Personalized Human-Robot Interaction: Learning Depth Vision-Based Navigation with User Preferences

Jorge de Heuvel Nathan Corral Benedikt Kreis Maren Bennewitz

Abstract— **For the best human-robot interaction experience, the robot's navigation policy should take into account personal preferences of the user. In this paper, we present a learning framework complemented by a perception pipeline to train a depth vision-based, personalized navigation controller from user demonstrations. Our perception pipeline enrolls a variational autoencoder. It compresses the perceived depth images to a latent state representation to enable efficient reasoning of the learning agent about the robot's environment. In a detailed analysis and ablation study, we evaluate different configurations of the perception pipeline. We discuss the robot's navigation performance in various virtual scenes and demonstrate the first personalized robot navigation controller that solely relies on depth images.**

I. Introduction

The personalization of robots will be a key factor for comfortable and satisfying human-robot-interactions. As the integration of robots at home or at work will inevitably increase, the number one goal should be a naturally collaborative experience between users and the robot. However, users might have personal preferences about specific aspects of the robot's behavior that define the personal golden standard of interaction. Falling short of user's preferences could lead to negative experiences and consequently frustration [1].

Where humans share the same environment with a mobile robot, the robot's navigation behavior significantly influences the comfort of interaction [2]. Consequently, basic obstacle avoidance approaches are insufficient to address individual preferences regarding proxemics, trajectory shape, or area of navigation in a given environment, while being a key component to successful navigation without question. Instead, a robot's navigation policy should be aware of humans [3] and reflect the users' personal preferences.

In our previous work [2], we demonstrated that pairing a virtual reality (VR) interface with a reinforcement learning (RL) framework enables the demonstration and training of highly customizable navigation behaviors. Ina user study, the presented personalized controller significantly outperformed traditional, established local navigation approaches with regards to comfort of user in the vicinity of the moving robot. However, a key assumption of this work is the known pose of static human and obstacles. To overcome these assumptions, enrolling a 3D perceiving depth vision sensor to sense both human and obstacles is a possible solution [4]. However, depth vision cameras come at the cost of high-dimensional, complex, and redundant output, from which it is challenging to learn [5]. The question crystallizes, how do we learn from preferences of moving users in realistic 3D environments, while relying on state-of-the art sensor modalities?

To solve the challenges above, we introduce a depth vision-based perception pipeline that is both lightweight, human-aware and, most importantly, provides the robot with a low-dimensional representation of the 3D scene. This pipeline i) detects the human and obstacles, ii) compresses the perceived depth information, and iii) enables efficient reasoning about the robot's dynamic environment to the learning framework. Our new system is able to learn personalized navigation preferences from VR demonstrations for dynamic scenes in which both robot and human move.

The **main contributions** of our work are: 1) Learning a preference-reflecting navigation controller that relies solely on depth vision. 2) An qualitative and quantitative analysis of different perception configurations. 3) An ablation study to investigate the learned 3D scene understanding.

II. Related Work

Adjusting or learning the navigation behavior of a robot based on feedback or demonstration has been the focus of various studies [6], [7]. Especially, deep learning-based approaches shine by their ability to learn from subtle and implicit features in their environment [8], [9], [10]. Fusing the potential of user demonstrations with a learning architecture led to promising results in the field of robotic manipulation tasks [11] and has successfully been applied the field of robot navigation for personalization [2].

Vision-based sensor modalities for navigation appeal due to their cost-efficiency. For human-aware navigation, the detection and explicit localization of pedestrians enabled socially conforming navigation controllers [4], [12].

Recent advances in the field of depth vision-based navigation in combination with RL have been made by Hoeller *et al.* [13], who study a latent state representation of depth-images to efficiently learn navigation in dynamic environments. Our proposed perception pipeline is built upon their successful architecture.

While in our previous work [2] we presented one of the first approaches at the intersection of navigation and robot personalization, we now enhance the system by using only depth vision as input.

III. Our Approach

In this work, we consider a robot navigating in the same room as a single, human user. The user has personal preferences about the way the robot circumnavigates him/her while pursuing a local goal in the same room. Such preferences could lie in the approaching behavior or the robot's trajectory.

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 will be published at IEEE IROS 2023.

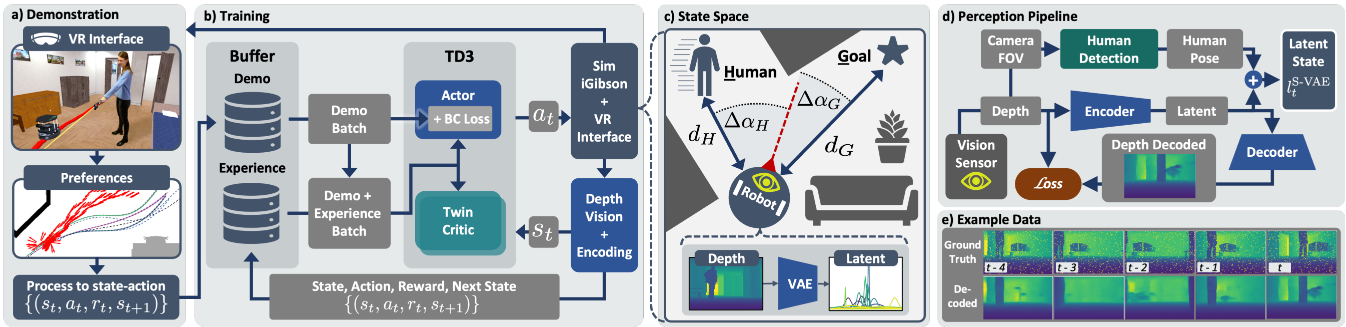

Fig. 1. Schematic of our architecture. **a)** Demonstration trajectories are drawn by the user in VR onto the floor using the handheld controller. **b)** Our TD3 RL architecture with an additional demonstration buffer and behavioral cloning (BC) loss on the actor trains a personalized navigation policy that outputs linear and angular velocities. **c)** The robot-centric state space relies on a depth vision perception pipeline plus explicit relative human and goal position. A variational autoencoder (VAE) compresses the raw images to a latent state representation.**d)** Visualization of the VAE with ground truth depth data before encoding (top) and the decoder's reconstruction (bottom).

We assume the robot to be provided a local goal from a global planner. To achieve user-preferred and collision-free navigation behavior, the robot relies on a depth vision camera to sense human and obstacles. We formulate personalized navigation as a learning task from VR demonstrations of the user, where the robot learns a controller outputting linear and angular velocity.

A. Reinforcement Learning Robot Navigation with Demonstrations

The learning approach is a hybrid of reinforcement learning and behavior cloning, see. Fig. 1a+b.

In short, we enroll an off-policy twin-delayed deep deterministic policy gradient (TD3) reinforcement learning architecture with an additional behavioral cloning (BC) loss on the actor, similar to Nair *et al.* [11]. Furthermore, a separate buffer for demonstration data is introduced.

The robot-centric **state space** consists of three main parts, compare Fig. 1c: 1) The relative goal position $(d_G, \Delta\alpha_G)$, 2) the human position $(d_H, \Delta\alpha_H)$ and presence $k_H \in \{0, 1\}$ in the robot's FOV, and 3) the latent representation of the depth data. The human's FOV-presence and position are obtained from simulation directly. When no human is observed in the FOV $d_H^* = -1$ m and $\Delta\alpha_H^* = 0$ rad.

TD3's continuous **action space** ensures smooth robot control, as the actor network outputs forward and angular velocity $(v \in [0, 0.5], \omega \in [-\pi, +\pi]\ \mathrm{rad\,s^{-1}})$.

The **reward** design features penalties for collision $(r_{\text{collision}} = -\frac{1}{2}c_{\text{rew}})$ and timeout $(r_{\text{timeout}} = -\frac{c_{\text{rew}}}{4})$, and rewards goal reaching $(r_{\text{goal}} = +\frac{c_{\text{rew}}}{2})$, where $c_{\text{rew}} = 10$. Special rewards apply for demonstration states, where the goal reaching reward is increased to $r_{\text{goal}} = +c_{\text{rew}}$ and an additional $+\frac{c_{\text{rew}}}{100}$ is added to each demonstration state . The explicitly higher reward of the demonstration data to boost the value of demonstration-like behavior for the critics during learning. In short, a higher value of demonstration-like behavior encourages user-preference-like navigation whenever possible, while preventing the agent from taking more efficient, shorter trajectories.

To teach and train our navigation controller in a realistic environment for the Kobuki Turtlebot 2 robot, we use the iGibson **simulator** [14] that provides a set of interactive indoor scenes and a VR interface that we used for immersive demonstration. iGibson renders the robot's forward facing depth-camera with a 87° horizontal FOV specified, which serve as input to our perception pipeline during training. Generally, our approach is applicable to other robots with similar control modalities.

To collect preference-reflecting **demonstration data**, the user demonstrates a trajectory for the robot by drawing it onto the floor using the beam-emitting handheld controller, see Fig. 1a. For this study, we recorded dynamic and static navigation scenarios by ourselves. The dataset contains nine scene configurations, with around three demonstration trajectories each.

During **training** in the randomized iGibson scenes, start and goal location of the robot are randomly sampled in the same room, while ensuring a goal distance with $1.5\ \mathrm{m} < d_G < 6\ \mathrm{m}$, equivalent to the depth sensing range.

To simulate the human, four different behaviors modes are sampled: 1) Human walks in the opposite direction from the robot's goal to its start on an A* path. 2) Random human start and goal location. 3) The Human is static. 4) No human in scene. 5) Human moves according to recorded demonstrations. For modes 1+2, the human speed is sampled from a standard distribution $\mathcal{N}(\mu = 0.5\ \mathrm{m\,s^{-1}}, \sigma = 0.3\ \mathrm{m\,s^{-1}})$.

B. Representation Learning

Reinforcement learning on raw high-dimensional depth vision data is unfeasible. Ideally, a dimensionality-reduced state representation is used [13]. Thus, we compress the depth data to a latent representation l using a β-variational autoencoder (VAE) with six relu-activated convolutional layers, see Fig. 1c-d. The dimensionality reduction is factor 320 from a 128 x 80 pixel depth image to a latent space of dimensionality 32. To make the model robust against sensor noise that a depth camera would exhibit, we apply a $5\ \%$ dropout noise to the depth frames during VAE training. The VAE learns to filter the noise, as the VAE's reconstruction loss is computed between the decoded and the noise-free depth-frame. A visualization of the VAE's performance is depicted in Fig. 1e.

To train the autoencoder, we generated an extensive dataset of depth-frames in the iGibson simulator [14] according to the training environment, using a simple obstacle avoidance

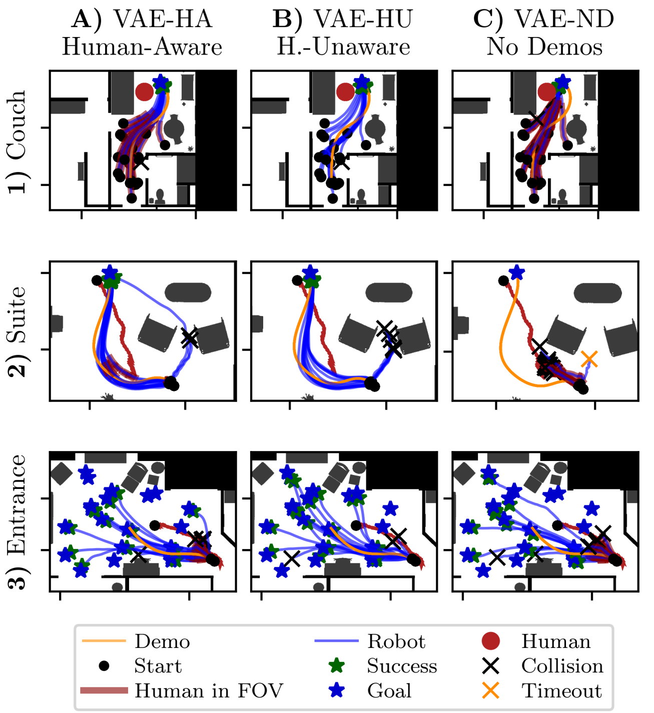

Fig. 2. The robot's learned navigation behavior (blue lines) for scenes with demonstrated preferences (orange line) (**rows 1-3**) and various controller configurations (**columns A-C**) are depicted. The human (red) is either static (red circle) or moving through the scene (red arrow). Goal (blue star) and start location (black dot) are either taken from the demonstration or sampled in the room. In short, the VAE-HA approach (A) exhibits navigation behavior which is most reliable and closest to the demonstrated preferences.

controller trained with TD3 RL. During training of the RL agent, the autoencoder model is frozen.

IV. Experimental Evaluation

This section highlights the performance of our learned preference-reflecting navigation controller under different configurations. A qualitative analysis in Sec. IV-B discusses the navigation behavior on for selected scenes. This is followed by a quantitative analysis targeting the robustness with success metrics in Sec. IV-C.

A. Perception Pipeline Configurations

We first evaluate different perception pipeline and learning configurations against each other, compare Fig. 2.A-D and Fig. 4.A-C. Their key differences lie in the state space as input to the RL policy.

The standard **h**uman-**a**ware VAE-HA (Fig. 2A) state space configuration S-VAE contains the current latent depth encoding, goal position, the human presence binary and human position: $s_t^{\text{VAE-HA}} = (l_t, d_G, \Delta\alpha_G, k_H^t, d_H^t, \Delta\alpha_H^t)$.

The **h**uman-**u**naware VAE-HU (Fig. 2B) is the same controller as the VAE-HA, but the human detection in the robot's field of view is disabled during evaluation.

The **n**o-**d**emonstration VAE-ND controller does not rely on the learning architecture as shown in Fig. 1. It has neither a demonstration buffer, nor a behavioral cloning loss, making it a standard TD3 architecture. Therefore, it has learned its navigation behavior without user demonstrations.

Our ablation study introduces two more configurations, see Sec. IV-D: VAE-FOV-120 implements a widened FOV at 120° over the standard 87°, as it can be found on wide-angle depth cameras such as the Microsoft Azure Kinect. VAE-NG discards the goal distance d_G from the state space.

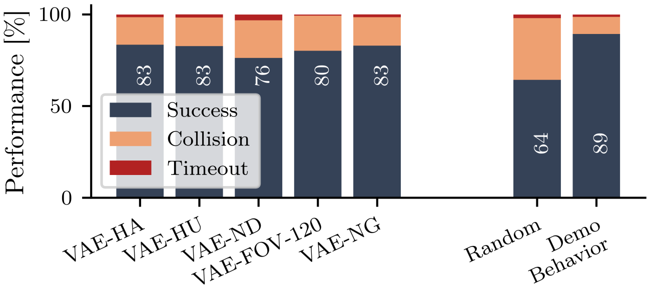

Fig. 3. The performance of the different controllers is averaged over all demonstration scenarios and other scenes. For each combination of scene, human behavior modes, and demonstration preferences (if available), 50 trajectories were generated. "Random behavior" refers to behavior modes 1-4, while "demo behavior" refers to mode 5, both evaluated with controller VAE-HA.

B. Qualitative Navigation Analysis

Fig. 2 shows the learned navigation behavior of our controller and highlights resulting differences between the perception pipeline configurations introduced above.

In **Fig. 2.1**, the human is static and located at the couch. The robot's start location is randomized, while keeping the goal at the end of the demonstration trajectory. As the robot traverses the living room, it shall navigate on the opposite side of the room close to the dining table and along the cupboard. With VAE-HA, the robot learned to navigate closely to the demonstrated preference. It exhibits a similar, smooth, S-shaped curve while passing by the couch. Interestingly, a pronounced difference in the robot's trajectory shape can be observed between VAE-HA and VAE-HU (Fig. 2.A1+B1). As the human is not explicitly observed in the state space, VAE-HU's approaching behavior to the human rather resembles shortest-path trajectories, while cutting short on the demonstrated S-shaped curve. Note that the robot trajectories are shaded in red in Fig. 2, whenever the human is observed on the FOV.

In **Fig. 2.2**, the moving human encounters the robot with an opposite direction of travel at the living room's suite. As a preference, the robot should take a wide turn of avoidance around the armchair to make space for the approaching human. Among all controllers except VAE-HA, the navigation of the situation is challenging, leading to collisions around the armchair's corner.

As the human walks out of the room in **Fig. 2.3**, the robot enters. When the robot detects the approaching human, it shall take a left turn and make room for the human to pass. Afterwards, the robot can continue traversing the living room to its goal. In this scenario the effect of demonstration trajectories strikes: The VAE-ND controller without access to demonstrations mostly exhibits direct goal-oriented, straight-path navigation.

Qualitatively, the VAE-HA configuration results in the best-performing personalized robot navigation controller.

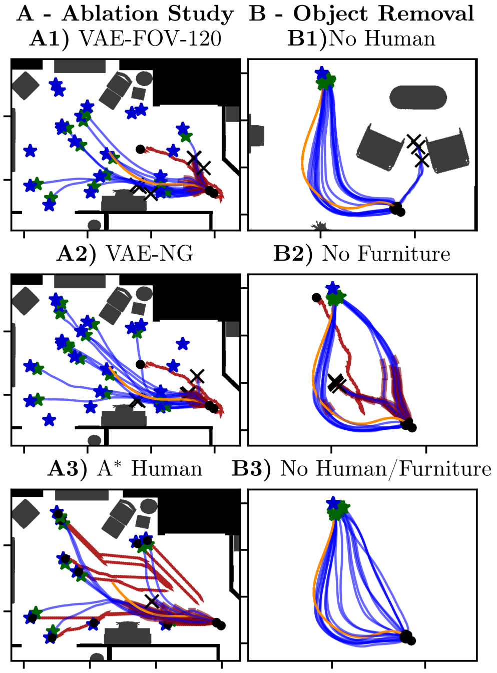

Fig. 4. **A)** In our ablation study, we investigate (**A1**) the effect of increased camera field of view with VAE-FOV-120, (**A2**) the removal of the goal distance from the state space VAE-NG in comparison to the original approach (Fig. 2.**A3**). Also, a differently moving human is implemented (**A3**). **B)** To identify relevant environment features for the agent, the human (**B1**), the furniture (**B2**), or both (**B3**) were removed from the scene, compared to the original setup (Fig. 2.**A2**). For a legend, please refer to Fig. 2.

C. Quantitative Analysis: Robustness

Fig. 3 shows the performance of different controller setups and human behavior modes (see Sec. III-A) in terms of success, collision, and timeout. We determine the demonstration-aware VAE architectures (VAE-HA, -HU, -NG) most capable of avoiding collisions. The VAE-ND controller without demonstration access perform worse than the demonstration-based VAE architectures. Regarding different human behavior sampling modes (Sec. III-A), as expected the demonstration-related mode 5 perform best. Increasing the RGB-D camera's FOV (VAE-FOV-120), e.g., for better perception of pedestrians approaching from the side, does not lead to better collision avoidance. More observed collisions than timeouts could be a consequence of the agent being encouraged to drive by the BC loss from demonstration data.

D. Ablation Study

Finally, we perform an ablation study. Investigate effects of an increased camera field of view (Fig. 4.A1), VAE-FOV-120 rather deteriorates the collision avoidance capabilities. This is in line with the obtained overall performance results, see Fig. 3. Removing the goal distance from the state space (Fig. 4.A2) (VAE-NG) interestingly does not deteriorate the performance, but also results in robust and preference-reflecting navigation.

Demonstrating the ability for generalization, in Fig. 4.A3 we showcase a scenario where humans follows an A* path in the opposite direction to the robot (compare behavior mode 1 in Sec. III-A). In most cases, the robot intuitively gives way to the approaching human.

To learn which environment features the agent uses for navigation and preference reproduction, we removed either the human, furniture, or both from the scene, see Fig. 4.B. Interestingly, as no human approaches from behind the armchair (Fig. 4.B1), the robot navigates closer to the chair with similar trajectory shape. As all furniture is removed from the scene (Fig. 4.B2), the robot either exhibits preference navigation or a shorter path on the other side of the approaching human. With everything removed (Fig. 4.B3), the small deviation around the human collapses to a shortest path. But still the robot is able to reflect preferences. We attribute this behavior to the perception of walls and room layout that are still observable for the robot, or a learned guidance by relative goal position in the state space.

V. Conclusion

To summarize, we presented a learning approach to personalized navigation based on depth vision. A VAE compresses the perceived 3D scene to an efficient latent representation used as input by the learning framework. As demonstrated with our results, we successfully learned a personalized navigation controller that reflects user preferences from few VR demonstrations in dynamic human-robot navigation scenarios. We furthermore find the inclusion of demonstrations to improve the overall navigation performance in terms of success rate. In conclusion, our research has demonstrated the feasibility of personalized robot navigation utilizing depth vision sensors and presents a promising avenue for the development of more user-oriented robot controllers.

References

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

[2] J. de Heuvel, N. Corral, L. Bruckschen, and M. Bennewitz, "Learning Personalized Human-Aware Robot Navigation Using Virtual Reality Demonstrations from a User Study," in *2022 31th IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*, pp. 898–905.

[3] 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, p. 103837.

[4] C. Theodoridou, D. Antonopoulos, A. Kargakos, I. Kostavelis, D. Giakoumis, and D. Tzovaras, "Robot Navigation in Human Populated Unknown Environments based on Visual-Laser Sensor Fusion," in *The15th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, pp. 336–342.

[5] M. Laskin, A. Srinivas, and P. Abbeel, "CURL: Contrastive Unsupervised Representations for Reinforcement Learning," in *Proceedings of the 37th International Conference on Machine Learning*. PMLR, pp. 5639–5650.

[6] 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, pp. 11 025–11 031.

[7] X. Gao, X. Zhao, and M. Tan, "Modeling Socially Normative Navigation Behaviors from Demonstrations with Inverse Reinforcement Learning," in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, pp. 1333–1340.

[8] 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, pp. 4423–4430, Oct. 2018.

[9] H. Karnan, A. Nair, X. Xiao, G. Warnell, S. Pirk, A. Toshev, J. Hart, J. Biswas, and P. Stone, "Socially Compliant Navigation Dataset (SCAND): A Large-Scale Dataset of Demonstrations for Social Navigation." [Online]. Available: http://arxiv.org/abs/2203.15041

[10] 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)*, pp. 6015–6022.

[11] 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, pp. 6292–6299.

[12] 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, pp. 1111–1117.

[13] D. Hoeller, L. Wellhausen, F. Farshidian, and M. Hutter, “Learning a State Representation and Navigation in Cluttered and Dynamic Environments,” *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5081–5088.

[14] B. Shen, F. Xia, C. Li, R. Martín-Martín, L. Fan, G. Wang, C. Pérez-D’Arpino, S. Buch, S. Srivastava, L. Tchapmi, M. Tchapmi, K. Vainio, J. Wong, L. Fei-Fei, and S. Savarese, “iGibson 1.0: A Simulation Environment for Interactive Tasks in Large Realistic Scenes,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7520–7527.