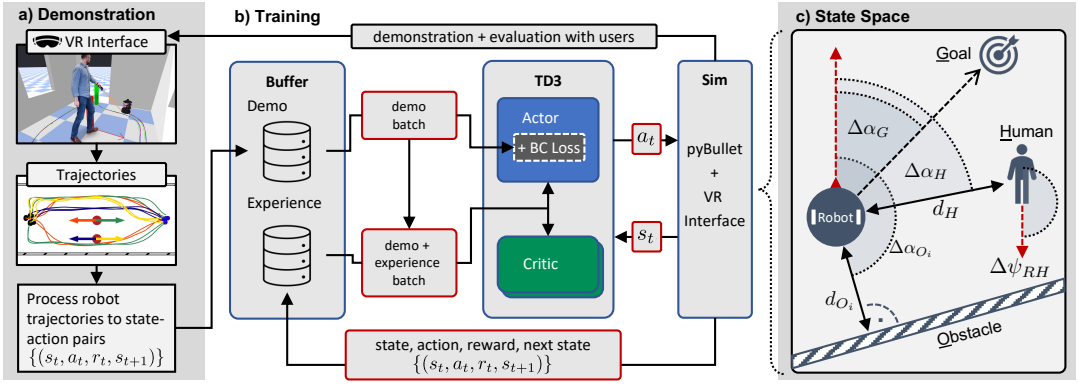

Fig. 2. Schematic representation of the used architecture. **a)** Demonstration trajectories are drawn by the user and fed into the demonstration buffer. **b)** A TD3 reinforcement learning architecture with an additional behavioral cloning (BC) loss on the actor trains a personalized navigation policy for the human-robot interaction with continuous control. The learned policy is then evaluated in VR and subsequently transferred to a real robot. **c)** The robot-centric state space captures the vicinity and orientation of the human and the obstacles as well as the goal direction.

of demonstrations.

In order to solve the aforementioned challenges, we propose a novel navigation learning approach together with a virtual reality (VR) interface to intuitively demonstrate robot navigation preferences by drawing trajectories onto the floor with a handheld controller, see Fig. 1. Importantly, the interface does not require expert-level knowledge on robotics, facilitating personalized navigation to a wide range of users. Our demonstration process is time-efficient, as only few demonstrations are required. The demonstrations are leveraged to successfully train a personalized human-aware navigation controller, by combining deep reinforcement learning and behavioral cloning. We show that our navigation policy closely reflects user preferences from only a few demonstrations. But at the same time, it generalizes to unseen states. In an extensive user study, we evaluate the personalized navigation behavior against classical navigation approaches both in VR and on a real robot.

The threefold **main contributions** of our study are:

- A VR demonstration interface for teaching navigation preferences to robots intuitively.
- Learning a user-personalized, context-based navigation policy based on the combination of RL and BC.
- An interactive user study recording user specific navigation preferences, evaluating both the presented interface and learned personalized navigation policies.

II. Reinforcement Learning from Demonstrations

We adapted a twin-delayed deep deterministic policy gradient (TD3) architecture consisting of an actor and two critic networks [18]. TD3 was chosen for two reasons: i) It has a continuous action space allowing smooth robot control and ii) it is off-policy, thus is a perfect candidate for use with demonstration data. The actor network outputs two continuous robot control commands, i.e., forward and angular velocity. We introduce two modifications to classic TD3, similar to Nair *et al.* [19]: i) a behavioral cloning loss on demonstration data for the actor network and ii) a separate buffer to hold demonstration data, see Fig. 2.

A wheeled robot that has a local navigation goal navigates in the vicinity of a single human. A visualization of our robot-centric state space capturing the human pose, orientation and obstacles is shown in Fig. 2c. All of those parameters can play a role for the robot navigation preferences of the user, as they encode proxemics, the human's field of view and relative positioning in the room. We assume that the positions and orientations of the human, the robot, and all obstacles are known. The functionality of our approach is proven for a single human in the vicinity of the robot, however, it is expandable to more than one person.

We aim to teach the user-specific navigation preferences not by complex reward shaping, but only via demonstration data. Consequently, we keep the reward as sparse as possible, besides basic collision penalties and goal rewards: $r = r_{\text{collision}} + r_{\text{goal}} + r_{\text{timeout}}$. Upon collision with the human or an obstacle $r_{\text{collision}} = -5$. When the goal is reached, we provide $r_{\text{goal}} = +5$ exclusively in the demonstration data. This is to boost the value of demonstration-like behavior over more efficient, shortest-path navigation behavior. Inefficient actions that lead to an episode timeout result in $r_{\text{timeout}} = -\frac{5}{2}$ if episode timeout $(n > N_{\text{ep}})$. The rewards are 0 in all other cases, respectively.

III. Demonstration and Training Environment

We propose a novel VR demonstration setup, where the user teaches the robot personal navigation preferences in a virtual reality environment, see Fig. 2a.

*1) **Simulator and Robot**:* Our robotic platform is the Kobuki *Turtlebot 2*. As a VR and physics simulator we use Pybullet [20].

A key challenge in using demonstrations for reinforcement learning is bridging the gap between the agent's and the demonstrator's state space. Given a desired forward velocity v, we analytically calculate action commands along a demonstration trajectory, so that the robot follows discrete segments $(\Delta d, \Delta\alpha)$ along a trajectory by executing successive actions calculated at the control frequency f. By integrating the forward and angular velocities v and ω over time $t = \frac{1}{f}$, we can derive the relation $\frac{v}{\omega} = \frac{\Delta d}{\Delta\alpha}$ for the finite distance $\Delta d = v\Delta t$ and angle $\Delta\alpha = \omega\Delta t$.

*2) **Collecting and Processing Demonstration Trajectories**:* We use the following steps to process raw demonstration

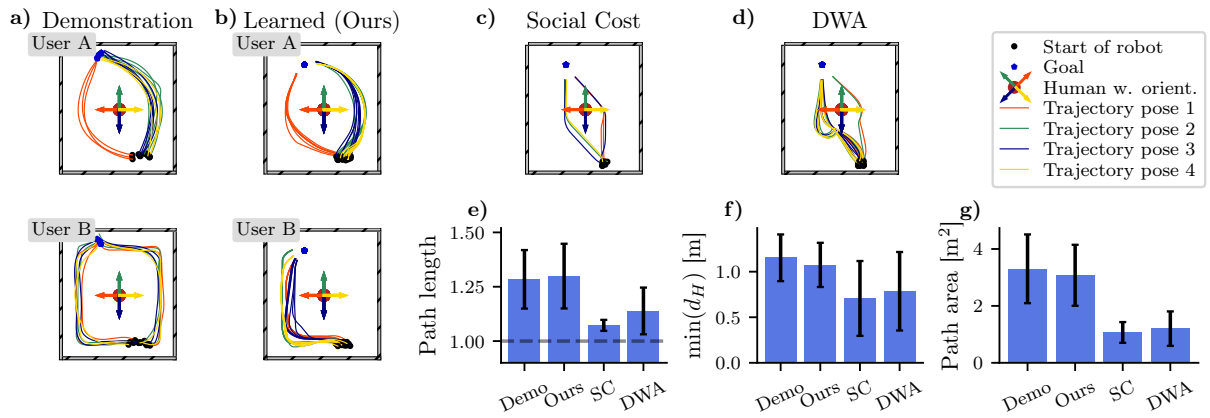

Fig. 3. **a)** The demonstrated robot navigation preference trajectories of two participants A and B are shown for different human position-orientation pairs (color-coded). Note the wall-following preference of user B, whereas user A prefers a smooth curve navigation style. **b)** The personalized controller successfully learned to reflect the individual user preferences. Note that when no specific side preference is given as in the demonstrations in the corridor, the controller reproduces trajectories mainly on one side. We evaluated our approach against **c)** the social cost model and **d)** the Dynamic Window Approach. A quantitative comparison of the different approaches reveals **e)** a higher relative path length (normalized by linear distance) and **f)** a higher preferred minimum distance. **g)** The increased path area for our controller (between the learned trajectory and linear distance) also points to a general preference for earlier deviation from the shortest path in favor for more comfortable trajectories.

trajectories into state-action pairs (s_t, a_t, r_t, s_{t+1}) contained in the demonstration buffer:

1) In VR, a user draws a trajectory using the handheld controller, emitting a beam of light. The analogue trigger on the controller backside allows to control the robot speed linearly in the range from $v_{\text{min}} = 0.1\,\text{m}\,\text{s}^{-1}$ to $v_{\text{max}} = 0.25\,\text{m}\,\text{s}^{-1}$ at the drawing location.
2) The drawn trajectory is interpolated and smoothed with a 2D spline, parameterized by $k \in [0, 1]$. Also, the speed information is spline-interpolated.
3) Based on the speed along the spline $v(k)$, we consecutively extract the locations at which the robot receives a new control command, using $\Delta d = v(k)\Delta t$.
4) Given $v(k)$ for all control command locations, the corresponding angular velocities ω are calculated.
5) The robot is placed and oriented according to the trajectory's starting point.
6) Successively, the control command tuples $a_t = (v_t, \omega_t)$ are executed and the robot follows the trajectory.
7) Before and after executing an action a_t, we record the corresponding states s_t, s_{t+1} and the reward r_{t+1}.
8) Finally, all state-action-reward pairs $(s_t, a_t, s_{t+1}, r_{t+1})$ are stored in the demonstration buffer.

We use data augmentation ($N_{\text{aug}} = 15$) to increase the data output from a single demonstration trajectory by slightly shifting the start position, while preserving its original character ($\max(\Delta d) = 5\,\text{cm} \ll$ environment scale).

IV. Experimental Evaluation

This section highlights the results of our user study and provides a qualitative and quantitative analysis of the learned personalized navigation controller.

*1) **User Study**:* We conducted a two-session user study with 24 non-expert participants (13 male, 11 female) to i) record individual navigation preferences (demonstration session), ii) evaluate the navigation behavior learned by our personalized controller (evaluation session). The user study featured a room and a corridor environment. However, this short paper only shows the room environment results.

a) Demonstration Session: During the demonstration session, preference trajectories were recorded, see Fig. 3a. The environment featured four position-orientation pairs (color-coded) for the participant. For each pair, between three and five trajectories were recorded. The total time investment was about 20 min for each participant.

b) Evaluation Session: During the second session, our personalized navigation approach was evaluated against two approaches in virtual reality in unknown order: The Dynamic Window Approach (DWA) [21] using the ROS *move_base* package [22] in combination with a 2D lidar sensor, and a social cost model (SC) based on the configuration of [23]. Each navigation approach was shown in VR for all four position-orientation pairs (cf. Fig. 3b-d), followed by an evaluation survey. The survey questions and results are shown in Fig. 5a). All of the following findings are statistically significant: Regarding the comfort and closeness perception of the robot trajectories, our approach outperformed both the SC and DWA. Participants saw their preference reflected mainly in our personalized controller.

c) Real Robot Evaluation: Our personalized controller was demonstrated on the real robot (room environment) to investigate the participant's transition experience from the simulated to the real robot. The real robot evaluation was also complemented by a survey, see Fig. 5b). As in VR, the navigation of the real robot was predominantly experienced comfortable and participants saw their preferences mostly reflected. Furthermore, the transition from the simulated robot experience in VR to the real robot was mostly experienced as *very natural*.

*2) **Qualitative Navigation Analysis**:* Fig. 3a shows demonstration data from two participants. The preference of participant A is a smooth curve around their position, while the robot drives in their field of view when approaching from either side. Interestingly, participant B's preference is a wall-following robot that navigates at higher distance to the

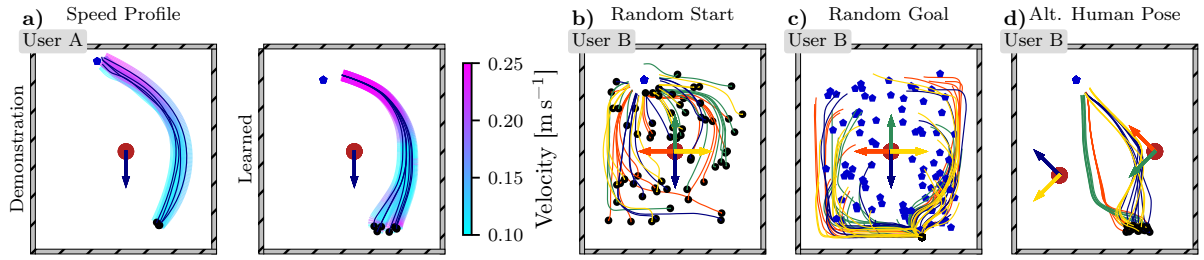

Fig. 4. **a)** User A demonstrated a distinct speed profile when facing the robot start position. It was successfully adapted by the learned controller. Furthermore, we tested the ability for generalization of the learned controller threefold by showcasing state configurations not covered by the demonstration data: **b)** When the robot starts at a random position in the environment, its navigation behavior still reflects the characteristics of the trajectory from the user demonstrations (cf. Fig. 3a). **c)** Even when its goal is randomly placed in the room, the robot exhibits the distinct user preferences. **d)** The user's position and orientation was altered to non-demonstration configurations. When the human is obstructing the robot's path while facing the wall, the robot traverses behind the human. In all cases, a distinct distance is kept to the human, as demonstrated by both users. This shows nicely how the navigation agent improved beyond the limits of the demonstration data provided. For a legend, please refer to Fig. 3.

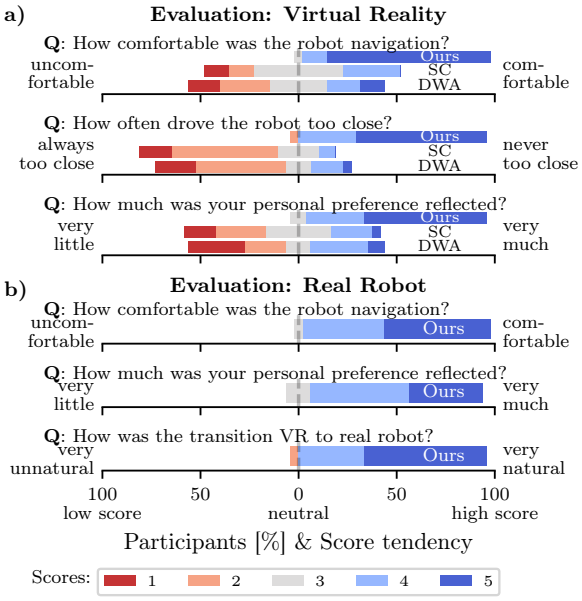

Fig. 5. User study survey results of the evaluation session. **a)** Evaluation: In virtual reality, both the Dynamic Window Approach and the social cost model were outperformed by our personalized controller in various aspects. **b)** On the real robot, our novel personalized controller was perceived predominantly positive as well. The plot-bar's positions are aligned to the neutral score (3) to indicate overall rating.

human, compared to participant A.

Fig. 3b shows trajectories of the learned navigation behavior. The learned policy clearly reflects the characteristics of the demonstration trajectories. Furthermore, the robot adjusts its navigation trajectory according to the human orientation. For user A, it learned to traverse in the field of view, compare yellow orientation and trajectories. In participant B's demonstration, trajectories from a single position-orientation pair traverse both in front and behind the participant. Here, no specific side preference is given and the controller reproduces trajectories mainly on one side.

Beside trajectory shape, users demonstrated speed profiles along the demonstration trajectories. As an example, Fig. 4a depicts how user A demonstrated a distinct speed profile when directly facing the robot start position in the room environment. After the robot slowly approached and passed by, it was allowed to accelerate. As can be seen, the behavior is picked up by the controller during training.

*3) **Quantitative Navigation Analysis**:* Fig. 3e-g compare quantitative properties of all three evaluation approaches and demonstrations from all 24 study participants. The personalized navigation trajectories are on average longer than those executed by DWA or SC, while maintaining a higher minimal distance to the human, averaged at $(1.1 \pm 0.2)\,\mathrm{m}$. The path area is calculated between the trajectory and linear distance from start to goal. A higher path area reveals earlier deviation from the linear path in favor of personalization, as it is the case for our personalized controller, compare Fig. 3g. This clearly indicates that users prefer personalized navigation trajectories over shortest path navigation. Furthermore, the large standard deviation of the path area indicates a high trajectory shape variability among the participants.

*4) **Generalization**:* Finally, we tested the ability for generalization of the learned navigation policy with states not covered by the demonstrations. First, the robot started at random positions in the environment (cf. Fig. 4b). Second, goal positions were randomized (cf. Fig. 4c). Thirdly, we tested altered human positions and orientations (cf. Fig. 4d). In all three cases reflect distinct demonstration characteristics, compare Fig. 3a.

As demonstrated with these results, our framework can successfully learn a personalized navigation controller that improves beyond the limits of few demonstration trajectories.

V. Conclusion

To summarize, we presented both a learning framework and an intuitive virtual reality interface to teach navigation preferences to a mobile robot. From a few demonstration trajectories, our context-based navigation controller successfully learns to reflect user-preferences and furthermore transfers smoothly to a real robot. The conducted user study provides evidence that our personalized approach significantly surpasses standard navigation approaches in terms of perceived comfort. Furthermore, the study verifies the demand for personalized robot navigation among the participants. Our results are a first important step towards personalized robot navigation, made possible by our interface and user study. As a next logical step, we will transfer the framework to more complex and diverse environments.

References

[1] N. Gasteiger, M. Hellou, and H. S. Ahn, "Factors for personalization and localization to optimize human–robot interaction: A literature review," *International Journal of Social Robotics*, 2021.

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

[3] X. Xiao, B. Liu, G. Warnell, J. Fink, and P. Stone, "Appld: Adaptive planner parameter learning from demonstration," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, 2020.

[4] H. Zender, P. Jensfelt, and G.-J. M. Kruijff, "Human-and situation-aware people following," in *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2007.

[5] M. Luber, L. Spinello, J. Silva, and K. O. Arras, "Socially-aware robot navigation: A learning approach," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012.

[6] R. Möller, A. Furnari, S. Battiato, A. Härmä, and G. M. Farinella, "A survey on human-aware robot navigation," *Robotics and Autonomous Systems*, vol. 145, 2021.

[7] M. Hellou, N. Gasteiger, J. Y. Lim, M. Jang, and H. S. Ahn, "Personalization and localization in human-robot interaction: A review of technical methods," *Robotics*, vol. 10, no. 4, 2021.

[8] M. Kollmitz, T. Koller, J. Boedecker, and W. Burgard, "Learning human-aware robot navigation from physical interaction via inverse reinforcement learning," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020.

[9] A. Y. Ng, D. Harada, and S. Russell, "Policy invariance under reward transformations: Theory and application to reward shaping," in *Icml*, vol. 99, 1999.

[10] K. Bungert, L. Bruckschen, S. Krumpen, W. Rau, M. Weinmann, and M. Bennewitz, "Human-aware robot navigation based on learned cost values from user studies," in *2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*, 2021.

[11] N. Pérez-Higueras, F. Caballero, and L. Merino, "Teaching robot navigation behaviors to optimal rrt planners," *International Journal of Social Robotics*, vol. 10, no. 2, 2018.

[12] 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.

[13] M. Pfeiffer, S. Shukla, M. Turchetta, C. Cadena, A. Krause, R. Siegwart, and J. Nieto, "Reinforced imitation: Sample efficient deep reinforcement learning for mapless navigation by leveraging prior demonstrations," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, 2018.

[14] M. Vecerik, T. Hester, J. Scholz, F. Wang, O. Pietquin, B. Piot, N. Heess, T. Rothörl, T. Lampe, and M. Riedmiller, "Leveraging demonstrations for deep reinforcement learning on robotics problems with sparse rewards," *arXiv:1707.08817 [cs]*, 2018.

[15] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, 2009.

[16] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, "Recent advances in robot learning from demonstration," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, no. 1, 2020.

[17] A. L. Thomaz, "Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance."

[18] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proceedings of the 35th International Conference on Machine Learning*. PMLR, 2018.

[19] A. Nair, B. McGrew, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Overcoming exploration in reinforcement learning with demonstrations," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018.

[20] E. Coumans and Y. Bai, "Pybullet: physics simulation for games visual effects robotics and reinforcement learning," 2016.

[21] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics Automation Magazine*, vol. 4, no. 1, 1997.

[22] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3. Kobe, Japan, 2009.

[23] M. Kollmitz, K. Hsiao, J. Gaa, and W. Burgard, "Time dependent planning on a layered social cost map for human-aware robot navigation," in *2015 European Conference on Mobile Robots (ECMR)*. IEEE, 2015.