EmoBipedNav: Emotion-aware Social Navigation for Bipedal Robots with Deep Reinforcement Learning

Wei Zhu[1], Abirath Raju[1], Abdulaziz Shamsah[2], Anqi Wu[3], Seth Hutchinson[4], and Ye Zhao[1]

***Abstract*— This study presents an emotion-aware navigation framework – EmoBipedNav – using deep reinforcement learning (DRL) for bipedal robots walking in socially interactive environments. Specifically, social navigation scenarios are represented using sequential LiDAR grid maps (LGMs), from which we extract latent features, including collision regions, emotion-related discomfort zones, social interactions, and the spatio-temporal dynamics of evolving environments. The extracted features are directly mapped to the actions of reduced-order models (ROMs) through a DRL architecture. Furthermore, the proposed framework incorporates full-order dynamics and locomotion constraints during training, effectively accounting for tracking errors and restrictions of the locomotion controller while planning the trajectory with ROMs. The hardware videos and open-source code are available at `https://gatech-lidar.github.io/emobipednav.github.io/`.**

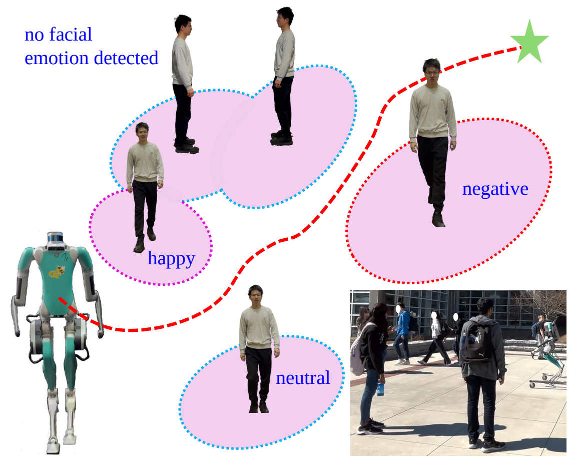

Fig. 1. Emotion-aware social navigation with the bipedal robot Digit. Digit is required to maintain customized comfort distances from pedestrians with specific emotions while navigating toward a designated goal location.

I. INTRODUCTION

The research and development of bipedal robots have garnered significant interest within the robotics community, mainly due to their human-like morphology and versatile locomotion capabilities [1]–[6]. While a great number of studies have focused on developing stable locomotion controllers for bipedal robots, one goal is to enable bipedal robots to autonomously and courteously navigate pedestrian-populated environments within social contexts, so as for practical service in human societies. However, achieving safe and socially aware navigation remains a formidable challenge, originating from the intricate dynamics of bipedal robots, limited maneuvering capabilities in dynamic environments, and the complexities of understanding the intentions and interactions of pedestrians.

Although significant efforts have been dedicated to the use of wheeled mobile robots for social navigation [7], [8], few studies address bipedal social navigation due to challenges in developing stable and precise locomotion controllers for their hybrid, highly nonlinear, and high-degree-of-freedom dynamics. Furthermore, complex social interactions increase the level of difficulty because they are too implicit to be accurately modeled. In addition, social cues, such as human emotions, psychologically influence the preferred personal space during interactions with other agents, including pedestrians and robots [9], [10], as illustrated in Fig. 1. These social interactions and cues remain underexplored for bipedal robot navigation.

Cutting-edge bipedal navigation studies in crowded environments decouple navigation and locomotion, planning navigation trajectories with reduced-order models (ROMs) at first, and then tracking the trajectories with locomotion controllers based on full-order dynamics [11], [12]. Such a planning architecture cannot take into account the tracking errors and constraints of low-level controllers, such as joint torque limits, for bipedal navigation. In addition, pedestrian trajectory prediction is also individually executed [12] because prediction models are pretrained using specific datasets only including pedestrian-pedestrian interactions, ignoring pedestrian-robot interactions. Such pretrained models might further cause deployment discrepancies in bipedal robots.

In this study, we present an emotion-aware navigation framework – EmoBipedNav – using deep reinforcement learning (DRL) for bipedal robots walking in socially interactive environments. The main contributions of this study are as follows.

- **Novel DRL observation representation**. We introduce LiDAR grid maps (LGMs) as a novel representation of highly dynamic environments integrated with pedestrian emotions.
- **Integration of pedestrian emotions**. Our framework incorporates pedestrian emotions to design discomfort

[1]The authors are with the Laboratory for Intelligent Decision and Autonomous Robots, Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30313, USA. {wzhu328, araju60, yezhao}@gatech.edu

[2]College of Engineering and Petroleum, Kuwait University, PO Box 5969, Safat, 13060, Kuwait. abdulaziz.shamsah@ku.edu.kw

[3]School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA. anqiwu@gatech.edu

[4]Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA. s.hutchinson@northeastern.edu

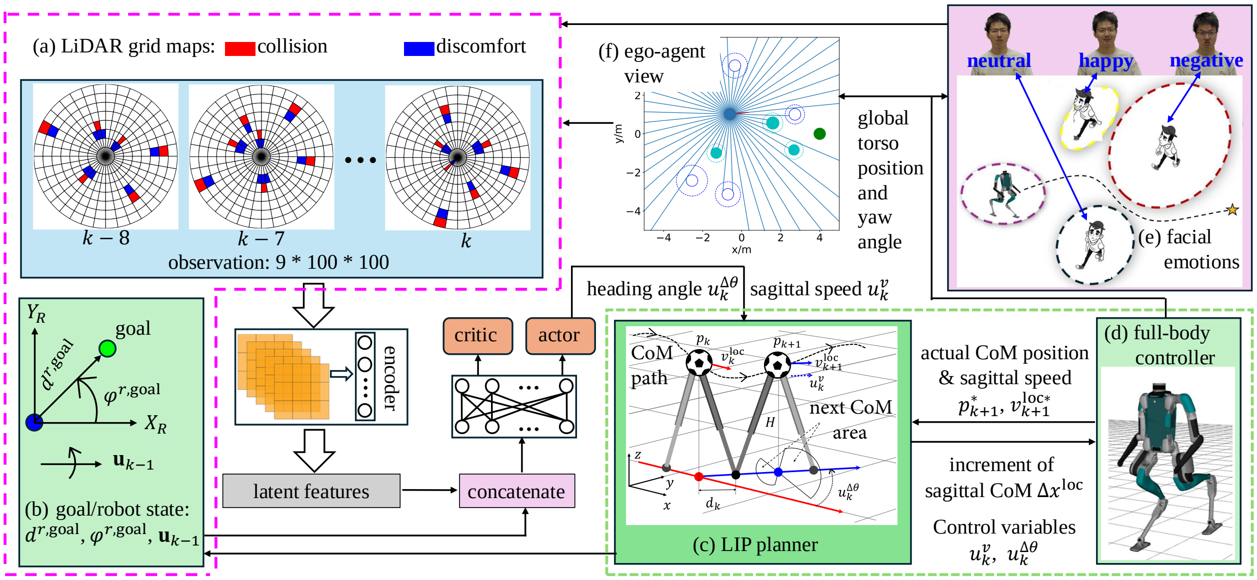

Fig. 2. Overview of the proposed EmoBipedNav framework using the bipedal robot Digit. Our framework begins by obtaining estimated facial emotions (Fig. (e)) using pre-trained CNN models. Simultaneously, we transform raw LiDAR scans (Fig. (f)) into sequential pie-shape LGMs (Fig. (a)), where the red cells denote collision areas and the blue patches highlight discomfort zones associated with pedestrian emotions. These grid maps are converted into stacked pixel images which are further processed through an encoder constructed using CNNs to extract socially interactive and emotionally aware features. The resulting latent features are concatenated with the robot's last command and target position (Fig. (b)), which are fed into an actor-critic DRL structure implemented with multi-layer perceptrons (MLPs). The action output from the actor network is derived from the ROM (Fig. (c)) and practically applied to a bipedal robot with full-body dynamics and constraints (Fig. (d)). The torso position and yaw angle obtained from Digit correspond to the ego-agent state in Fig. (f). We use the Angular momentum LIP planner (ALIP) [13] and a passivity full-body controller with ankle actuation [14] to track the desirable ROM trajectory.

zones and further enable emotion-aware navigation.

- **End-to-end navigation framework**. We propose an end-to-end navigation pipeline that directly maps latent features to bipedal robot actions and is trained with full-body bipedal robot dynamics, effectively mitigating model discrepancies induced by ROMs.

II. APPROACH

In this section, we first introduce the reduced-order model (ROM) of bipedal locomotion and the Partially Observable Markov Decision Process (POMDP) of the social navigation problem. Subsequently, we present our DRL-based approach, detailing the novel observation space, the DRL network architecture, and the design of the integrated reward functions. The overall framework is illustrated in Fig. 2.

A. Preliminaries

The dynamics of bipedal robots is commonly approximated using ROMs, such as the linear inverted pendulum (LIP) model [13], [15], as depicted in Fig. 2-(c). The discrete state transition equation for the LIP model at the k^{th} walking step in the local sagittal frame is given as follows:

$$\begin{aligned} d_k &= \frac{v_k^{\text{loc}} \cosh(\omega T) - u_k^v}{\omega \sinh(\omega T)}, \\ \Delta x^{\text{loc}}(d_k) &= v_k^{\text{loc}} \frac{\sinh(\omega T)}{\omega} + d_k\left(1 - \cosh(\omega T)\right), \end{aligned} \tag{1}$$

where the control input u_k^v represents the desired sagittal velocity at the next foot placement switch instant, the output state $\Delta x^{\text{loc}}(d_k)$ is the local sagittal center-of-mass (CoM) position increment between two consecutive walking steps, v_k^{loc} denotes the sagittal velocity in the local coordinate at the k^{th} walking step switch instant, d_k is an auxiliary variable that specifies the sagittal foot distance relative to CoM, T is the constant footstep time, the term $\omega = \sqrt{g/H}$ is derived from the gravitational constant g and the desired CoM height H.

Since u_k^v influences only the translational motion, we introduce a heading angle $u_k^{\Delta\theta}$ to steer the CoM orientation in the horizontal plane, as illustrated in Fig. 2-(c). Combining the translational and rotational motions of the CoM, we define the action of our navigation policy as:

$$\mathbf{u}_k = \{u_k^v, u_k^{\Delta\theta}\},\ 0 \le u_k^v \le \bar{u}^v,\ \mid u_k^{\Delta\theta} \mid \le \bar{u}^{\Delta\theta}, \tag{2}$$

where $\bar{u}^v$ and $\bar{u}^{\Delta\theta}$ represent the maximum allowable sagittal speed and heading angle, respectively.

Given the control $\mathbf{u}_k$, the CoM displacement Δx^{loc}, and the LIP model, the CoM position is propagated from p_k to p_{k+1}, and the sagittal speed changes to v_{k+1}^{loc} from v_k^{loc}. However, there are discrepancies between the actual CoM position p^*_{k+1} and p_{k+1}, and between the actual sagittal speed $v_{k+1}^{\text{loc}*}$ and v_{k+1}^{loc}. More specifically, the full-body controller generates joint torques to track the desired CoM displacement Δx^{loc} and the control $\mathbf{u}_k$, as illustrated in Fig. 2-(d). However, discrepancies between the LIP model and the full-body dynamics can cause the robot's states to deviate from the desired ones, which may lead to collisions and intrusions into discomfort areas of pedestrians, as demonstrated in Section III.

B. Problem Formulation

We formulate the social navigation problem as a POMDP. More specifically, we represent social environments with sequential LiDAR grid maps (LGMs). Despite only including partial information, LGMs allow the extraction of implicit features such as collision zones, pedestrian-pedestrian and pedestrian-robot interactions, emotion-related discomfort areas, and the spatio-temporal dynamic evolution of the navigation system. In addition to LGMs which only contain partial information, we have two other fully observable states. For one thing, the goal position state is responsible for the target-directed navigation task. In addition, since pedestrians interact with the robot and the robot's actions influence pedestrian decisions, we also include the robot's actions as part of the states.

Given the observation of LGMs and the fully observable states, our objective is to optimize a navigation policy, straightforwardly projecting the observation and states into the robot actions. The problem is formulated as a POMDP defined by a 6-tuple $(S, O, A, \Gamma, R, \gamma)$, where S is the fully observable state space including the goal position and the robot's actions, O represents the observation space consisting of sequential LGMs, A stands for the action space, Γ denotes the state transition model fully propagated with the bipedal robot Digit, R is the reward function with an instantaneous scalar reward r_k at the k^{th} walking step and γ is the discount factor. To derive the optimal navigation policy $\pi(a \in A|s \in S, o \in O)$, our goal is to maximize the expected sum of discounted rewards over an infinite time horizon, as follows:

$$V = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k r_k \right]. \tag{3}$$

We will leverage an encoder composed of convolutional neural networks (CNNs), followed by a soft actor-critic (SAC) DRL algorithm [16] to iteratively optimize the navigation policy π.

C. Observation and State Spaces

We propose a novel representation for the observation space, referred to as pie-shape LGMs (Fig. 2-(a)), to comprehensively encapsulate the features of all entities around the ego-agent, including static obstacles having variable sizes and dynamic pedestrians with different emotions. Sequential pie-shape LGMs are built on the basis of raw LiDAR scans. More specifically, a 2-D LiDAR scan with K beams is emitted from the ego-agent to detect collision and discomfort margins, as illustrated in Fig. 2-(f).

The detected obstacles are projected onto a pie-shape grid map (shown in Fig. 2-(a)) with a maximum detection radius of L. This map is divided into M radial segments, forming M concentric rings. Each ring is further segmented angularly into M equal sections. The resolution of each grid increases as the ring gets closer to the center, ensuring that more details around the ego-agent are detected.

In particular, we define a discomfort margin (blue grids in Fig. 2-(a)) that expands according to the collision area (red cells in Fig. 2-(a)). The size of this discomfort zone is adjusted according to the emotional state of the pedestrian. We consider three emotional states: `happy`, `neutral`, and `negative` including angry, sad, fearful, and surprised. According to psychological research [10], happy emotions encourage approach behaviors and therefore shorter distances, whereas negative emotions promote avoidant behaviors and longer distances. Based on the studies in [10], [17], we assign the following distances from the robot margin to the pedestrian edge: 0.2 m for `happy`, 0.35 m for `neutral`, and 0.5 m for `negative` emotions.

The observation $\mathbf{o}_k$ at the current walking step is the LGM capturing only the static immediate environment in the spatial dimensio. To capture dynamic information in the temporal dimension, we incorporate previous LGMs from time step $k-N+1$ to $k-1$ as illustrated in Fig. 2-(a). Thus, the observation in both the spatial and temporal dimensions is represented as $\mathbf{o}_k^{\text{sp-tem}} = \{\mathbf{o}_{k-N+1}, ..., \mathbf{o}_{k-1}, \mathbf{o}_k\}$.

In addition to the observation only including partial information, we define another two fully observable states as illustrated in Fig. 2-(b): the position state of the goal $\mathbf{s}_k^{\text{goal}} = \{d^{r,\text{goal}}, \varphi^{r,\text{goal}}\}$, where $d^{r,\text{goal}}$ represents the distance between the goal and the robot and $\varphi^{r,\text{goal}}$ indicates the orientation of the goal in the robot frame; the last robot command $\mathbf{u}_{k-1}$.

D. Network Structure

The encoder, as shown in Fig. 2, consists of CNNs followed by multi-layer perceptrons (MLPs) to fully extract the latent features $\mathbf{f}_k^o$ from the sequential LGMs $\mathbf{o}_k^{\text{sp-tem}}$. Given these latent features $\mathbf{f}_k^o$, we concatenate them with the goal state $\mathbf{s}_k^{\text{goal}}$ and the action state $\mathbf{u}_{k-1}$. The combined inputs are fed into the SAC DRL framework [16] to optimize the navigation policy $\pi\left(\mathbf{u}_k|\mathbf{o}_k^{\text{sp-tem}}, \mathbf{s}_k^{\text{goal}}, \mathbf{u}_{k-1}\right)$. The actor and value modules, as shown in Fig. 2, are made up of MLPs that take the same inputs $\{\mathbf{f}_k^o, \mathbf{s}_k^{\text{goal}}, \mathbf{u}_{k-1}\}$. In summary, the network structure can be expressed as follows:

$$\mathbf{f}_k^o = f_\vartheta(\mathbf{o}_k^{\text{sp-tem}}), \tag{4a}$$
$$v_k = f_\phi(\mathbf{f}_k^o, \mathbf{s}_k^{\text{goal}}, \mathbf{u}_{k-1}), \tag{4b}$$
$$\mathbf{u}_k \sim f_\psi(\mathbf{f}_k^o, \mathbf{s}_k^{\text{goal}}, \mathbf{u}_{k-1}), \tag{4c}$$

where the network parameters ϑ, ϕ, ψ are updated using the SAC algorithm [16] and the exploration samples collected from the bipedal robot Digit, v_k represents the deterministic value function predicted by the critic network, and $\mathbf{u}_k$ is the action sampled from a Gaussian distribution with its mean and standard deviation parameters output from the actor network.

E. Reward Integration

The objectives of our navigation task are threefold: avoidance of collisions, reaching goals, and awareness of emotions. Accordingly, we define the reward function as follows:

$$r_k = r_k^{\text{col}} + r_k^{\text{goal}} + r_k^{\text{emo}}, \tag{5}$$

where r_k^{col} penalizes the collisions with all surrounding objects, r_k^{goal} rewards progress toward reaching the goal, and r_k^{emo} denotes an emotion-related penalty. Specifically, these three rewards are defined as follows.

1) Collision avoidance reward:

$$r_k^{\text{col}} = \begin{cases} -0.6 & \text{if } d_{\min}^{\text{all}} \leq r^{\text{robot}}, \\ -0.1 & \text{else if } d_{\min}^{\text{static}} < d^{\text{static}} + r^{\text{robot}}, \\ 0.0 & \text{else}, \end{cases} \tag{6}$$

where $d_{\min}^{\text{all}}$ represents the minimum distance from the ego-agent to the margins of all surrounding objects, $d_{\min}^{\text{static}}$ denotes the minimum distance from the ego-agent to all static obstacles, d^{static} stands for a safety distance expanded from the collision margin of the obstacle, and r^{robot} is the ego-agent radius.

2) Goal reaching reward:

$$r_k^{\text{goal}} = \begin{cases} 0.5 & \text{if } d_k^{\text{goal}} \leq d^{\text{goal}}, \\ 0.3 \cdot (d_{k-1}^{\text{goal}} - d_k^{\text{goal}}) & \text{else}, \end{cases} \tag{7}$$

where d_k^{goal} stands for the distance from the ego-agent to the goal at the k^{th} walking step, and d^{goal} represents the goal-reaching threshold. In particular, the second term encourages the robot to move toward the goal, thus alleviating the sparsity issue of the reward function and promoting efficient training.

3) Emotion-aware reward:

$$r_k^{\text{emo}} = \begin{cases} -0.1 & \text{if } \mathcal{C}, \\ 0.0 & \text{else}, \end{cases} \tag{8}$$

where $\mathcal{C}$ is a distance condition with respect to pedestrian emotions. Specifically, the discomfort distance of the i^{th} pedestrian is defined as $d_{p_i}^{\text{emo}} \in \{0.2, 0.35, 0.5\}$m, depending on the corresponding emotion of the pedestrian $\{\texttt{happy}, \texttt{neutral}, \texttt{negative}\}$. r^{ped} is a constant pedestrian radius. Given the distance d_{p_i} from the margin of the i^{th} pedestrian to the ego-agent, the penalty is -0.1 if $\exists\, i$, such that $d_{p_i} < d_{p_i}^{\text{emo}} + r^{\text{robot}}$. Note that, to avoid learning a conservative navigation policy, we only count once if the ego-agent enters the discomfort zones of several objects, including static obstacles and dynamic pedestrians, at the same walking step. That means we add -0.1 only once in Eq. (5) although it appears in both Eq. (6) and Eq. (8).

III. Experiments

We first train and test our navigation policy in MuJoCo [18] with a full-body bipedal robot Digit. Next, we demonstrate various deployments on the robot hardware.

A. Simulation Results

Evaluation Metrics. We evaluate each benchmark method over 500 random trials. The success rate (`SR`) is defined as the ratio of successful episodes that occur without collision or timeout. Individual discomfort times (`IDT`) quantifies the discomfort times for individual entities. Specifically, the four values in the `IDT` column of TABLE I correspond to the `IDT` of static obstacles and pedestrians exhibiting `happy`, `neutral`, and `negative` emotions, respectively.

TABLE I

Performance evaluation w/o emotions and w/o full-body robot dynamics with 500 random tests.

DD (Train)	Train/Test	`IDT` (Test)	`SR`
0.2 m	Digit/Digit	[**51**; 50, 181, 1352]	0.950
0.5 m	Digit/Digit	[395; 81, **138**, **1167**]	0.938
Var-DD	Digit/Digit	[213; 56, 166, 1198]	0.940
Var-DD	LIP/LIP	[202; **37**, 143, 1227]	**0.962**
Var-DD	LIP/Digit	[642; 66, 177, 1266]	0.900

DD: discomfort distance (m); Var: variable; Obs: observation
Var-DD= $[0.2, 0.35, 0.5]$ m $\leftrightarrow$ [`happy`, `neutral`, `negative`]

Compared to maintaining a constant discomfort distance of 0.2 m, our approach of using variable discomfort distances of $[0.2, 0.35, 0.5]$ m decreases `IDT` for `neutral` and `negative` pedestrians while increasing `IDT` for `happy` pedestrians, and static obstacles in particular. This indicates that our approach prioritizes `neutral` and `negative` pedestrians by assigning them longer discomfort distances. However, this comes at the cost of reduced comfort for `happy` pedestrians and, notably, static obstacles. The performance degradation with respect to static obstacles may result from the reduction in the available free areas. When the constant discomfort distance is increased to 0.5 m, the ego-agent further improves `IDT` for `neutral` and `negative` pedestrians, yet compromising comfort for `happy` pedestrians and static obstacles. Additionally, while the pipeline trained and tested with the LIP model achieves the highest overall performance, its `SR` drops to 0.9, and `IDT` increases substantially when tested with Digit, highlighting the limitations of ROMs in maintaining robust navigation performance.

B. Hardware Implementations

We begin by transferring the simulated policy learned in MuJoCo [18] to the physical robot Digit. Furthermore, we include various pedestrian motion patterns to verify the sim-to-real capabilities of our emotion-aware navigation policy. For example, a pedestrian intentionally walks in front of Digit and remains stationary to obstruct its path. Additionally, pedestrians move ingroup together in the right column. More pedestrian motion patterns are attached in the accompanying video to further validate the sim-to-real capabilities of our pipeline.

IV. CONCLUSIONS

In this study, we present a DRL-based navigation framework designed to enable bipedal robots to navigate socially interactive environments integrated with pedestrian emotions. Future research will focus on incorporating more realistic social norms, such as adhering to conventions like “walking on the right”, and conducting DRL training in increasingly complex social scenarios, including walls and grouped pedestrians. The ultimate goal is to facilitate natural and socially compliant locomotion for bipedal robots in crowded environments.

References

[1] Z. Gu, J. Li, W. Shen, W. Yu, Z. Xie, S. McCrory, X. Cheng, A. Shamsah, R. Griffin, K. Liu, and others, "Humanoid locomotion and manipulation: Current progress and challenges in control, planning, and learning," *arXiv preprint arXiv:2501.02116*, 2025.

[2] F. Wu, Z. Gu, H. Wu, A. Wu, and Y. Zhao, "Infer and adapt: bipedal locomotion reward learning from demonstrations via inverse reinforcement learning," in *IEEE International Conference on Robotics and Automation*, 2024.

[3] K. Muenprasitivej, J. Jiang, A. Shamsah, S. Coogan, and Y. Zhao, "Bipedal safe navigation over uncertain rough terrain: unifying terrain mapping and locomotion stability," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2024.

[4] Z. Li, J. Zeng, S. Chen, and K. Sreenath, "Autonomous navigation of underactuated bipedal robots in height-constrained environments," *The International Journal of Robotics Research*, vol. 42, no. 8, pp. 565–585, 2023.

[5] K. S. Narkhede, A. M. Kulkarni, D. A. Thanki, and I. Poulakakis, "A sequential mpc approach to reactive planning for bipedal robots using safe corridors in highly cluttered environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11831–11838, 2022.

[6] G. Gibson, O. Dosunmu-Ogunbi, Y. Gong, and J. Grizzle, "Terrain-adaptive, alip-based bipedal locomotion controller via model predictive control and virtual constraints," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 6724–6731.

[7] P. T. Singamaneni, et al., "A survey on socially aware robot navigation: Taxonomy and future challenges," *The International Journal of Robotics Research*, vol. 43, no. 10, pp. 1533-1572, 2024.

[8] K. Zhu and T. Zhang, "Deep reinforcement learning based mobile robot navigation: A review," *Tsinghua Science and Technology*, vol. 26, no. 5, pp. 674-691, 2021.

[9] J. R. Martinez, A. Spalanzani, and C. Laugier, "From proxemics theory to socially-aware navigation: A survey," *International Journal of Social Robotics*, vol. 7, pp. 137-153, 2015.

[10] G. Ruggiero, et. al., "The effect of facial expressions on peripersonal and interpersonal spaces," *Psychological research*, vol. 81 pp. 1232-1240, 2017.

[11] M. Ciocca, P. B. Wieber, and T. Fraichard, "Effect of planning period on MPC-based navigation for a biped robot in a crowd," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 491-496.

[12] A. Shamsah, K. Agarwal, N. Katta, A. Raju, S. Kousik, and Y. Zhao, "Socially acceptable bipedal robot navigation via social zonotope network model predictive control," *IEEE Transactions on Automation Science and Engineering*, doi: 10.1109/TASE.2024.3519012, 2024.

[13] Y. Gong, and J. Grizzle, "One-step ahead prediction of angular momentum about the contact point for control of bipedal locomotion: Validation in a lip-inspired controller," in *IEEE International Conference on Robotics and Automation*, 2022, pp. 2832-2838.

[14] A. Shamsah, Z. Gu, J. Warnke, S. Hutchinson, and Y. Zhao, "Integrated task and motion planning for safe legged navigation in partially observable environments," *IEEE Transactions on Robotics*, vol. 39, no. 6, pp. 4913-4934, 2023.

[15] S. Kajita, F. Kanehiro, K. Kaneko, K. Yokoi, and H. Hirukawa, "The 3D linear inverted pendulum mode: a simple modeling for a biped walking pattern generation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2001.

[16] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor", in *International Conference on Machine Learning*, 2018.

[17] V. Narayanan, B. M. Manoghar, R. P. RV, and A. Bera, "EWareNet: Emotion-aware pedestrian intent prediction and adaptive spatial profile fusion for social robot navigation", in *IEEE International Conference on Robotics and Automation*, 2023.

[18] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *IEEE/RSJ international conference on intelligent robots and systems*, 2012.