

Following Is All You Need: Robot Crowd Navigation Using People As Planners

Yuwen Liao^{1*}, Xinhang Xu^{1*}, Ruofei Bai¹, Yizhuo Yang¹,
Muqing Cao², Shenghai Yuan¹, Lihua Xie¹, *Fellow, IEEE*

Abstract—Navigating in crowded environments requires the robot to be equipped with high-level reasoning and planning techniques. Existing works focus on developing complex and heavyweight planners while ignoring the role of human intelligence. Since humans are highly capable agents who are also widely available in a crowd navigation setting, we propose an alternative scheme where the robot utilises people as planners to benefit from their effective planning decisions and social behaviours. Through a set of rule-based evaluations, we identify suitable human leaders who exhibit the potential to guide the robot towards its goal. Using a simple base planner, the robot follows the selected leader through short-horizon subgoals that are designed to be straightforward to achieve. We demonstrate through both simulated and real-world experiments that our novel framework generates safe and efficient robot plans compared to existing planners. Our method also brings human-like robot behaviours without explicitly defining traffic rules and social norms. Code will be available at <https://github.com/centiLinda/PeopleAsPlanner.git>.

I. INTRODUCTION

Crowd navigation is a challenging problem in robotics research as it involves not only static environmental obstacles, but also dynamic agents such as humans. To handle such complex scenarios, recent research aims to develop intelligent and sophisticated robotic systems using techniques such as predictive planning and reinforcement learning [1]. On the other hand, humans, perhaps the most intelligent agents in the scene, are solving similar navigation problems as the robot. If we think outside the box, can the robot take advantage of human intelligence to simplify its task?

Inspired by how children follow their parents when walking in crowded streets, we find that following trustworthy leaders is an alternative to planning independently, as illustrated in Fig. 1. We can decompose the crowd navigation problem into high-level decisions (E.g. how to cut through a crowd, whether to overtake the person in front) and low-level short-horizon movements. The former can be solved using the aforementioned intelligent systems or humans, while the latter only requires simple local planners. Therefore, we propose to rethink human-robot relationships in crowd navigation scenarios by handing over high-level decisions to humans. Apart from being dangerous obstacles that the robot needs to avoid, previous works have explored using people as

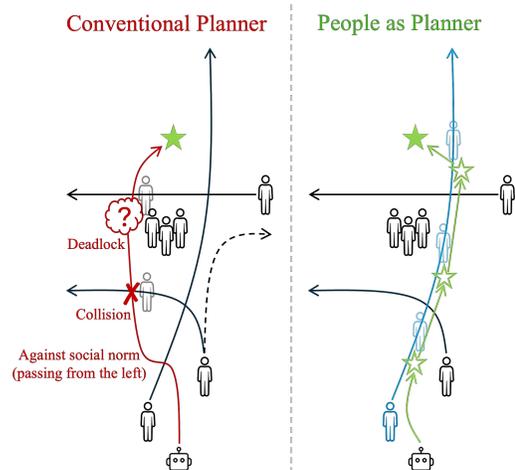


Fig. 1: Instead of developing more sophisticated planners, we propose an alternative scheme where the robot follows suitable human leaders to simplify the problem into straightforward subgoal planning.

sensors [2]–[5] to extend and enrich robot perception. In this work, we use people as intelligent planners that can help the robot solve the more difficult part of the navigation problem.

Through a set of rule-based criteria, we aim to identify suitable human leaders and find a subgoal such that it allows the robot to follow the selected leader closely with minimal interruption from other humans. The subgoals are designed to be easily achievable using a simple base planner and we choose Social Force (SF) [6] in our experiments. The leader selection process is performed continuously so the robot can switch between different humans for maximized efficiency.

Of relevance to our work is the literature on human-following robots [7]–[12]. In these works, a specific user is predefined for the robot to follow without additional context or objectives. Our work is fundamentally different as 1) the robot *actively* chooses a suitable leader to follow, 2) by following humans, we effectively address the challenge of crowd navigation.

Our contributions are threefold. First, we propose a novel crowd navigation framework that utilises human intelligence to decompose the original problem into straightforward subgoal planning. Second, we design a set of evaluation processes to find subgoals near suitable human leaders, which can be effectively reached using a simple SF planner. Third, we demonstrate through experiments that the proposed People-as-Planner scheme contributes to safe and efficient robot behaviours in crowded and safety-critical scenarios.

*Equal contribution.

¹Authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. {yuwen001, xu0021ng, ruofei001, yizhuo001, shyuan, elhxie}@ntu.edu.sg

²Author is with the Robotics Institute, Carnegie Mellon University. muqingc@andrew.cmu.edu

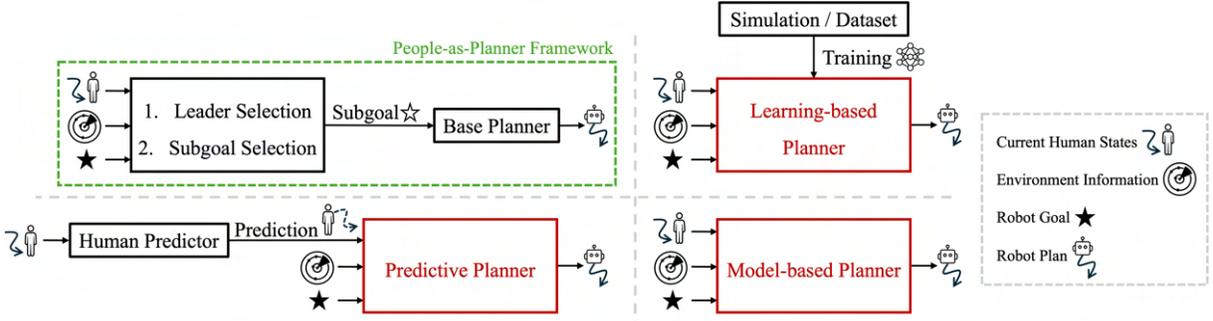


Fig. 2: The proposed People-as-Planner framework, compared with common planning methods in crowd navigation.

II. METHODOLOGY

Problem Definition. The state of an agent consists of its position and velocity $\mathbf{x} = [\mathbf{p}^\top, \mathbf{v}^\top]^\top \in \mathbb{R}^4$. We use subscripts H, R to represent human and robot, and superscript t to represent timesteps. The robot's goal position and environmental obstacles are represented by \mathbf{p}_g and \mathcal{O} . In a crowd navigation setting, the robot plan is generated by a planner \mathcal{P} , denoted as

$$\mathbf{x}_R^{t+1} \leftarrow \mathcal{P}(\mathbf{x}_R^t, \mathbf{x}_H^t, \mathcal{O}, \mathbf{p}_g). \quad (1)$$

Overview. In our proposed scheme, we first identify a human leader H_L^t and subsequently define a subgoal \mathbf{p}_g^t :

$$H_L^t = f_{\text{leader}}(\mathbf{x}_R^t, \mathbf{X}_H^{0:t}, \mathcal{O}, \mathbf{p}_g), \quad (2a)$$

$$\mathbf{p}_g^t = f_{\text{subgoal}}(\mathbf{x}_{H_L^t}^t, \mathbf{x}_H^t), \quad (2b)$$

where $f_{\text{leader}}(\cdot)$ and $f_{\text{subgoal}}(\cdot)$ are the leader and subgoal selection processes. The robot plan is then generated by a base planner \mathcal{P} :

$$\mathbf{x}_R^{t+1} \leftarrow \mathcal{P}(\mathbf{x}_R^t, \mathbf{x}_H^t, \mathcal{O}, \mathbf{p}_g^t), \quad (3)$$

and the process repeats until the robot reaches its goal. Since each subgoal \mathbf{p}_g^t is much easier to reach than the original goal \mathbf{p}_g , we can use a simple and lightweight local planner as the base planner. The overall framework is illustrated in Fig. 2.

A. Leader Selection

We first evaluate the reachability of each human. As we aim to use simple base planners, the selected leader needs to be directly accessible without the need for complex obstacle avoidance. We adopt the Line-of-Sight (LoS) distance metric proposed by [13] to define the reachability, which measures the distance from each human to the boundary of the robot's visible region constructed from the robot's LiDAR scans. We represent the reachability score as f_{reach} and only consider Human i as a potential leader if

$$f_{\text{reach}}(H_i, \mathbf{x}_R^t, \mathbf{x}_H^t, \mathcal{O}) \geq \tau_{\text{reach}}, \quad (4)$$

where τ_{reach} is the reachability threshold.

We then evaluate each qualified human i using three criteria. Based on T previous steps, we first evaluate if the human's average heading is towards the robot's goal \mathbf{p}_g :

$$S_{\text{head}} = \begin{cases} \frac{\tilde{\mathbf{v}}_i \cdot \overrightarrow{\mathbf{p}_i \mathbf{p}_g}}{\|\tilde{\mathbf{v}}_i\| \|\overrightarrow{\mathbf{p}_i \mathbf{p}_g}\|} & \text{if } \frac{\tilde{\mathbf{v}}_i \cdot \overrightarrow{\mathbf{p}_i \mathbf{p}_g}}{\|\tilde{\mathbf{v}}_i\| \|\overrightarrow{\mathbf{p}_i \mathbf{p}_g}\|} \geq \cos \frac{\pi}{4}, \\ -1, & \text{otherwise.} \end{cases} \quad (5)$$

where $\tilde{\mathbf{v}}_i = \frac{1}{T} \sum_{k=t-T+1}^t \mathbf{v}_i^k$.

Second, we compare the human's average speed to the robot's ideal speed v_{pref} :

$$S_{\text{vel}} = \begin{cases} \frac{\tilde{v}_i - v_{\text{pref}}}{v_{\text{pref}}}, & \text{if } \tilde{v}_i < v_{\text{pref}}, \\ \max\left(0, 1 - \frac{\tilde{v}_i - v_{\text{pref}}}{v_{\text{pref}}}\right), & \text{otherwise.} \end{cases} \quad (6)$$

where $\tilde{v}_i = \frac{1}{T} \sum_{k=t-T+1}^t \|\mathbf{v}_i^k\|$. We penalize slower speeds to encourage active progression.

Third, we compare relative positions to identify nearby humans between the robot and its goal:

$$S_{\text{pos}} = \begin{cases} \max\left(0, 1 - \frac{\|\overrightarrow{\mathbf{p}_R \mathbf{p}_i}\|}{r}\right), & \text{if } \frac{\overrightarrow{\mathbf{p}_R \mathbf{p}_i} \cdot \overrightarrow{\mathbf{p}_R \mathbf{p}_g}}{\|\overrightarrow{\mathbf{p}_R \mathbf{p}_i}\| \|\overrightarrow{\mathbf{p}_R \mathbf{p}_g}\|} > 0, \\ -1, & \text{otherwise,} \end{cases} \quad (7)$$

where r is the robot's observable range.

We can now calculate a weighted score S_i by

$$S_i = w_{\text{head}} S_{\text{head}} + w_{\text{vel}} S_{\text{vel}} + w_{\text{pos}} S_{\text{pos}}. \quad (8)$$

where w_{head} , w_{vel} , and w_{pos} are the weights for each score. We add an adjustment term to the previous leader $S_{H_L^{t-1}}$ to avoid fluctuation between candidates with similar scores. The candidate with the highest score is selected as H_L^t .

B. Subgoal Selection

We aim to define a subgoal \mathbf{p}_g^t that allows the robot to follow H_L^t via a straightforward path. We first sample a set of position candidates \mathbf{p}_m between the robot and H_L^t , defined as

$$\mathbf{p}_m = \mathbf{p}_{H_L^t} - R(\theta_m) \cdot \overrightarrow{\mathbf{p}_R \mathbf{p}_{H_L^t}} \cdot \frac{d}{\|\overrightarrow{\mathbf{p}_R \mathbf{p}_{H_L^t}}\|}, \quad (9)$$

where $\theta_m \in \{-\frac{\pi}{4} + m\Delta\theta \mid m = 0, 1, \dots, \lceil \frac{\pi}{2\Delta\theta} \rceil\}$; d is the safe distance from humans; and

$$R(\theta_m) = \begin{bmatrix} \cos(\theta_m) & -\sin(\theta_m) \\ \sin(\theta_m) & \cos(\theta_m) \end{bmatrix}. \quad (10)$$

We select the position that is furthest from neighbours:

$$\mathbf{p}_g^t = \operatorname{argmax}_{\mathbf{p}_m} \left(\min_{H_i \neq H_L^t} \|\overrightarrow{\mathbf{p}_m \mathbf{p}_{H_i}}\| \right), \quad (11)$$

so to minimize the possibility of collision avoidance.

We adjust the robot’s speed in the planner \mathcal{P} to encourage the robot to follow the leader closely. In our experiments, we update the speed limit v_{\max} in SF planner directly:

$$v_{\max} = \begin{cases} \|v_{H_L^t}\|, & \text{if } \|\overrightarrow{\mathcal{P}R\mathcal{P}H_L^t}\| \leq \tau_{\text{catchup}}, \\ v_{\text{catchup}}, & \text{otherwise,} \end{cases} \quad (12)$$

where v_{catchup} is a faster speed to catch up with the leader.

If no H_L is identified in the previous leader selection step, we will set $\mathbf{p}_g^t = \mathbf{p}_g$ where it will plan towards the robot’s goal directly. We will show in Sec. III-C that this happens when there are only a few surrounding humans and using a simple SF planner is sufficient.

III. EXPERIMENTS

A. Simulation Settings



Fig. 3: Three test scenes from the SDD dataset. The inter-changeable initial and goal positions are shown in red stars.

Scenarios. We first perform experiments in simulation environments with real-world trajectory data from the Stanford Drone Dataset (SDD) [18]. SDD was collected mainly from three scenes and we name them Promenade, Crossing, and Roundabout, as shown in Fig. 3. For each scene, we select 10 crowded segments from the longest video. We use the densest segment from each scene for comparison study.

Implementation Details. All parameters can be found in our code repository. We use the same set of tunable parameters for all experiments. Since not all baseline methods distinguish heterogeneous road users, we set the radius of all agents (robot, humans, bicycles, and cars) to 0.5m for fair comparisons. For collision evaluation, we provide two sets of agent dimensions: 1) all agents have the same radius of 0.5m, 2) bicycles and vehicles are given more realistic dimensions of $1.9\text{m} \times 1\text{m}$ and $4.5\text{m} \times 1.9\text{m}$ respectively. We report the results for both settings in Sec. III-B.

Evaluation Metrics. We perform 100 repeated runs for each segment and evaluate the average performance. The following three metrics are used for evaluation:

- 1) Total Collision Count (**TCC**) is the number of frames where collision occurs. Following existing work [19], we allow the experiment to continue after a collision.
- 2) Average Time (\mathbf{T}_{avg}) taken to reach the goal.
- 3) Average Distance (\mathbf{D}_{avg}) taken to reach the goal.

B. Quantitative Evaluation

As shown in Table. I, our method shows the best safety awareness with high efficiency across all three scenes. Our method allows the robot to adjust its speed according to the leader, so it can reach the goal fast even when additional

manoeuvres are required for effective collision avoidance. Notice that although we set the radius for all agents to 0.5m during the experiments, when we calculate collisions using real dimensions for other road users (namely bicycles and vehicles), our method maintains a low TCC, while the values increase dramatically for all other baselines. This implies that our method can keep a larger distance from dangerous road users without explicitly considering agent types in the inputs, which benefits from the human leader’s planning decisions. We also show that by simplifying complex crowd navigation problems using people as planners, we can achieve the best performance using a simple base planner SF, which is less capable when planning independently.

We summarise the types of information and procedures required to deploy each method in Table. II. We can see that our method only needs the basic inputs to achieve the best overall performance. Our method is further evaluated over all 30 segments from the three scenes, as shown in Table. III. We demonstrate strong robustness as the performance is consistent under different traffic conditions and crowd densities.

C. Qualitative Evaluation

Using people as planners, our method demonstrates efficient and socially-aware behaviours that contribute to the outstanding results presented above. For the Promenade scene in Fig. 4-1a, the robot switched to the second leader when the first leader was temporarily occluded, and switched back to the first leader when the second leader stopped its movement. For the Crossing scene, both SF and HEIGHT resulted in collisions with a fast-moving bicycle, while our method, as shown in Fig. 4-2a, avoided this encounter by switching between the two leaders. For the Roundabout scene in Fig. 4-3a, an interesting observation is that although traffic rules are not explicitly defined, our method follows the human to cross the road from the appropriate point on the sidewalk, which avoids a dangerous encounter with the bicycle. Other baselines, however, move towards the goal directly, which inevitably increases the difficulty of avoiding fast-moving road users on the main road. This makes our method especially useful in real-world deployments, as it is difficult to define every rule in every scenario, our method can simply follow the surrounding agents to respect the “unknown rules”.

D. Real-World Experiment

We further demonstrate our method when interacting with real-world humans in a 150m crowded campus corridor.

In Fig. 5, we illustrate four events, most of which have been observed in the simulation experiments. In the first event, the robot switched to a new leader when the previous leader slowed down. This ensures the robot continues to progress efficiently without being stuck behind static humans. In the second and fourth events, the robot moved through a complex crowd and a narrow gate without deadlocks. In the third event, the robot followed the social norm to keep to the left, which avoided encounters with humans walking from the opposite direction. The robot benefits from

Method	Promenade			Crossing			Roundabout		
	TCC↓	T _{avg} (s)↓	D _{avg} (m)↓	TCC↓	T _{avg} (s)↓	D _{avg} (m)↓	TCC↓	T _{avg} (s)↓	D _{avg} (m)↓
SF [6]	2.66 / 50.13	50.11	62.99	125.46 / 234.13	39.19	54.57	13.42 / 77.21	22.53	31.37
DWA [14]	44.14 / 114.79	48.56	67.03	8.00 / 182.59	29.98	42.43	29.85 / 141.55	27.69	39.16
ORCA [15]	8.00 / <u>43.11</u>	56.62	74.81	0.00 / 80.07	<u>29.70</u>	41.87	<u>2.94</u> / <u>53.07</u>	28.41	35.03
Pred2Nav+CV [16]	122.71 / 125.68	48.45	57.34	18.9 / <u>40.68</u>	32.64	43.50	16.27 / 58.12	19.84	29.37
Pred2Nav+SGAN [16]	138.36 / 139.65	<u>43.92</u>	<u>56.06</u>	71.47 / 117.06	31.98	44.22	23.87 / 83.66	19.40	30.05
HEIGHT [17]	605.68 / 605.68	<u>56.08</u>	43.35	171.45 / 270.46	29.91	44.45	12.17 / 71.76	<u>19.41</u>	26.17
Ours	0.25 / 1.94	40.39	60.48	0.00 / 0.00	29.40	45.10	0.54 / 5.44	19.87	<u>29.17</u>

TABLE I: Quantitative comparison results. For TCC, the first and second values are calculated by setting all agents with the same radius of 0.5m, and setting bicycles and vehicles with realistic dimensions, respectively. TCC is the frames with collisions under 30Hz simulation. The best performance is in **bold** and the second-best is underlined.

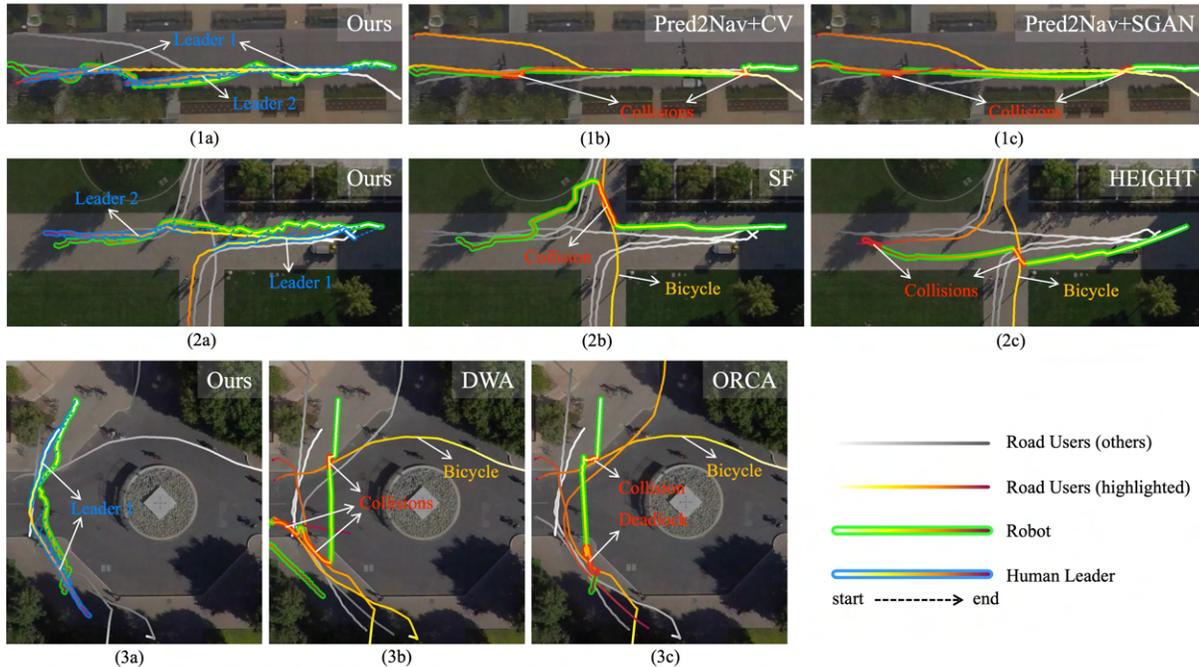


Fig. 4: Qualitative comparisons in simulation experiments. For simplicity, we select a few agents to draw for each scene to highlight the interactions. The full experiment recordings can be found in the supplementary video.

Method	Current human states	Future human states	Environment information	Training
SF [6]	✓	×	✓	×
DWA [14]	×	×	✓	×
ORCA [15]	✓	×	✓	×
Pred2Nav+CV [16]	✓	✓	×	×
Pred2Nav+SGAN [16]	✓	✓	×	✓
HEIGHT [17]	✓	×	✓	✓
Ours	✓	×	✓	×

TABLE II: Comparison of deployment requirements.

Scene	TCC↓	T _{avg} (s)↓	D _{avg} (m)↓
Promenade	2.14 ± 4.53	41.09 ± 3.32	59.04 ± 1.80
Crossing	3.74 ± 5.22	30.53 ± 2.73	46.64 ± 1.89
Roundabout	3.97 ± 3.37	19.09 ± 2.58	28.80 ± 2.46

TABLE III: Quantitative results on all 30 segments. For TCC, we use realistic dimensions for bicycles and vehicles.

the proposed people-as-planner scheme and makes intelligent decisions in real-world crowded environments. Our method is robust to disturbances such as inaccurate human tracking, hardware delays, etc., and successfully completes the run without collisions. The complete experiment can be found

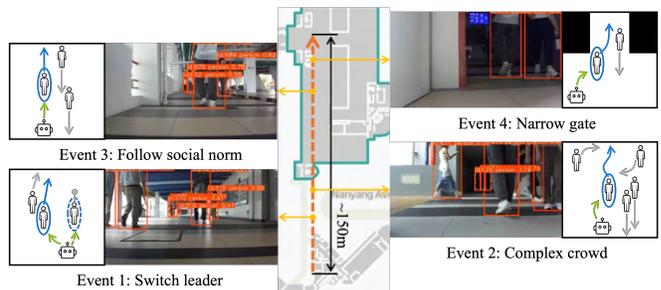


Fig. 5: Events observed in the real-world experiment.

in the supplementary video.

IV. CONCLUSIONS

This paper presents a novel crowd navigation scheme which utilises people as external intelligent planners. The more complex planning decisions are handed over to humans so the robot's task is simplified into straightforward subgoal planning. The experiments demonstrate robot behaviours that are safe, efficient, and socially compliant in various crowded environments. Our proposed framework offers new insights into human-robot relationships in social navigation.

REFERENCES

- [1] C. Mavrogiannis, F. Baldini, A. Wang, D. Zhao, P. Trautman, A. Steinfeld, and J. Oh, "Core challenges of social robot navigation: A survey," *ACM Transactions on Human-Robot Interaction*, vol. 12, no. 3, pp. 1–39, 2023.
- [2] M. Lewis, H. Wang, P. Velagapudi, P. Scerri, and K. Sycara, "Using humans as sensors in robotic search," in *2009 12th International Conference on Information Fusion*. IEEE, 2009, pp. 1249–1256.
- [3] O. Afolabi, K. Driggs-Campbell, R. Dong, M. J. Kochenderfer, and S. S. Sastry, "People as sensors: Imputing maps from human actions," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 2342–2348.
- [4] M. Itkina, Y.-J. Mun, K. Driggs-Campbell, and M. J. Kochenderfer, "Multi-agent variational occlusion inference using people as sensors," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 4585–4591.
- [5] Y.-J. Mun, M. Itkina, S. Liu, and K. Driggs-Campbell, "Occlusion-aware crowd navigation using people as sensors," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 031–12 037.
- [6] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, p. 4282, 1995.
- [7] S. Leisiazar, S. R. R. Rohani, E. J. Park, A. Lim, and M. Chen, "Adapting to frequent human direction changes in autonomous frontal following robots," *IEEE Robotics and Automation Letters*, 2025.
- [8] C.-P. Lam, C.-T. Chou, K.-H. Chiang, and L.-C. Fu, "Human-centered robot navigation—towards a harmoniously human–robot coexisting environment," *IEEE Transactions on Robotics*, vol. 27, no. 1, pp. 99–112, 2010.
- [9] K. Morioka, Y. Oinaga, and Y. Nakamura, "Control of human-following robot based on cooperative positioning with an intelligent space," *Electronics and Communications in Japan*, vol. 95, no. 1, pp. 20–30, 2012.
- [10] A. Antonucci, P. Bevilacqua, S. Leonardi, L. Paolopoli, and D. Fontanelli, "Humans as path-finders for mobile robots using teach-by-showing navigation," *Autonomous Robots*, vol. 47, no. 8, pp. 1255–1273, 2023.
- [11] R. Algabri and M.-T. Choi, "Deep-learning-based indoor human following of mobile robot using color feature," *Sensors*, vol. 20, no. 9, p. 2699, 2020.
- [12] M. Gupta, S. Kumar, L. Behera, and V. K. Subramanian, "A novel vision-based tracking algorithm for a human-following mobile robot," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 7, pp. 1415–1427, 2016.
- [13] R. Bai, S. Yuan, K. Li, H. Guo, W.-Y. Yau, and L. Xie, "Realm: Real-time line-of-sight maintenance in multi-robot navigation with unknown obstacles," 2025. [Online]. Available: <https://arxiv.org/abs/2502.15162>
- [14] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.
- [15] J. Van Den Berg, S. J. Guy, M. Lin, and D. Manocha, "Reciprocal n-body collision avoidance," in *Robotics Research: The 14th International Symposium ISRR*. Springer, 2011, pp. 3–19.
- [16] S. Poddar, C. Mavrogiannis, and S. S. Srinivasa, "From crowd motion prediction to robot navigation in crowds," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 6765–6772.
- [17] S. Liu, H. Xia, F. C. Pouria, K. Hong, N. Chakraborty, and K. Driggs-Campbell, "Height: Heterogeneous interaction graph transformer for robot navigation in crowded and constrained environments," *arXiv preprint arXiv:2411.12150*, 2024.
- [18] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VIII 14*. Springer, 2016, pp. 549–565.
- [19] M. Cao, X. Xu, Y. Yang, J. Li, T. Jin, P. Wang, T.-Y. Hung, G. Lin, and L. Xie, "Learning dynamic weight adjustment for spatial-temporal trajectory planning in crowd navigation," *arXiv preprint arXiv:2412.00555*, 2024.