SACSoN: Scalable Autonomous Control for Social Navigation

Noriaki Hirose[1,2], Dhruv Shah[1], Ajay Sridhar[1] and Sergey Levine[1]

Abstract— **Machine learning provides a powerful tool for building socially compliant robotic systems that go beyond simple predictive models of human behavior. By observing and understanding human interactions from past experiences, learning can enable effective social navigation behaviors directly from data. In this paper, our goal is to develop methods for training policies for socially unobtrusive navigation, such that robots can navigate among humans in ways that don't disturb human behavior. We introduce a definition for such behavior based on the *counterfactual* perturbation of the human: if the robot had not intruded into the space, would the human have acted in the same way? By minimizing this counterfactual perturbation, we can induce robots to behave in ways that do not alter the natural behavior of humans in the shared space. Instantiating this principle requires training policies to minimize their effect on human behavior, and this in turn requires data that allows us to model the behavior of humans in the presence of robots. Therefore, our approach is based on two key contributions. First, we collect a large dataset where an indoor mobile robot interacts with human bystanders. Second, we utilize this dataset to train policies that minimize counterfactual perturbation. We provide supplementary videos and make publicly available the largest-of-its-kind visual navigation dataset on our project page[1].**

I. Introduction

Even the simplest forms of interaction between humans, such as how to pass someone in a hallway, are governed by complex non-verbal cues, and may be challenging to script. In order for robots to inhabit the same environments as people, they must also be cognizant of basic social cues and etiquette, even for seemingly simple navigational tasks. While a range of prior works have proposed approaches for modeling human behavior [1, 2], the complexity of such interactions often defies analytic modeling techniques.

We approach this challenge from a data-driven perspective: acquiring policies for navigation around humans by leveraging data of human-robot interactions to *learn* how to navigate in socially unobtrusive ways. We propose a definition for such behavior, which is based on the *counterfactual* perturbation of humans. Specifically, we consider whether humans would have acted in the same way if the robot had not intruded into their space. By minimizing this counterfactual perturbation, we can guide robots to behave in a manner that does not alter the natural behavior of humans in the shared space. To instantiate this principle, we train the SACSoN (**S**calable **A**utonomous **C**ontrol for **So**cial **N**avigation) policy to minimize the impact on human behavior. This requires us to both formalize the notion of counterfactual perturbation into an objective, and to collect a dataset that has the kinds of human-robot interactions that can allow our model to learn to predict human behavior in the presence of robots. Thus, our work focuses on two complementary technical components: the design of a policy learning method that can utilize predictive models of humans for unobtrusive navigation, and the collection of a large dataset of human-robot interactions to train these predictive models.

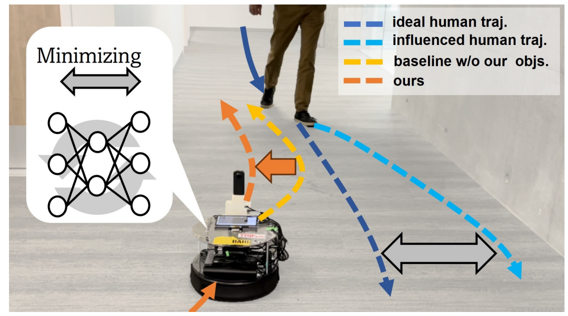

Fig. 1: **SACSoN is a socially unobtrusive vision-based navigation policy in the human-occupied spaces.** We penalize counterfactual perturbations (gray) from the intended human trajectory (navy) and generate the compliant commands (orange).

To collect such a dataset, we propose a data collection system, which we call HuRoN (**Hu**man-**Ro**bot interaction data collection for vision-based **N**avigation) system. In contrast to previous social navigation datasets that involve expensive manual tele-operation [3, 4], or simple scripted policies that fail to capture data diversity [5]. Instead, we devise an intelligent system that can *autonomously* collect rich interaction data with little-to-no human intervention, and can improve its data collection policy over time as the ever-growing dataset is reused to further train our policy. Due to the page limitation, the details of HuRoN system is shown in our project page.

Our work makes the following contributions: (i) a method for learning a socially compliant SACSoN policy for visual navigation around humans, (ii) an autonomous data collection system, HuRoN, that encourages rich interactions with human pedestrians using a novel training objective, and (iii) the HuRoN dataset, a large and diverse dataset comprising over 4000 human-robot interactions of an autonomous robot operating in a densely populated office-space environment. Please see the project page for the dataset and videos.

II. Related Work

Social navigation has been widely studied in the literature [6–8]. Model-based approaches based on the dynamic pedestrian model have clasically been applied for behavior modeling [1, 2, 9]. These methods determine the robot's actions in a virtual space with the predicted pedestrians' behavior [10–17], considering social momentum [14], a maximum entropy model [13], a model predictive controller [15], or a classical planner [16, 17]. Social navigation has also

[1]UC Berkeley, [2]Toyota Motor North America

[1] sites.google.com/view/SACSoN-review

been viewed through the lens of model-free data-driven learning such as reinforcement learning [18–22].

Our method using the pedestrians' predictive model belongs to the former. However, different from prior works, we apply the predictive model to estimate the counterfactual perturbation from the intended pedestrians' trajectory and train the control policy by penalizing the perturbation on offline. Hence our control policy enables the robot to navigate to the target position while allowing the pedestrian to walk as intended. Moreover, since our approach is end-to-end learning, the robot actions can be derived from raw images without detecting and predicting pedestrians in inference.

III. Preliminaries

We propose a method and dataset for social compliant robotic navigation with a learning-based approach. The design of our method extends ExAug [23], a control policy for vision-based navigation that optimizes a goal-directed cost function (but does not by itself consider interaction with humans). This system can navigate to user-specified goal images using a combination of a topological graph and a learned low-level control policy, and its design is related to a number of recent works on vision-based navigation with learned policies and topological maps [23–28]. We build our data collection system, HuRoN, on top of the same visual navigation system.

The control policy in ExAug predicts control velocities $\{v_i, \omega_i\}_{i=1 \ldots N_s} = \pi^c_\phi(I_t, I_g)$ from the current image I_t and subgoal image I_g, and commands the linear velocity v_1 and the angular velocity ω_1 to the robot to reach the position of I_g, similar to receding horizon control. Here, N_s is the control horizon and t is the current step number. We commonly show the learnable parameters (e.g., ϕ) as a subscript on the model function (e.g., π^c_ϕ). The control policy is paired with a topological memory that contains images as nodes and temporal distance between them as the edges. The ExAug control policy π^c_ϕ is trained to minimize the objective

$$J_{\text{nav}}(\phi) := J_{\text{pose}}(\phi) + w_c J_{\text{col}}(\phi) + w_r J_{\text{reg}}(\phi), \qquad (1)$$

where J_{pose} corresponds to the prediction error in the relative pose estimates, J_{col} penalizes collisions, and J_{reg} is a regularization term for predicted velocities. ExAug uses a geometric and kinematics model to estimate the relevant states of the robot in a virtual space and calculate these objectives, akin to the model predictive control. These objectives enable us to train the policy by minimizing the differentiable cost J_{nav} without imitating the ground truth values. Please refer to the original paper for implementation details of this system [23].

Overview: Section IV introduces our method to train the SACSoN policy, which aims to enable robotic navigation among humans with minimal disruption. In addition to J_{nav}, we introduce two new objectives using the counterfactual human trajectories. We pre-train the predictive model of the pedestrians' future trajectory to estimate the counterfactual human trajectories in training.

IV. Learning a Socially Compliant Policy

We posit that a possible way to achieve "social compliance" is for robots to avoid disrupting the *intended behavior*

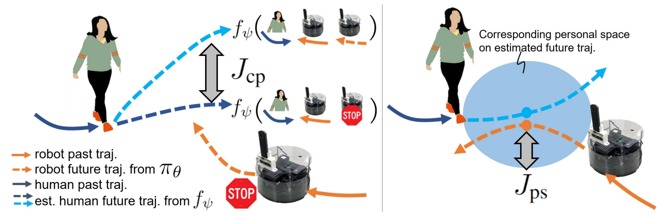

Fig. 2: **Our proposed objectives J_{cp} and J_{ps} for training SACSoN policy.** J_{cp} penalizes the counterfactual perturbation from the estimated intented pedestrian's trajectory (left). J_{ps} penalizes the personal space violation in the future space (right).

of pedestrians, i.e., allow humans to carry on with their activities without disruption. In our proposed method, we penalize the counterfactual perturbation of the intended trajectories of the pedestrians. We define the intended trajectory of a pedestrian as the predicted trajectory of the pedestrian from our predictive model conditioned on the robot being stationary and non-intrusive. Our method aims to control the robot so that the humans in the environment do not act differently than they would have if the robot had been stationary. This principle could be further generalized to minimize the difference to other counterfactual situations, such as ones where the robot is absent all together, but we focus on the stationary robot counterfactual as a simple instantiation of the principle. For safety, the complete design of our full objective function also includes a term to penalize the predicted distance between the human and the robot, to encourage the robot to maintain clearance, as well as the standard navigation terms described in the preceding section. Thus, we add two terms to J_{nav}, forming our full objective:

$$\min_\theta J(\theta) := J_{\text{nav}}(\theta) + w_{cp} J_{\text{cp}}(\theta) + w_{ps} J_{\text{ps}}(\theta), \qquad (2)$$

where J_{cp} is an objective to suppress the counterfactual perturbation (Fig. 2 left) and J_{ps} is an objective to penalize the penetration of the personal space of the pedestrians (Fig. 2 right), where w_{cp} and w_{ps} are weights for each objective. Here, our control policy π_θ predicts velocity commands $\{v_i, \omega_i\}$ from $I_{t:t-N_p}$ and I_g, defined as follows:

$$\{v_i, \omega_i\}_{i=1 \ldots N_s} = \pi_\theta(I_{t:t-N_p}, I_g) \qquad (3)$$

Concatenating the past image frames gives the robot additional context that can be useful to avoid obstacles, detect pedestrians in the environment, and reduce partial observability [29].

J_{cp}: To train the policies without distracting pedestrians, we desgin J_{cp} using counterfactual pedestrians trajectories,

$$J_{\text{cp}}(\theta) = \frac{1}{N_s} \sum_{i=1}^{N_s} (\hat{h}^{gw}_{t+i} - \hat{h}_{t+i})^2, \qquad (4)$$

where $\hat{h}^{gw}_{t+i}$ is the estimated pedestrian's 2D trajectory conditioned on the the robot virtually stopping at the current position to give way and $\hat{h}_{t+i}$ is the estimated pedestrian's 2D trajectory conditioned on the robot future action. By minimizing J_{cp} with the other objectives to train our control policy, the pedestrian can walk a path similar to what they would have taken when the robot stopped and gave way, while allowing the robot itself to move toward the goal position. Here, we estimate $\hat{h}_{t+i}$ as

$$\hat{h}_{t+1:t+\beta} = f_\psi(h_{t-\alpha:t}, r_{t-\alpha:t}, r_{t+1:t+\beta}) \tag{5}$$

where f_ψ is a trained predictive model of a pedestrian's future trajectory, conditioned on their past trajectory $h_{t-\alpha:t}$, as well as the robot's past trajectory $r_{t-\alpha:t}$ and future trajectory $r_{t+1:t+\beta}$. All trajectories in Eqn. 5 are on the current robot coordinate. The values for $r_{t-\alpha:t}$ are obtained from past wheel odometry, and $r_{t+1:t+\beta}$ is derived by integrating the velocity commands $\{v_i, \omega_i\}_{i=1\dots N_s}$ from our control policies. To obtain $h_{t-\alpha:t}$, we use YOLO [30, 31] and DeepSORT [32] to detect and track pedestrians in the images (processed into a panorama) from the recorded observations of the robot [33], and project these detections in 3D using the depth and scale estimates obtained from the ExAug perception module [34].

For the other counterfactual trajectory, we input a zero vector instead of $r_{t+1:t+\beta}$ to estimate $\hat{h}^{gw}_{t+1:t+\beta}$ as $f_\psi(h_{t-\alpha:t}, r_{t-\alpha:t}, \mathbf{0})$. Giving the zeros vector as the robot future trajectory corresponds to stopping at the current pose. Note that we only consider scenes involving a single pedestrian for simplicity; for scenes with multiple pedestrians, we consider the nearest non-stationary pedestrians for training, since they are most likely to interact with the robot. To obtain an accurate predictive model f_ψ, we collect an interaction-enriched dataset using the HuRoN system and train f_ψ before training π_θ.

J_{ps}: We design J_{ps} to encourage the robot to avoid the personal space of the pedestrians.

$$J_{\text{ps}}(\theta) = \min_i\{|r_h + r_r - c(d_i)|\}, \tag{6}$$

where r_h is the personal space, r_r is the robot radius, d_i is the distance on 2D plane between the future pedestrians' position $\hat{h}_{t+i}$ and the future robot position r_{t+i}, and c is the function to limit d_i between 0 and $r_h + r_r$ to penalize the robot trajectories only penetrating the personal space. J_{ps} may be alternatively defined as the mean of the set $\{|r_h + r_r - c(d_i)|\}$, but empirically, we found the min formulation of Eqn. 6 to better capture the desired behavior.

Implementation details: Following ExAug [23], we set the control horizon $N_s = 8$ and the past observations $N_p = 5$ (see Eqn. 3). For pedestrian detection and tracking, we use the spherical camera on the robot to allow detection and interactions with pedestrians behind it. We use a batch size of 80, with the training pair (past observations and subgoal images) sampled from the same trajectory for one half of the batch, and the pair coming from different trajectories in the other half of the batch. We empirically set the weights $w_{cp} = 10.0$, and $w_{ps} = 100.0$ for each objective, after analyzing closed-loop navigation performance.

For training π_θ, we pre-train f_ψ with $\alpha = N_s - 1$ and $\beta = N_s$ by minimizing the MSE loss using supervised learning and frozen f_ψ while training π_θ. We calculate the gradient of π_θ by back-propagation via f_ψ for updating π_θ. To be more accurate predictive model, we generate the human and robot trajectories by social force model [1] and mix them with our real data in the batch. One half of the batch is from

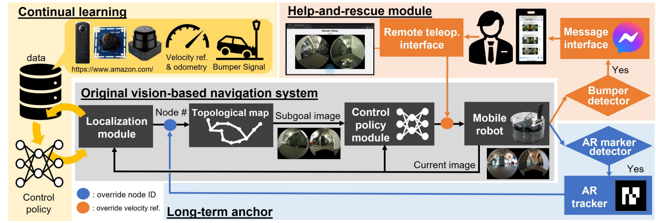

Fig. 3: **HuRoN System overview.** We design our autonomous data collection platform around a vision-based navigation system (gray) that uses a topological graph and a learned control policy. Our proposed system has three key components: a help-and-rescue module for collision recovery (orange), long-term anchors for localization (blue), and continual learning (yellow).

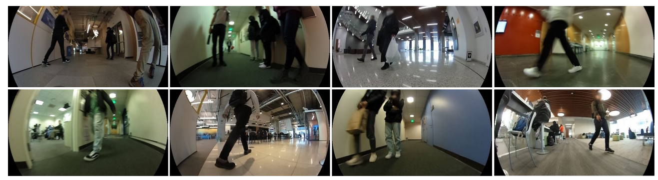

Fig. 4: **Example scenes from the HuRoN Dataset.** We collected our dataset in 5 different environments, spanning over 75 hours of data collection and 4000 rich human interactions, containing raw visual observations (cropped spherical images shown here).

our real dataset and the other half of the batch is from the social force model. Please see our supplemental materials for more information on the simulation data. Following [35], we set the personal space r_h as 0.45 and the robot radius r_r as 0.25 including a small margin. All other hyperparameters are replicated from ExAug [23].

V. Autonomous Data Collection System

For our counterfactual objective to effectively supervise the robot's policy, we rely on the predictive model f_ψ to make accurate predictions about hypothetical human-robot interactions. This requires training f_ψ on a diverse dataset that contains many interactions between pedestrians and our robot. Therefore, the second major contribution of our work is an autonomous data collection system that can collect such a dataset. During collection, we wish to *maximize* interactions between the robot and pedestrians, while also maintaining autonomy, to collect high-quality data.

We propose the HuRoN system (Fig. 3) with the data collection policy π_ρ to autonomously collect over 75 hours of robot navigation data in 5 diverse human-occupied environments, capturing over 4000 rich interactions with humans. Due to the page limitation, We describes the key characteristics of the collected dataset. Details of the HuRoN system are shown in our project page.

Dataset Characteristics: We collected the HuRoN dataset over the course of 24 days in 5 diverse environments, spread across 3 university buildings. The dataset spans 75 hours and 58 kilometers of autonomous robot navigation trajectories, containing over 4000 interactions with humans. The dataset includes visual observations (spherical and fisheye), 2D LiDAR scans, velocity information, and collision signals from the bumper. Figure 4 shows example images of rich human-robot interactions captured in our dataset.

To evaluate the efficacy of the proposed data collection system, our dataset contains two equal subsets: the *interaction-enriched dataset* corresponding to data collected by our collection policy π_ρ, and the *naïve dataset* collected without taking interaction. We have released this dataset publicly on our project page.

VI. Evaluation

We design our experiments to answer for the questions, **Q1** Does our proposed objectives J_{cp} and J_{ps} lead to better socially unobtrusive behavior? and **Q2** Does our proposed dataset lead to better control performance? The other detail evaluations for our dataset are shown in our project page.

Towards answering **Q1** and **Q2**, we train two different policies with and without our proposed objectives J_{cp} and J_{ps}. Here, the control policy without J_{cp} and J_{ps} corresponds to the most relevant baseline method, ExAug [23]. In addition, we train different social navigation policy on the naïve dataset without enriched human-robot interactions. We conduct fifteen experiments using the real robot in three different real environments. The distance between the start and goal positions ranges from 13.0 to 37.8 meters, which is considered relatively long for vision-based navigation in indoor settings. In order to ensure equivalent experimental conditions, we request during the evaluation that the pedestrians navigate around the robot, creating similar interaction scenarios for each control policy. If the robot collides with a pedestrian or obstacle, we request the pedestrian to distance themselves from the robot's perimeter, and we allow the the robot to continue navigation.

Table I presents the comparison of our method to the above baselines along several metrics: Goal arrival Rate (GR), Success weighted by Path Length (SPL) [36], Success weighted by Time Length (STL) [37], Collision count for Pedestrians (CP), Collision count for static Objects (CO), and Personal Space Violation duration (PSV). Our control policy trained on our proposed dataset with J_{cp} and J_{ps} shows a clear improvement over ExAug. In particular, our method decreases the collision counts for pedestrians by more than 80%, reduces PSV by over 30%, and successfully leads the robot to the goal position. The comparison suggests that our proposed objective improves the robot's ability to navigate unobtrusively in the presence of humans, and our proposed dataset collected via an interaction-seeking policy leads to better performance for our method.

In Fig. 5, we qualitatively observe the robot's behavior to be significantly more "compliant" when trained with the interaction-enriched dataset (left). Even in the narrow corridors, our control policy makes space for the pedestrians while still maintaining clearance from the walls. The control policy trained on the naïve dataset does not take avoidance action when a pedestrian approaches the robot, so the robot often violates personal space, collides with the pedestrian (top right), or fails to reach the goal (bottom right).

VII. Discussion

In this paper, we proposed a method for training the SACSoN policy for vision-based navigation to build the socially unobstrusive navigation system. In training SACSoN policy, we introduced novel objectives using the predictive model of the pedestrians' future trajectories to suppress the counterfactual perturbation from the intended human trajctories. To obtain an accurate predictive model for a better SACSoN policy, we proposed the HuRoN system, a scalable data collection system, to autonomously collect a dataset with enriched human-robot interactions. We used this data collection system to collect the HuRoN dataset: the largest-of-its-kind publicly available dataset of visual navigation around humans, spanning over 75 hours of data collected in 5 different environments and comprising over 4000 rich human-robot interactions. Our experiments show that policies trained on the collected dataset enables the real robot to navigate with the socially unobtrusive behavior.

Our SACSoN policy trained on the dataset with enriched human-robot interactions does have some limitations. Our autonomous robot is restricted to operate at slow speeds (capped at 0.4 m/s) to limit damage due to policy errors in safety-critical environments with humans; however, this inherently limits the robot behavior. When a pedestrian is approaching the robot at a casual pace, even if the robot takes actions to avoid the pedestrian, it may not be fast enough. Additionally, our current system only learns simple social interactions such as avoiding a pedestrian's personal space and giving way to the pedestrians; learning richer social interactions will need better objectives incorporated more tightly in the data collection and deployment policies.

Method	Training dataset	GR ↑	SPL ↑	STL ↑	CP ↓ [#]	CO ↓ [#]	PSV [s] ↓
ExAug [23]	int.-enriched (ours)	0.800	0.692	0.595	20	6	85.248
Ours	naïve (baseline)	0.667	0.517	0.365	8	11	84.915
Ours	int.-enriched (ours)	**1.000**	**0.888**	**0.692**	**1**	**2**	**57.609**

TABLE I: **Closed-loop Evaluation of trained control policies.** We find a policy trained with the interaction-enriched dataset results in less collisions and personal space violations for social navigation and enables the robot to reach the goal position at higher accuracy.

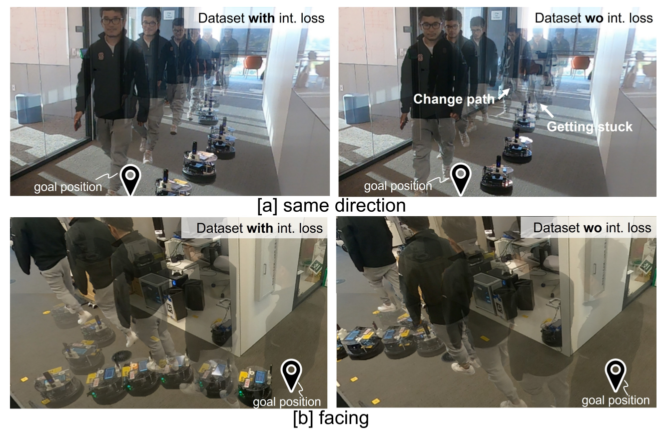

Fig. 5: **Qualitative Examples of Learned Behavior.** A social navigation policy trained on the interaction-enriched subset of HuRoN (left) leads to better handling of human pedestrians while successfully reaching the goal, without intruding in their personal space. Training on the naïve dataset results in a conservative policy (right) that gets stuck and collides with pedestrian.

REFERENCES

[1] D. Helbing *et al.*, "Social force model for pedestrian dynamics," *Physical review E*, 1995. 1, 3

[2] G. Ferrer *et al.*, "Robot companion: A social-force based approach with human awareness-navigation in crowded environments," in *IROS*. IEEE, 2013, pp. 1688–1694. 1

[3] R. Martin-Martin *et al.*, "Jrdb: A dataset and benchmark of egocentric robot visual perception of humans in built environments," *IEEE transactions on pattern analysis and machine intelligence*, 2021. 1

[4] H. Karnan *et al.*, "Socially compliant navigation dataset: A large-scale dataset of demonstrations for social navigation," *RA-L*, 2022. 1

[5] A. Rudenko *et al.*, "Thör: Human-robot navigation data collection and accurate motion trajectories dataset," *RA-L*, vol. 5, no. 2, 2020. 1

[6] C. Mavrogiannis *et al.*, "Core challenges of social robot navigation: A survey," *arXiv preprint*, 2021. 1

[7] E. A. Sisbot *et al.*, "A human aware mobile robot motion planner," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 874–883, 2007.

[8] J. Mumm and B. Mutlu, "Human-robot proxemics: physical and psychological distancing in human-robot interaction," in *HRI*, 2011. 1

[9] D. Mehta *et al.*, "Autonomous navigation in dynamic social environments using multi-policy decision making," in *IROS*, 2016. 1

[10] M. Luber *et al.*, "Socially-aware robot navigation: A learning approach," in *IROS*. IEEE, 2012, pp. 902–907. 1

[11] B. D. Ziebart *et al.*, "Planning-based prediction for pedestrians," in *IROS*. IEEE, 2009.

[12] A. Bajcsy *et al.*, "A scalable framework for real-time multi-robot, multi-human collision avoidance," in *ICRA*. IEEE, 2019.

[13] M. Pfeiffer *et al.*, "Predicting actions to act predictably: Cooperative partial motion planning with maximum entropy models," in *IROS*. IEEE, 2016. 1

[14] C. I. Mavrogiannis *et al.*, "Social momentum: A framework for legible navigation in dynamic multi-agent environments," in *HRI*, 2018. 1

[15] X. Xiao *et al.*, "Learning model predictive controllers with real-time attention for real-world navigation," *arXiv preprint*, 2022. 1

[16] X.-T. Truong and T.-D. Ngo, ""to approach humans?": A unified framework for approaching pose prediction and socially aware robot navigation," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 3, pp. 557–572, 2017. 1

[17] H.-Y. Chen *et al.*, "Social crowd navigation of a mobile robot based on human trajectory prediction and hybrid sensing," *AURO*, 2023. 1

[18] Y. F. Chen *et al.*, "Socially aware motion planning with deep reinforcement learning," in *IROS*. IEEE, 2017. 2

[19] ——, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in *ICRA*. IEEE, 2017.

[20] M. Everett *et al.*, "Motion planning among dynamic, decision-making agents with deep reinforcement learning," in *IROS*. IEEE, 2018.

[21] C. Chen *et al.*, "Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning," in *ICRA*, 2019.

[22] Y.-J. Mun *et al.*, "Occlusion-aware crowd navigation using people as sensors," *arXiv preprint*, 2022. 2

[23] N. Hirose *et al.*, "ExAug: Robot-Conditioned Navigation Policies via Geometric Experience Augmentation," in *ICRA*, 2023. 2, 3, 4

[24] N. Savinov *et al.*, "Semi-parametric topological memory for navigation," in *ICLR*, 2018.

[25] X. Meng *et al.*, "Scaling local control to large-scale topological navigation," in *ICRA*. IEEE, 2020.

[26] D. Shah *et al.*, "Viking: Vision-based kilometer-scale navigation with geographic hints," *RSS*, 2022.

[27] N. Hirose *et al.*, "Deep visual mpc-policy learning for navigation," *RA-L*, 2019.

[28] N. Kim *et al.*, "Topological Semantic Graph Memory for Image Goal Navigation," in *CoRL*, 2022. 2

[29] D. Shah *et al.*, "GNM: A General Navigation Model to Drive Any Robot," in *ICRA*, 2023. 2

[30] J. Redmon *et al.*, "You only look once: Unified, real-time object detection," in *CVPR*, 2016. 3

[31] "Object detection by yolov5," https://github.com/ultralytics/yolov5. 3

[32] N. Wojke *et al.*, "Simple online and realtime tracking with a deep association metric," in *ICIP*. IEEE, 2017. 3

[33] "Pedestrian detection and tracking by yolov5 and deepsort," https://github.com/HowieMa/DeepSORT_YOLOv5_Pytorch. 3

[34] N. Hirose *et al.*, "Depth360: Self-supervised learning for monocular depth estimation using learnable camera distortion model," in *IROS*. IEEE, 2022, pp. 317–324. 3

[35] N. Tsoi *et al.*, "An approach to deploy interactive robotic simulators on the web for hri experiments: Results in social robot navigation," in *IROS*, 2021. 3

[36] P. Anderson *et al.*, "On evaluation of embodied navigation agents," *arXiv preprint arXiv:1807.06757*, 2018. 4

[37] A. Francis *et al.*, "Principles and guidelines for evaluating social robot navigation algorithms," *arXiv preprint*, 2023. 4